University of Connecticut

Department of Economics Working Paper Series

Who Pays for Training? Theory and Evidence on Firm-Level Differences in Training Investments

by

Xiao Ma
Peking University

Alejandro Nakab
Universidad Torcuato Di Tella

Daniela Vidart
University of Connecticut

Working Paper 2025-05
April 2025

365 Fairfield Way, Unit 1063
Storrs, CT 06269-1063
Phone: (860) 486-3022
Fax: (860) 486-4463
http://www.econ.uconn.edu/

Who Pays for Training? Theory and Evidence on Firm-Level Differences in Training Investments*

Xiao Ma
Peking University

Alejandro Nakab
Universidad Torcuato Di Tella

Daniela Vidart
University of Connecticut

April 2025

Abstract

We investigate how on-the-job training varies with firm characteristics and how this informs the distribution of training costs between firms and workers. Using data from over 100 countries, we document that smaller firms consistently offer fewer training opportunities to their workers. Drawing on administrative data from China and Mexico, we identify differences in labor share and productivity levels as key factors explaining this pattern. We then build a general equilibrium model with various training cost-sharing schemes and show that only those in which firms bear a substantial share of training costs after hiring align with the empirical evidence. A quantitative version of the model calibrated to the US reveals significant inefficiencies in training provision, particularly among smaller firms, and suggests that (1) optimal training subsidies are higher for smaller firms, though even a uniform subsidy can raise net output by 7%; and (2) increasing the labor market share of larger firms can significantly impact on-the-job human capital formation.

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

JEL codes: E24, J24, M53

*Email: xiaoma@phbs.pku.edu.cn, anakab@utdt.edu, and daniela.vidart@uconn.edu. 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, and PHBS Meeting on Theoretical and Empirical Macroeconomics.

1 Introduction

Numerous studies in labor economics have documented that on-the-job training has large and persistent impacts on workers' productivity.[1] Despite this, in his seminal work, Becker (1964) argues that firms have little incentive to invest in general training for their employees, as the skills acquired can be used in other firms. However, subsequent research has shown that firms may still be willing to invest in general training due to frictional labor markets, which allow them to capture some of the returns on this investment (Acemoglu and Pischke (1999)). Yet, the extent to which firms actually bear the costs of training compared to workers remains poorly understood because cost-sharing arrangements are difficult to observe. For example, it is unclear whether a low observed wage reflects a worker taking a pay cut to finance part of their training, or if other factors at the worker or firm level are driving it. Relatedly, it is unclear what characteristics influence the amount of training that firms are willing to provide and fund. In this paper, we empirically analyze how training patterns change with firm characteristics. We then build a model to interpret these findings and show that they offer key insights into how training costs are shared between firms and workers. With this, this paper aims to further our understanding of the roles that both workers and firms play in on-the-job learning and career advancement, and to inform the design of policies aimed at increasing investments in on-the-job training and improving worker productivity.

In the empirical portion of the paper, we explore how firm-level training investments vary with firm size, productivity, and labor shares. First, we construct a harmonized and consistent definition of training that encompasses any organized or sustained on-the-job learning activity occurring outside of the formal education system, and thus captures several important sources of workers' human capital acquisition such as participation in seminars or workshops, along with task-related learning arising from coworker instruction. Using data from over 100 countries, we then show that on-the-job training opportunities are consistently lower in smaller firms, and that this is robust to controlling for year, country, and industry fixed effects.

Using administrative firm-level data from China and Mexico, we then show that differences in the labor share and productivity levels across firms are key to explain the positive correlation between firm size and training. In particular, we find that firms with larger labor shares are less likely to offer training, while firms with higher TFP levels are more likely to do so. These patterns are robust to different TFP measures and to controlling for firm size, as well

[1]See Leuven and Oosterbeek (2004) and Bassanini et al. (2005) for reviews of this literature.

as year, industry, and firm fixed effects.

The positive correlation between TFP and training suggests that there are complementarities between firm-level productivities and workers' human capital, and is consistent with the production function being supermodular (Acemoglu and Pischke (1998) and Bagger et al. (2014)). The negative correlation between the labor share and training, on the other hand, indicates that the share of the revenue received by workers negatively impacts the training opportunities available to them by reducing firms' willingness to invest in their human capital. This suggests that learning-cost allocation schemes where workers decide the level of training, or where the division of value is less important to the level of training chosen, are not supported by the data.

We then develop an analytical model that formalizes this idea and sheds light on the mechanisms mediating our empirical findings by examining how the incentives for training change with firm characteristics. Our model explicitly considers the rich interactions between firms and workers in training investments, and accounts for the incentives faced by each. The model economy is characterized by labor market frictions and firm heterogeneity à la Burdett and Mortensen (1998). Firms differ in their labor productivity, and post vacancies and wages to meet workers by random search. Workers can be separated from firms for two reasons: an exogenous separation shock leading workers to unemployment, and job-to-job transitions as employed workers look for new job offers.

After matching, workers and firms jointly write a contract that stipulates training investments and the share of the training costs that each party will finance.[2] We consider four cost-sharing scenarios for these training costs, which are motivated by the classic question of how training costs are shared between firms and workers, and encompass established approaches in the literature. In the first case, firms bear all explicit training costs and fully determine training (Acemoglu and Pischke (1998)). In the second case, workers bear all explicit training costs and fully determine training (Ben-Porath (1967), Manuelli and Seshadri (2014)). In the third case, firms and workers choose training to maximize the joint match value, and the shares of the explicit training costs allocated to each party correspond to the shares of this value they each receive (Acemoglu (1997), Moen and Rosén (2004)). We label this case the joint internal efficiency case. In the fourth case, both workers and firms pay a constant share of the explicit training costs, and the level of training is determined by the

[2]Firms can offload some of the expected training costs by posting lower wages. We consider different cases for the allocation of explicit (or out-of-pocket) training costs borne after the match is formed and the human capital of the worker is revealed.

party with lower affordability, i.e. the party desiring the lower level of training (Ma et al. (2024)). In this last case, training is generally determined by the firm, since at all levels of productivity, the marginal returns to training are lower for firms than workers due to the existence of a hold-up problem arising due to the possibility of workers' leaving the firm after being trained.

We then characterize training outcomes in each of these scenarios, and show that only cases where firms pay a significant portion of explicit training costs are consistent with our empirical findings regarding the positive correlation between TFP and training, and the negative correlation between the labor share and training.[3] In particular, we first find that training rises unequivocally with firm productivity only in cases where the firm determines the training level. 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 workers will be less likely to be poached from more productive firms.[4] In addition, we also find that the negative correlation between training and the labor share only arises in cases where the firm determines the training level, since the labor share is inversely correlated with the returns the firm receives from the match.[5] Thus, the data does not support a cost-sharing structure where workers choose the training level, or following joint internal efficiency where the joint match value is maximized. Instead, the data supports a cost-sharing structure where firms choose the training level. This occurs when firms bear all or a high enough share of explicit training costs.[6]

[3]Although firm productivity and the labor share are linked in our model setup, our closed-form training expressions allow us to consider the roles of firm productivity and the labor share separately, which matches our empirical analysis.

[4]The cases where the worker chooses the training level and of joint internal efficiency may yield training levels that decrease with firm productivity. This follows because (1) workers' returns from training change slowly with the current firm's productivity since they also incorporate the benefits after leaving the firm; and (2) workers lose a significant portion of compensated time when training, and this opportunity cost is higher in more productive firms.

[5]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 match value. In the case of joint internal efficiency, a higher labor share also increases the optimal training level, as it reduces the incidence of job-to-job transitions which reduce the benefits firms' receive from training.

[6]We provide two further pieces of evidence supporting the importance of firms in deciding and paying for training investments. First, we show that the share of formally trained workers in each country-year in the EU-CVT decreases with job turnover rates. This is consistent with firms playing a key role in deciding and paying for training investments, since job turnover depresses the incentives for firms to provide training, but not for workers. Second, we show that a sizeable share of workers receive training even when not wanted, suggesting that firms are in charge of training decisions.

We then extend the analytical model for quantitative analysis, and calibrate it to the US economy under each of the four cost-sharing scenarios to assess which training cost-sharing scheme best fits the data. Consistent with the analytical model, we find that only when firms pay a significant share of explicit training costs by either financing all of these, or a high share of them, training levels will be higher in more productive firms. This matches key evidence in the literature showing that workers in more productive firms exhibit faster rates of skill acquisition (Engbom (2021), Arellano-Bover (2020), Arellano-Bover and Saltiel (2023)), and follows from the joint effects of productivity and the labor share documented in the analytical model in these settings, particularly since the labor share is lower in higher productivity firms. In addition, we find that the scenario where firms pay a constant share of explicit training costs generates the most reasonable training returns matching the literature. In the other three scenarios, the returns to training are either too low or too high to match these empirical findings. For example, when firms bear all explicit training costs, the returns to training must be exceedingly high in order to reconcile the model with the training time data. This contrasts with the scenario where firms pay a share of explicit training costs since the reduced cost burden allows for more reasonable training returns when matching the training time data.[7] This suggests that the scenario where workers and firms pay a given share of explicit training costs matches the data best.

We then quantify the size of training inefficiencies in this cost-sharing scenario, and examine the behavior of these training inefficiencies along the productivity distribution of firms. To do this, we characterize the training choices of a constrained social planner and compare them to those present in our calibrated economy.[8] Since firms fail to internalize the gains from training to workers and other employers following separation, large inefficiencies exist in the provision of training in the constant cost-share scenario given that firms fully determine the level of training investments in this case. We find that these inefficiencies are more marked in smaller firms, due to the higher labor share which reduces the direct benefits of training for firms, along with the larger likelihood of workers being poached by other firms, which aggravates the hold-up problem.

Motivated by these results, we then consider the scope of policies that subsidize training to correct these inefficiencies and promote aggregate human capital accumulation and output

[7] When workers bear all explicit training costs or when the joint match value is maximized, on the other hand, the required training returns are low relative to the data since workers enjoy all future wage returns from training.

[8] This constrained social planner chooses the optimal level of training for each firm taking the vacancy and wage distributions in the competitive equilibrium as given.

growth. We find that the optimal training subsidy rate is larger for smaller firms due to the more marked inefficiencies prevalent in these, but is still quite substantial in larger firms due to (1) the inefficiencies in the provision of training that still prevail among these firms; and (2) the need to curtail labor reallocation towards small unproductive firms arising from heavily subsidizing these enterprises. Nevertheless, we find that even a policy providing the same subsidy rate to all firms can generate an 7% increase in net output in the US, and that the current subsidy rates provided by US states are low relative to the optimal policy.

Finally, we examine the influence of labor market concentration on training dynamics within our quantitative model. We find that as employment more heavily concentrates in higher productivity firms, the average training level in the economy first increases and then decreases. This stems from two countervailing forces: higher concentration raises overall training since more productive firms exhibit higher training levels, but it also promotes greater wage compression reducing training incentives. These results suggest that an increase in the labor market share of larger firms stemming, for instance, from the rise of superstar firms as characterized by Autor et al. (2020), can have important repercussions to on-the-job human capital formation and worker productivity dynamics which crucially depend on training inefficiencies and wage dispersion along the productivity distribution of firms.

The paper is organized as follows. In Section 2 we present a literature review. In Section 3, we describe the data and empirical evidence. In Section 4 we present the analytical model and results. In Section 5 we present the quantitative model extensions and calibration. In Section 6, we use the quantitative model to quantify the size of training inefficiencies; consider the effects of policies that subsidize training; and characterize the scope of labor market concentration in shaping aggregate training investments. We conclude in Section 7.

2 Related literature

Through its focus on the role of firms in driving and funding on-the-job training, our paper is most closely related to the theoretical literature on general training investments first proposed by Becker (1964), and later developed by others (Acemoglu (1997), and Moen and Rosén (2004)). A fundamental problem highlighted in this literature is that firms may have no incentives to fund general training investments due to the portability of the skills acquired, which implies that general human capital gains are immediately priced into wages. Other work in this literature has shown that firms may be willing to invest in general training due to the existence of frictional labor markets which allow firms to extract partial rents from training (Acemoglu and Pischke, 1999). However, the share of training costs that firms actually sponsor relative to workers is poorly understood. We contribute to this literature by building a model with firm heterogeneity and frictional labor markets that considers different training cost-sharing schemes between workers and firms, and sheds light on which of these are consistent with the data and empirical facts.[9]

Through its focus on firm-level differences in training investments, our paper is also related to the literature that examines the factors driving firms' training decisions (Black et al. (1999) and Braga (2018)).[10] Studies in this literature have documented a positive correlation between firm size and training investments in the US (Barron et al. (1987), Frazis et al. (1995)), and the UK (Harris (1999)). We contribute to this literature in two distinct ways. First, we use data from over 100 countries to show that this positive correlation between firm size and training expenditures is also prevalent among low- and middle-income countries. Second, we use administrative firm-level data in China and Mexico to examine the role of productivity and the labor share in giving rise to this pattern.

Our focus on productivity is rooted in the theoretical literature suggesting that more productive firms have higher returns from training investments since human capital and firm productivity are complements (Acemoglu and Pischke (1998) and Bagger et al. (2014)). This complementarity is empirically validated through studies showing positive assortative matching patterns between employers and employees in the US (Barth et al. (2016), Abowd et al. (2018), Song et al. (2019)). The link between productivity and training is further motivated by the widely discussed hold-up problem (Acemoglu (1997), Acemoglu and Pischke (1998), Moen and Rosén (2004)), per which firms underinvest in training due to the possibility of workers leaving the firm after being trained. This problem is aggravated in low productivity

[9]Our paper also relates to the vast labor literature that examines 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), and Ma et al. (2024) for reviews on this evidence). This literature documents overwhelmingly positive effects of work-related training on wages and productivity. Some work in this area has focused on disentangling the impact of on-the-job training on wages from its impact on productivity (e.g, Dearden et al. (2006), Conti (2005), and Konings and Vanormelingen (2015)). These studies generally find that the productivity gain from firm-sponsored training is substantially higher compared to the wage gain, indicating both that on-the-job training is linked to human capital acquisition, and that firms have an incentive to pay for training investments.

[10]A related literature has explored 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), and Arellano-Bover and Saltiel (2023)). This literature has focused on showing that there is substantial heterogeneity in firms' promotion of human capital accumulation, and that this is an important determinant of lifecycle earning dynamics.

firms, since workers are more likely to leave these types of jobs. Our focus on the labor share, on the other hand, is rooted in the fact that this directly determines the revenue workers and firms receive from human capital acquisition, and thus shapes their optimal training decisions. Thus, our paper is also related to the literature that explores the consequences of labor market concentration on wages and related outcomes (Amodio et al. (2021), Azar et al. (2022), Berger et al. (2022)), and the literature documenting a decline in the aggregate labor share stemming from the rise in concentration among larger and more productive firms (Karabarbounis and Neiman (2014), Grullon et al. (2019), Barkai (2020), Autor et al. (2020), Gouin-Bonenfant (2022)). In a related paper, Amodio et al. (2021) use data from Peru to document that the average earnings and education of workers are lower in labor markets where concentration is higher, and build a model where rent-sharing between firms and workers depends on labor market power. By highlighting the importance of the labor share in training decisions, we show that the consequences of the rise in concentration and the consequent labor share decline extend to on-the-job human capital accumulation.

3 Data and empirical evidence on firm-level differences in training investments

We now turn our attention to analyzing how firm-level training investments differ with firm size, productivity, and labor shares. To do this, we use enterprise-level data from more than 100 countries from the World Bank and the European Union, along with detailed administrative firm-level data from China and Mexico. In this section, we first describe the data sources and carefully define on-the-job training, and then present our 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 first rely on firm-level data from more than 100 countries. We primarily rely on the World Bank Enterprise Survey (WB-ES) for this analysis, but also complement our findings with the European Union Continuing Vocational Training (EU-CVT) enterprise survey. These two sources jointly encompass developing and developed economies with per-capita GDP levels ranging from \$1,000 to \$60,000, thus suggesting that the patterns we document between training and firm size are not unique to particular settings. In addition, we also use cross-country worker-level data from the OECD Program for the International Assessment of Adult Competencies (PIAAC) to further support our findings.

The WB-ES is a collection of firm-level surveys of a representative sample of an economy's private manufacturing and service sectors covering approximately 136,000 firms across 140 low- and middle-income countries.[11] The ES consists of interviews with establishments' owners and top managers, who can request assistance of their firms' accountants or human resources managers to answer certain questions. Firms that are fully state-owned are omitted. In addition, firms with fewer than 5 employees are also usually omitted, though for some particular surveys in some countries, the ES includes informal firms and/or firms with fewer than 5 employees. We rely on the two ES waves conducted between 2002 and 2005 and between 2006 and 2017 since these have standardized questions on workers' training. For further details on this data please see Appendix A.1.

The EU-CVT provides information on European enterprises' investments in continuing vocational training for their staff, providing information on participation, time spent, and the costs of such training.[12] 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 and Norway. For further details on this data please see Appendix A.2.

PIAAC is an international survey conducted in over 40 OECD countries in 2011–2017.[13] The survey aims to assess and compare the learning environments, skills, and competencies of adults aged 16 to 65 in these countries. In total, this survey covers a sample of around 230,000 individuals. PIAAC collects information about workers' learning investments in skills, along with information on how adults utilize these skills in various settings. Further details on this data are available in Appendix A.3.

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

To investigate the drivers behind the relationship between firm size and training—specifically the roles of firm productivity and labor share—we rely on administrative data from Chinese

[11] Please see Table A.1 for a list of the countries and years covered by the WB-ES.

[12] Continuing vocational training refers to educational activities which are planned in advance, and organized with the specific goal of learning.

[13] Please see Appendix A.3 for a list of the countries covered by PIAAC.

and Mexican firms. The Chinese data corresponds to the Chinese Annual Survey of Manufacturing which contains detailed financial information from all manufacturing firms with revenues exceeding 5 million RMB. We rely on data from 2005–2007, which contain information on expenditures in training fees. The Mexican data corresponds to the Economic Census, which contains detailed financial information from all establishments in sectors outside of agriculture and forestry operating in a permanent location (thus excluding peddlers and vending carts, for example). We rely primarily on data from 2019 which contains information on the share of workers that received training in the past year. For both of these sources, we measure firm-level training investments, stock of capital, firm size, labor share, and different measures of productivity. For further details on these data sources please see Appendix A.4 and Appendix A.5, respectively.

3.2 Defining on-the-job training

Before turning our attention to the empirical evidence, we first carefully define training and its characteristics to ensure consistency across different data sources. Following Ma et al. (2024), we define *training* following the definition of "Non-formal Education and Training" category from ISCED (2011),[14] stating that training is any organized and structured learning activity outside the formal education system. Our definition encompasses two broad types of training: *formal training* and *informal training*. *Formal training* has a structured and defined curriculum and includes classroom work, seminars, and workshops, among others. Formal training activities are typically separate from the active workplace and show a high degree of organization by a trainer or an institution. Furthermore, this type of training is typically broader and not geared towards tasks, machinery, or equipment specific to certain jobs or workers. *Informal training* involves task-related learning connected to the active workplace and often arising from coworker instruction. It encompasses guided on-the-job training, job rotation, exchanges, and other forms of learning arising from participation in learning circles.[15]

Our definition of training includes all organized and structured on-the-job learning activities that take place outside the formal education system. This encompasses various important sources of human capital acquisition, such as participation in seminars or workshops, and learning from coworker instruction. However, our definition excludes formal schooling, as less

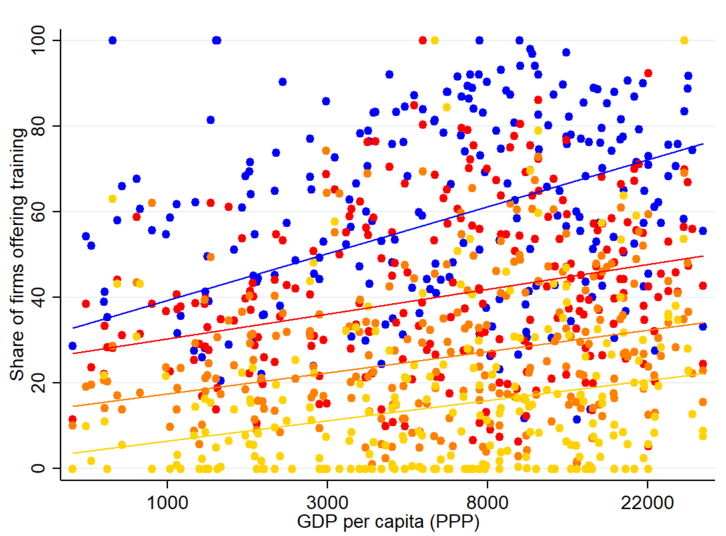
[14]The International Standard Classification of Education (ISCED) adopted by the UNESCO provides "uniform and internationally agreed definitions to facilitate comparisons of education systems across countries".
[15]For more information on these definitions, please see Ma et al. (2024).

than 10% of adult education involves formal schooling (Ma et al. (2024)). It also excludes informal learning activities like reading journals, visiting museums, or learning through media in an unstructured or unplanned manner, as these are primarily self-directed and typically do not involve employers.[16]

3.3 Training by firm size using cross-country data

We first show that the proportion of firms offering formal training increases with firm size. Figure 3.1 plots the share of firms providing formal training by firm size across countries with varying levels of GDP per capita, based on the WB-ES data. The plot indicates that larger firms are more likely to offer formal training regardless of a country's development level.[17] In Table B.1, we perform a firm-level regression where we regress a variable indicating whether the firm offers formal training on firm size, and show that this positive correlation between firm size and formal training is highly robust to including year, country, and industry fixed effects.[18]

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



Notes: Each dot represents the share of firms in a specific firm size category offering formal training in each country. The firm sizes 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.

[16]Our definition also excludes learning-by-doing, as we are focused on forms of learning that incur a cost and/or involve a tradeoff between learning and working.

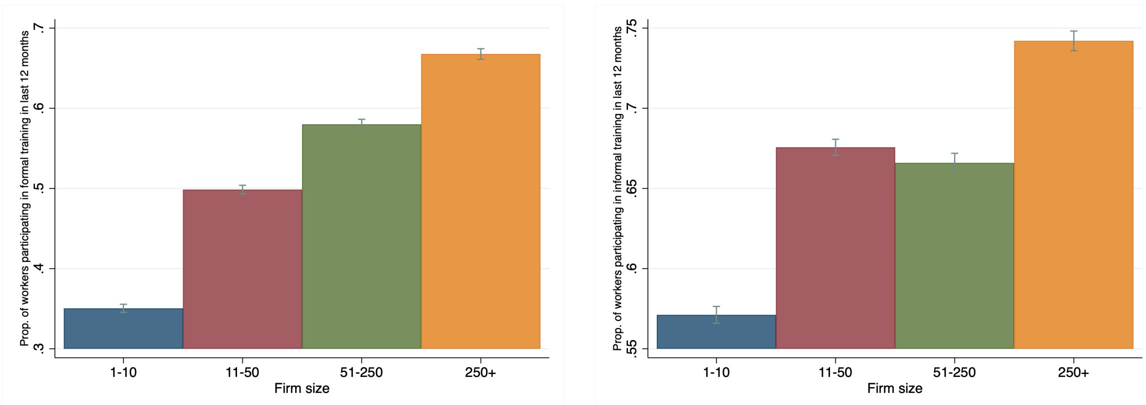
[17]Additionally, the plot shows that firms in countries with higher GDP per capita are more likely to offer formal training across all firm sizes, a finding considered in detail in Ma et al. (2024).

[18]These differences in training across firms of different sizes are also robust to considering different training types. Table B.2 presents the average shares of firms reporting formal training for different purposes (such as General IT, Management, Office Administration, etc) by firm size in the EU-CVT in 2010, and shows that larger firms report more training for virtually every purpose considered. In addition, we also find

In Figure B.3, we plot the difference in the share of firms offering training between medium and small firms, compared to the difference between large and medium firms. Panel (a) presents the data from the WB-ES, while panel (b) shows the EU-CVT data. In both datasets, 95% of country-year observations fall in the upper right quadrant, indicating that larger firms are much more likely to offer training than medium-sized firms, and medium-sized firms more than small firms.

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

(a) Formal training (b) Informal training



Notes: We plot 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.

In Figure 3.2, we use data from PIAAC to show that the positive correlation between firm size and on-the-job training is also prevalent when we look at worker-level data, implying that the results we observe are apparent from both workers' and firms' perspectives. This graph uses pooled data from all the countries considered and presents the share of workers reporting engaging in formal and informal training in the last 12 months, respectively, across different firm size bins. The results suggest similar patterns to the ones described above: workers in larger firms are more likely to engage in both formal and informal training. In Table B.5 we show that these results are robust to controlling for several demographic variables and wages, along with occupation, industry, and country fixed effects.[19]

that the positive correlation between firm size and training is also prevalent along the intensive margin: Figure B.1 shows that the hours spent in formal training courses by each participant increase with firm size using data from the EU-CVT. Finally, in Figure B.2, we show that the positive relationship between firm size and on-the-job training persists even when firms are divided into those above and below the median number of employees within each country-year, rather than using absolute firm size categories.

[19]In addition, in Figure B.4 we further show that these results extend to the intensive margin, as workers in larger firms spend more hours in formal training on average than workers in smaller firms. Please note that this intensive margin variable captures the number of hours workers spent in the most recent formal training activity, and not the hours spent in all training activities in the last 12 months.

3.4 Training by firm productivity and labor shares using administrative data

We now explore how firm productivity and the labor share contribute to the positive correlation between training and firm size we document. 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. These datasets include both financial and training information, allowing us to construct training, labor share, and TFP measures.

3.4.1 Construction of training, labor share, and TFP variables

In the Chinese data, training is measured by per-worker training expenditures, while in the Mexican data, training 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 TFP using various methods, which we detail below.

1. TFP à la Hsieh and Klenow (2009) We first calculate firm-level TFP using the approach outlined by Hsieh and Klenow (2009). This approach involves using firm-level data on revenue, payroll, and fixed capital stock, and then taking the residual of a constant-returns-to-scale Cobb-Douglas production function of capital and labor to retrieve TFP. Similar to Hsieh and Klenow (2009) and in order to control for human capital when estimating TFP, we measure the labor input using payroll, which captures both wage per unit of human capital and the number of units of human capital. We consider three different measures for the labor share in this Cobb-Douglas function: (1) labor share set at $^2/_3$, as suggested by cross-country evidence (Gollin, 2002) (HK1); (2) average industry-level labor shares following Hsieh and Klenow (2009) (HK2);[20] and (3) firm-level labor shares (HK3).

2. TFP via production function estimation à la Olley and Pakes (1996) and Levinsohn and Petrin (2003) We also calculate 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, cost of intermediate inputs, and fixed capital stock to address the simultaneity bias in productivity estimation, which occurs because

[20] Due to extensive labor market distortions in China, we follow Hsieh and Klenow (2009) and use the corresponding industry-level labor share from the US as a proxy for China's industry-level labor share.

productivity influences output both directly and indirectly by affecting input choices. Olley and Pakes (1996) tackle this issue by using investment as a proxy for productivity, while Levinsohn and Petrin (2003) instead use intermediate inputs as a proxy, addressing the fact that many plants have zero investment. These methods require panel data where each variable is deflated by its corresponding price index.[21] For example, we calculate real revenue for each firm using the three-digit industry-level producer price index, and real investment using the new capital expenditures price index. To control for human capital in the TFP estimation, we also include payroll in the estimation process. Finally, given differences in production technologies, we estimate these production functions separately for each 2-digit industry.

Our TFP measures aim to capture physical productivity (TFPQ) rather than revenue-based productivity (TFPR) given the distortions that lead to variations 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 using a constant elasticity of demand framework which separates a firm's prices and quantities based on a given elasticity of demand. Conversely, the methods of Olley and Pakes (1996) and Levinsohn and Petrin (2003) construct a traditional measure of TFP (TFPT) that falls between TFPQ and TFPR by using industry-level price deflators instead of plant-specific ones. As shown by Foster et al. (2008) using data from industries where physical output information is available, the correlation between TFPT and TFPQ is substantial.

Further details on the construction of the training, TFP, and labor share variables are provided in Appendix A.4 and Appendix A.5. In Figure A.1 and Figure A.2, we plot the histograms of all these measures, along with firm size, in both the Chinese and Mexican datasets, and show that there is substantial variation in them in both settings.

3.4.2 Specification and results

We examine the relationship between training, TFP, and the labor share using linear regressions applied to the Chinese and Mexican datasets, respectively,

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

[21] Although the only wave of the Mexican Economic Census with on-the-job training information is that of 2019, we also use data from 2009 and 2014 in order to estimate these TFP measures in the Mexican context.

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

$Y_{i,t}$ is an indicator for whether firm i offers training to its workers in period t.[22] Labor Share$_{i,t}$ and 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 age and industry fixed effects, with additional year and firm fixed effects in the Chinese regression.[23] However, since firms with different labor shares and productivity levels may also vary along other important dimensions, the estimated effects should be interpreted as correlations.

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	-0.151***	-0.027**	-0.188***	-0.033**	-0.204***	-0.106***	-0.038***	-0.11***	-0.11***
	(0.008)	(0.013)	(0.008)	(0.013)	(0.008)	(0.014)	(0.002)	(0.002)	(0.002)
TFP (HK1)	0.007***	0.010***					0.014***		
	(0.001)	(0.001)					(0.0003)		
TFP (HK2)			0.003***	0.009***				0.002***	
			(0.001)	(0.001)				(0.0002)	
TFP (HK3)					-0.003***	0.008***			0.001***
					(0.001)	(0.001)			(0.0002)
Log firm size	0.079***	0.038***	0.080***	0.037***	0.082***	0.035***	0.096***	0.104***	0.104***
	(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
Firm FE		Y		Y		Y			
Observations	772,764	658,597	768,476	654,831	773,885	660,187	1,561,690	1,558,774	1,561,672
R-squared	0.064	0.704	0.064	0.704	0.065	0.704	0.212	0.211	0.211

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.

In Tables 3.1 and 3.2, we present the results of these regressions when using the TFP measures constructed using the methodologies of Hsieh and Klenow (2009), and Olley and Pakes (1996) and Levinsohn and Petrin (2003), respectively. We find that in both China and Mexico, across almost all specifications, firms with higher labor shares are less likely to offer training,

[22] We focus on the extensive margin of training in the main text for 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.

[23] By incorporating firm fixed effects, we control for time-invariant firm characteristics and exploit within-firm variation in productivity and labor share over time for estimation.

while firms with higher TFP levels are more likely to do so. Specifically, a 10 percentage point increase in the labor share is associated with a 0.003–0.03 percentage point decline in the probability the firm offers training, and a 10% increase in TFP is associated with a 0.0001–0.004 percentage point increase in the probability the firm offers training.[24]

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***	-0.179***	-0.021	-0.10***	-0.10***
	(0.009)	(0.014)	(0.009)	(0.014)	(0.007)	(0.002)
TFP (OP)	-0.015***	0.006***			0.032***	
	(0.001)	(0.001)			(0.009)	
TFP (LP)			0.004***	0.013***		0.035***
			(0.001)	(0.001)		(0.002)
Log firm size	0.084***	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
Firm FE		Y		Y		
Observations	771,909	657,506	771,910	658,118	122,381	719,736
R-squared	0.065	0.704	0.062	0.703	0.311	0.199

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.

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 Section 3.3 is explained by these factors.[25] However, it is important to note that the positive

[24] We also explore the relationship between the labor share and training within the context of our cross-country data in Table B.3, where we regress an indicator of training provision on the labor share (measured as the ratio of total compensation to total sales), firm size, and additional controls using the WB-ES. This analysis indicates that a 10 percentage point increase in the labor share leads to approximately a 0.01 percentage point decrease in the likelihood of a firm offering training.

[25] The positive correlation between firm size and productivity has been documented in various settings (see, for example, Oi and Idson (1999), Hsieh and Klenow (2009, 2014), and Syverson (2011)). Similarly, the labor share has been shown to decrease with firm size in several countries and contexts (Karabarbounis and Neiman (2014), Grullon et al. (2019), Barkai (2020), Autor et al. (2020), Gouin-Bonenfant (2022)). We provide additional cross-country evidence for this latter pattern using the WB-ES. Table B.4 presents the results of regressing firm-specific labor shares (measured as the ratio of total compensation to total sales) on firm size in a pooled sample of firms from over 100 countries. Consistent with the existing literature, we find that as firm size increases, firms tend to have lower labor shares.

correlation between firm size and training remains positive and significant in all specifications of Tables 3.1 and 3.2, indicating that differences in productivity and labor shares do not fully explain this correlation, and that other factors are also at play.

The positive correlation between TFP and training suggests that firms' productivity and workers' human capital are complementary. Conversely, the negative correlation between the labor share and training indicates that as workers' share of revenue increases, firms become less willing to invest in their human capital, thereby reducing training opportunities. This finding suggests that models where workers decide the level of training or where the division of value is less critical to the level of training chosen are not supported by the data. We formalize this intuition in the analytical model presented in the next section.

4 Analytical model

We now construct an analytical model that sheds light on the mechanisms mediating our empirical findings and examines how different methods for allocating training costs between workers and firms influence training investments. This analysis also offers economic insights into the mechanisms driving the results in the quantitative model of Section 5.

The model features an economy with firm productivity heterogeneity and labor market frictions as in 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 in order to attract workers. This wage posting setup allows us to meaningfully consider different allocations of explicit (or out-of-pocket) training costs borne after matching, since although firms can always offload some of the expected training costs by posting lower wages, the question of who pays for costs after the match is formed remains. In addition, this setup allows the labor share to differ across firms, and thus allows us to study the differential impacts of the labor share and productivity on training decisions. After matching, workers and firms jointly write a contract that stipulates training investments and the share of the explicit training costs that each party will finance. In the second period of their lives, some workers engage in on-the-job search and switch firms if the new wage offer is higher than the current wage. By allowing for on-the-job search, we capture the fact that endogenous job transitions are a key driver of the lack of training investments (Acemoglu and Pischke, 1999). In what follows, we focus on the model's stationary equilibrium, in which firm-level distributions of workers' age and human capital levels remain constant through time, and firms with the same productivity level are symmetric.

4.1 Households

The model economy is populated by a continuum of workers, each living for two periods. All workers are born identical but accumulate human capital through training at potentially different rates.[26] Workers supply one unit of labor inelastically to the market in each period. Workers' utility is linear, and they aim to 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.[27] We treat the consumption good as the numeraire. In the steady state, $\rho = r$, and therefore workers are indifferent between consuming in each period. We normalize the population size of each generation to unity.

4.2 Production

There is a unit measure of firms that are heterogeneous in productivity $z \sim G(z)$ and produce a homogeneous good, which is used for consumption and paying training and vacancy costs. We consider human capital and firm productivity to be complements as commonly assumed in the literature (Acemoglu and Pischke, 1998; Bagger et al., 2014). Once workers and firms are matched, worker i's production in firm j is given by

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

where z_j is the firm-specific productivity level, and h_i is worker i's efficiency units of labor (human capital). By aggregating output across all workers within each firm and across all firms, we obtain total output:

$$Y = \int_j \int_{i \epsilon j} y_{i,j} \ di \ dj.$$

Since we focus on the stationary equilibrium of the model, in what follows we index firms and their choices using solely their productivity level.

[26] Workers may, in principle, differ in their initial skills, which could influence training choices. However, a caveat of our firm-level data is that we cannot observe labor composition within each firm. Given this data limitation, we opt to assume homogeneous workers in our baseline model. In Appendix C.5, we extend the quantitative model to include multiple education groups of workers and show that the results remain consistent with the baseline results.

[27] The wages w^Y and w^O for young and old ages entering the utility function are net of the training costs paid by the workers.

4.3 Job search and matching

Firms post vacancies $v(z)$ at the start of each period, with a contract stipulating the wage rate per efficiency unit $w(z)$ for all workers.[28] This wage rate determines the labor share in firm z, given by $\beta(z) = w(z)/z$.

The vacancy cost is specified as $c_v \frac{v(z)^{1+\gamma_v}}{1+\gamma_v}$, which is strictly convex (i.e., $\gamma_v > 0$), following Acemoglu and Hawkins (2014), to ensure the coexistence of firms with different productivity levels. The total number of vacancies is given by $V = \int v(z)\, dG(z)$, and the distribution of posted wage offers is $F(w) = \int_{w(z)<w} v(z) dG(z)/V$.

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—when workers become old—-there is a probability δ of exogenous separation. These separated workers enter the unemployment pool alongside 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 effective number of job seekers is thus given by $\tilde{U} = (1 + \eta(1 - \delta) + \delta)$.

For analytical tractability and since our focus is to characterize training investments under different cost-sharing schemes, in this analytical model we consider the matching function as $M(\tilde{U}, V) = \min\{\tilde{U}, V\}$, and assume c_v is small enough such that $V > \tilde{U}$, which ensures full employment in the equilibrium. As usual, market tightness is defined by $\theta = \frac{V}{\tilde{U}}$.

4.4 Training investments and cost-sharing scheme

A young worker has an initial human capital level of $h^Y = 1$ (normalization) and can be trained for s units of time to enjoy an increase in the next-period's human capital:

$$h^O = h^Y + \zeta s^{\gamma_s},$$

where ζ is a constant,[29] and $0 < \gamma_s < 1$ governs the diminishing returns of training. We assume that there are two components of training costs: (1) direct costs of $c_s \bar{w}$ per unit of training time, which are proportional to the average wage rate $\bar{w}$, and capture costs such

[28] Another strand of the literature models wages as the outcome of bargaining over the match surplus (Cahuc et al., 2006; Bagger et al., 2014). In such frameworks, training decisions typically maximize the joint value of the match, a mechanism we incorporate in the third training scenario discussed in Section 4.4. Empirical evidence suggests that both wage posting and bargaining practices are common (Hall and Krueger, 2012).

[29] In theory, ζ could vary across firms, reflecting heterogeneity in learning environments, as emphasized in recent work (Gregory, 2019; Arellano-Bover and Saltiel, 2021). However, Gregory (2019) shows that firm-provided training is a key driver of these environments, something our model explicitly captures by endogenizing variation in training intensity across firms.

as fees paid to trainers; and (2) the opportunity cost of training, with each unit of training time causing an equivalent decrease in production time.[30] We assume that training raises general human capital, so its benefits accrue even if the worker changes firms.[31]

4.4.1 Firms' and workers' benefits from training

The following proposition characterizes firms' and workers' benefits from human capital.

Proposition 1 (**Firms' and workers' gains from human capital increment**). *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*

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

and

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

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

Proof: See Appendix B.4.1.

Workers' gains from additional human capital depend on the expected wage flows if they stay in the firm or switch employers. These wage flows, in turn, depend on firms' productivity levels and on the share of match value captured by workers. Higher firm productivity of either current or future employers will incentivize workers to invest in training because their expected income per efficiency unit will increase. Moreover, if workers capture higher shares of the value through a higher labor share, they will also want more training as their expected income is also higher.[32]

[30] We assume that training costs are linear with respect to training time, as the costs of hiring trainers and the opportunity costs of lost production are likely proportional to the time spent on training.

[31] We focus on general human capital since the firm-specific components of human capital have been found to be much less important for wage growth than the general component (Altonji and Shakotko, 1987; Lazear, 2009; Kambourov and Manovskii, 2009), and given that the questions of training willingness by firms and importance of the hold-up problem only arise when training contributes to general human capital.

[32] In our wage posting framework, the labor share is identical across workers within the same firm. In contrast, a wage bargaining framework would generate heterogeneous labor shares within firms (e.g., Cahuc et al., 2006). Nevertheless, in such a setting, the worker's value would still increase with her labor share—

Firms' gains from additional human capital depend on the net revenues from the match with the worker. These net revenues increase with firm productivity and decrease with the share of match value that workers receive as wages (labor share). A key difference between workers and firms is that firms cannot reap the gains from training after the trained worker leaves. This creates a hold-up problem, per which firms underinvest in training due to the possibility of workers leaving the firm after being trained.

4.4.2 Allocation of training costs and optimal training levels

After matching, workers and firms jointly write a contract that stipulates training investments and the share of the training costs that each party will finance.[33] 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. Motivated by the classic question of how training costs are shared between firms and workers, and following established approaches in the literature, we consider four cases for allocating the explicit cost component, which in turn determines the level of training:

1. Firms bear all explicit training costs and fully determine training, which follows Acemoglu and Pischke (1998).

2. Workers bear all explicit 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 explicit training costs allocated to each party correspond to the shares of this value they each receive (Acemoglu (1997) and Moen and Rosén (2004)).

4. Firms and workers each pay constant shares of explicit training costs, denoted μ_F and μ_W, respectively, with $\mu_F + \mu_W = 1$. Given these cost shares, both parties independently determine their desired training levels. We assume that the actual training level is set by the party with the lower desired level—that is, the party with lower

reflecting greater bargaining power—while the firm's value would decline as its labor share rises.

[33] An alternative approach to modeling training decisions is to assume that firms and workers make fully independent training choices. However, as shown by Acemoglu and Pischke (1999), this setup is equivalent to either firms or workers fully bearing the training costs and making the training decisions, with the other party free-riding. This maps directly to two of the four cases discussed below.

affordability.[34]

Table 4.1 presents the optimal training levels under the different scenarios of training determination. We now describe each of the scenarios respectively.

Table 4.1: Optimal training levels in different scenarios

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

Training determined by firms. When the firm pays all explicit training costs, or when the firm pays a large enough share of the explicit 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

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

The first term on the right-hand side captures the human capital gains from training, and its costs. The second term reflects the potential for job turnover, which depresses firms' training incentives and corresponds to the classic hold-up problem. On-the-job search introduces differential worker attrition rates across firms, causing the severity of this hold-up to vary along the productivity distribution of firms.

The last two terms capture firms' returns from an additional unit of human capital in production. Higher firm productivity increases the revenue generated per unit of human capital,

[34] This assumption reflects the idea that training investments require mutual consent, even when costs are shared. For example, if the firm bears a large share of the cost, workers may prefer a six-month training course, but the firm may only be willing to support two months, a level workers are still willing to co-finance. Conversely, if workers bear a large share of the cost, they may forgo some training opportunities proposed by the firm, even though the firm may still be willing to contribute to the opportunities workers do pursue.

strengthening firms' incentives to invest in training. In contrast, a higher labor share raises the portion of match surplus paid as wages, reducing the firm's return on training and thus discouraging investment.

In our model, and since we use a wage posting setting, we allow the labor share to differ across firms, which generates lower labor shares in larger and more productive firms as observed in the data (Autor et al., 2020) and further shown using cross-country data in our empirical section. This further disincentivizes training in less productive firms, and thus suggests that smaller firms suffer from further underinvestments in training compared with larger firms.[35]

Training determined by workers. When the worker pays all explicit training costs, or when the worker pays a large enough share of the explicit 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

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

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

[35] To further illustrate how the labor share and productivity jointly impact training investments, in Figure C.3 we plot training investments for firms with different productivity levels, considering either decreasing labor shares from the calibrated model, or constant labor shares. This figure shows that in the case where the labor share is constant across firms, training increases in a linear fashion with respect to productivity. However, in the case where the labor share decreases with firm productivity, training will be considerably lower among more unproductive firms, and larger for the remaining firms.

$$s(z) = \left(\frac{\zeta\gamma_s(MR_W(z)+MR_F(z))}{(1+r)(c_s\bar{w}+z)}\right)^{\frac{1}{1-\gamma_s}}$$

$$= \Bigg[\underbrace{\frac{\zeta\gamma_s}{(1+r)(c_s\bar{w}+z)}}_{\text{returns \& costs from training}} \Bigg(\underbrace{(1-\delta)\left[1-\eta\bar{F}(w(z))\right]}_{\text{prop of keeping match}} \underbrace{z}_{\text{firm's+worker's share of value}} + \underbrace{(1-\delta)\eta\int \mathbf{1}_{\{w(u)>w(z)\}}\beta(u)uf(w(u))w'(u)du}_{\text{expected worker's share of value after J-J transitions}}$$

$$+ \underbrace{\delta\int \beta(u)uf(w(u))w'(u)du}_{\text{worker's share of value after exogenous separations}} \Bigg)\Bigg]^{\frac{1}{1-\gamma_s}}. \tag{5}$$

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. However, underinvestment in training still arises, since neither party internalizes the benefits that accrue to future employers when a trained worker leaves the firm.

4.4.3 Confronting cost-sharing scenarios with empirical evidence

Armed with these results, we now confront the model's training predictions in each scenario with our empirical findings. Although firm productivity and the labor share are linked in our framework,[36] we consider their effects separately, mirroring our empirical strategy, where we include both variables in the same regression to isolate their individual contributions. The effects of firm productivity (z) and the labor share ($\beta(z)$) on training under each cost-sharing scenario are formalized in Proposition 2 and discussed below.

Proposition 2 (**Determinants of training level**). *When firms determine training:*
(1) holding the labor share $\beta(z)$ constant, the training level $s(z)$ increases with productivity z; and
(2) holding productivity z constant, if on-the-job search intensity η is small enough, there is a negative correlation between the labor share $\beta(z)$ and the training level $s(z)$.

When workers determine training, or the joint match value is maximized,
(1) holding the labor share $\beta(z)$ constant, there is an ambiguous relationship between the training level $s(z)$ and productivity z; and
(2) holding productivity z constant, there is a positive relationship between the labor share $\beta(z)$ and the training level $s(z)$.

Proof: See Appendix B.4.2.

[36] Both firm productivity and the labor share are linked to wages, which are endogenous in our setup. As a result, their influence on training investments also reflects the underlying role of wages.

Effect of productivity on training. First, we find that when firms determine the training level, and holding the 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.

The cases where the worker chooses the training level and of joint internal efficiency, on the other hand, may yield training levels that 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.[37] 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 the labor share on training. In addition, we find that when the firm determines the training level, and holding the productivity level 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 and additional evidence These results suggest that only cases where firms pay a significant portion of explicit training costs are consistent with our empirical findings. In Appendix B.5, we provide further evidence supporting the importance of firms in deciding and paying for training investments. First, we regress the share of formally trained workers

[37]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.

in each country-year in the EU-CVT on the predicted probability of staying in the same firm after a quarter. We find that higher job turnover rates are associated with lower levels of training even after controlling for country income. This is consistent with firms playing a key role in deciding and paying for training investments, since job turnover depresses the incentives for firms to provide training, but not for workers. Second, we use worker-level data from the Adult Education Surveys conducted in the EU in 2011 and 2016 (EU-AES) to show that a sizeable share of workers employed at firms of all sizes receive training even when not wanted.[38] This further suggests that firms are in charge of training decisions.

4.5 Solving the firms' problem

In each period, given young workers' training $s(z)$ as discussed above, a firm with productivity z chooses the wage rate $w(z)$ and number of vacancies $v(z)$ to maximize the total value from hiring. This value can be written as:

$$\begin{aligned}\max_{\{w(z),v(z)\}} \quad & \underbrace{\frac{v(z)}{\theta}\frac{1}{1+\eta(1-\delta)+\delta}\left[z-w(z)-\mu_F(z)(c_s\bar{w}+z)s(z)+\frac{1}{1+r}MR_F(z)(1+\zeta s(z)^{\gamma_s})\right]}_{\text{profits from hiring young workers}} \\ & + \underbrace{\frac{v(z)}{\theta}\frac{\eta(1-\delta)+\delta}{1+\eta(1-\delta)+\delta}\frac{\eta(1-\delta)F(w(z))\bar{l}(w(z))+\delta\bar{l}}{\eta(1-\delta)+\delta}(z-w(z))}_{\text{profits from hiring old workers}} - \underbrace{\frac{c_v v(z)^{1+\gamma_v}}{1+\gamma_v}}_{\text{vacancy costs}} \\ \text{s.t.} \quad & w(z)\geq b\bar{w},\end{aligned} \tag{6}$$

The first term in this equation represents the net profits from hiring young workers, where $\frac{v(z)}{\theta}\frac{1}{1+\eta(1-\delta)+\delta}$ is the number of young workers met by the firm posting $v(z)$ vacancies, and $\left[z-w(z)-\mu_F(z)(c_s\bar{w}+z)s(z)+\frac{1}{1+r}MR_F(z)(1+\zeta s(z)^{\gamma_s})\right]$ is the sum of the current-period and the expected next-period profits from hiring a young worker. It is worth noting that these profits account for expected training costs, indicating that firms may offload some of the training costs via posting lower wage rates. Nevertheless, the explicit (or out-of-pocket) costs arising after the match is formed are divided according to the cases discussed above. The second term captures the profits from poaching old workers from other firms or hiring them from unemployment, where $\frac{v(z)}{\theta}\frac{\eta(1-\delta)+\delta}{1+\eta(1-\delta)+\delta}$ is the number of old workers met by the firm. Upon receiving a job offer, on-the-job movers have a probability $F(w(z))$ of moving to firm z and an average of $\bar{l}(w(z))$ efficiency units of labor (human capital). Unemployed old workers, on the other hand, have a probability of 1 accepting the offer and have an average of

[38] We provide further details about the EU-AES data in Appendix A.7.

$\bar{l}$ efficiency units of labor (human capital).[39] Finally, the third term in this equation captures total vacancy costs.

We solve $w(z)$ and $v(z)$ through the first-order conditions of equation (6). In particular, $w(z)$ is determined by a first-order differential equation, combined with the minimum wage $b\bar{w}$, as in Burdett and Mortensen (1998).[40] Intuitively, firms have incentives to increase wage offers to poach workers from other firms and to keep their own workers from being poached. Nevertheless, higher wages generate a higher labor share, which decreases profits. Thus, the wage distribution is determined by these two offsetting forces, with more productive firms paying higher wages since retaining workers is more profitable compared to the costs incurred. $v(z)$, on the other hand, is determined by balancing the gains and profits from vacancy posting. In particular, since hiring workers generates profits, firms want to post vacancies, but will stop posting eventually as the costs of additional vacancies increase.

4.6 Equilibrium

To close the model, we assume that firm owners spend the net profits (revenues net of wages, vacancy costs, and training costs borne by the firm) on consumption. We now define the model's general equilibrium in the steady state.

Definition 4.1. The general equilibrium for this economy is given by
(1) workers' decisions over consumption $\{c^Y, c^O\}$;
(2) firms' decisions over wages and vacancy posting $\{w(z), v(z)\}$;
(3) the decision of human capital accumulation $\{s(z)\}$; and
(4) offer distribution $F(w)$ and labor market tightness θ;
such that:
(i) given labor market tightness, offer distribution, and human capital accumulation, (1) solves the households' utility maximization problems;
(ii) given labor market tightness, offer distribution, and human capital accumulation, (2)

[39] The average number of efficiency units of labor for on-the-job movers is given by $\bar{l}(w) = 1 + \frac{\int_0^{w(z)} \zeta s(u)^{\gamma_s} dF(w(u))}{F(w(z))}$, whereas for unemployed workers this is given by $\bar{l} = 1 + \int \zeta s(z)^{\gamma_s} dF(w(z))$.

[40] As shown by Hornstein et al. (2011), search and matching models with reasonable unemployment benefits have difficulty in generating the amount of frictional wage dispersion present in the data. Thus, because of our focus on training decisions, we choose to match the frictional 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 that the unemployed will take any job offer, which can be rationalized by low, often negative, values of unemployment benefits. This assumption matches empirical findings of the offer acceptance rate being 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.

solves the firm's problem;
(iii) given offer distribution and labor market tightness, (3) solves the optimal training problem for firms and workers according to Table 4.1;
(iv) offer distribution $F(w)$ and labor market tightness θ are consistent with workers' job transitions and firms' wage and vacancy posting; and
(v) firms' total output equals the sum of consumption, vacancy costs, and training costs.

5 Quantitative model and calibration

In this section we extend the analytical model for quantitative analysis and take our model to the data. We calibrate the model to the US economy under each of the four cost-sharing scenarios to assess which training cost-sharing scheme best fits the data.

5.1 Setup

We extend our analytical model to more closely replicate key aspects of the labor market and economic environment. These extensions are described below.

Workers We consider that workers live for $J > 2$ periods and that human capital from 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 born with, which captures basic cognitive and physical skills.

Firms Firms' productivity follows a Pareto distribution, $G(z) = 1 - z^{-\kappa}$, consistent with empirical evidence (Axtell, 2001).

Labor market We use the widely employed matching function $M(\tilde{U}, V) = c_M \tilde{U}^{\psi} V^{1-\psi}$, which generates positive unemployment and delivers reasonable elasticities of the number of matches with respect to the number of searchers $\tilde{U}$ and vacancies V.

Conditions for simulations The optimality conditions for the quantitative model mirror the intuition from our analytical framework and are presented in Appendix C.1. There, we characterize the optimal training levels as functions of firm productivity and worker age. This appendix also details how firms set wages and post vacancies.

5.2 Calibration

We calibrate the model to the US economy using quarterly data. Some parameters are directly set based on values from the literature, while the remaining ones are calibrated to match key data moments. We consider the four training cost-sharing scenarios introduced earlier and calibrate the corresponding parameters, $\mu_F(z)$ and $\mu_W(z)$, accordingly.

5.2.1 Externally calibrated parameters

We draw some common parameters directly from the literature. These externally calibrated parameters are 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 US 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 US labor flow and vacancy data. Since we lack estimates for the US, 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 US.

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)

5.2.2 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. We perform this calibration separately for each of the four training cost-sharing scenarios: (1) firms bear all explicit training costs; (2) workers bear all explicit training costs; (3) costs are shared in proportion to the respective benefits firms and workers receive from training; and (4) costs are shared according to fixed shares, μ_F for firms and μ_W for workers.

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; and the exogenous separation rate, δ. In the fourth case with constant firms' and workers' shares of training costs, we additionally calibrate the share of training costs borne by the firm, μ_F.

To calibrate these parameters, we target the following moments of the US 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 firm 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, and the ratio of training costs to wage costs of training, as reported in the 1995 Survey of Employer Provided Training (US-SEPT).[41] Finally, in the fourth case with constant shares of explicit training costs, given that we have an additional parameter μ_F, we also target the percent wage growth at 40 years' experience for US workers, as estimated by Lagakos et al. (2018).

Calibration results In Table 5.2, we first present a comparison of the targeted model and data moments in each of the four cost-sharing scenarios. The model does fairly well in matching the targeted moments in all scenarios. The exception is the ratio of training intensity between firms of different sizes, which is too low when workers pay all explicit training costs and when training maximizes the joint match value. This inability to match the gradient of training levels with regard to firm size is because workers' marginal benefits

[41] 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.6.

of training grow more slowly with firm size relative to the marginal costs of training.[42]

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

Moments	Data	Model: Firms pay	Model: Workers pay	Model: Maximize match value	Model: Constant cost shares
Panel (a): Targeted moments					
Moments: labor market					
Unemployment rate (%)	6.5	6.6	6.5	6.4	6.3
Ratio of #Vacancies to #Unemployed	0.55	0.55	0.54	0.58	0.52
Pareto parameter of firm size distribution	1.06	1.05	1.12	1.03	1.02
Share of employed people remaining in the same firm after one quarter	0.88	0.89	0.88	0.88	0.89
Share of employed people remaining employed after one quarter	0.94	0.95	0.94	0.94	0.94
Moments: training intensity					
Average training intensity (% time)	2.20	2.17	2.21	2.13	2.19
Ratio of training costs to wage costs of training	0.24	0.24	0.24	0.23	0.25
Moments: training across firms					
Ratio of training intensity in firms with 100-499 employees to that with 50–99 employees	1.19	1.28	0.95	0.92	1.20
Percent wage increase of 40 years' experience (%)	89	-	-	-	89
Panel (b): Non-targeted moments					
Percent wage increase of 40 years' experience (%)	89	121	26	24	-

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

We report the calibrated parameters for each of the four scenarios in Table 5.3. Our parameters are generally 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 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 values, γ_s =0.21–0.44, are close to the estimates in the literature. For instance, Imai and

[42] We further reinforce this intuition in Appendix C.2 which characterizes the marginal returns to training for workers and firms of different productivity levels, and shows that only when firms pay a significant share of explicit training costs by financing either all or a high share of them, training levels will be higher in more productive firms. This matches key evidence in the literature showing that workers in more productive firms exhibit faster rates of skill acquisition (Engbom (2021), Arellano-Bover (2020), Arellano-Bover and Saltiel (2023)), and follows from the joint effects of productivity and the labor share documented in the analytical model, particularly since the labor share is lower in higher productivity firms.

Keane (2004) find this parameter to be 0.22, while Manuelli and Seshadri (2014) estimate this parameter to be 0.48.

Table 5.3: **Calibrated parameters**

Parameter	Firms pay	Workers pay	Maximize match value	Constant cost shares
c_M - Constant in matching function	0.78	0.87	0.82	0.85
η - On-the-job search intensity	0.22	0.32	0.23	0.27
c_s - Ratio of training costs per time to wage	0.28	0.24	0.22	0.28
c_v - Constant in vacancy function	0.68	0.27	0.31	0.52
ζ - Constant in training function	0.20	0.02	0.02	0.06
γ_s - Convexity of training function	0.44	0.28	0.31	0.21
κ - Parameter of Pareto productivity dist	5.14	9.69	4.92	5.96
δ - Exogenous separation rate	0.07	0.07	0.07	0.07
$\mu_F(z)$ - Share of training costs paid by firm	1	0	Value share	0.30

We now turn our attention to training returns, captured by ζs^{γ_s}. In the fourth scenario, where the share of explicit training costs borne by firms is constant and calibrated, training a young worker for the full quarter (480 working hours, or $s = 1$) increases hourly wages by 6% (captured by ζ). This lies within the range of empirical evidence on US training returns as reviewed by Leuven (2004) and Bassanini et al. (2005). For example, Frazis and Loewenstein (2005) find that 60 hours of formal training increases the wage by 3–5%, matching our calibration which implies 4% wage growth for 60 hours of training in a quarter. In the other three scenarios, the returns to training are either too low or too high to match these empirical findings, with a full quarter of training leading to wage growth of 2% (workers pay all explicit training costs, or match value is maximized) and 20% (firms pay all explicit training costs), respectively. When firms bear all explicit training costs, the returns to training must be exceedingly high in order to reconcile the model with the training time data. This contrasts with the scenario where firms pay a share of explicit training costs since the reduced cost burden allows for more reasonable training returns when matching the training time data. When workers bear all explicit training costs or when the joint match value is maximized, on the other hand, the required training returns are low relative to the data since workers enjoy all future wage returns from training.

Echoing these findings, in Panel (b) of Table 5.2, we also show that the percent wage increase from 40 years of experience is either too low or too high in these three scenarios compared with the data estimates. This suggests that the scenario where workers and firms pay a calibrated constant share of explicit training costs matches the data best.

6 Training inefficiencies, subsidies, and labor market concentration

Our analysis so far suggests that the constant cost-share scenario provides the best fit to our empirical evidence and specific targeted and non-targeted moments. Nevertheless, since this scenario allows firms to fully control the level of training investments, it also implies significant inefficiencies in the provision of training since firms face lower incentives for training compared with the social optimum. Specifically, firms fail to internalize the benefits of training to workers and other employers following separation.

This section first assesses the extent of training inefficiencies in our preferred calibrated cost-sharing scenario in the quantitative model and examines the behavior of these training inefficiencies along the productivity distribution of firms. To do this, we characterize the training choices of a constrained social planner and compare them to those present in our calibrated economy. Then, we examine the scope of different policies that subsidize training to correct these inefficiencies and promote aggregate human capital accumulation and output gains; and assess the scope of labor market concentration in shaping aggregate training investments.

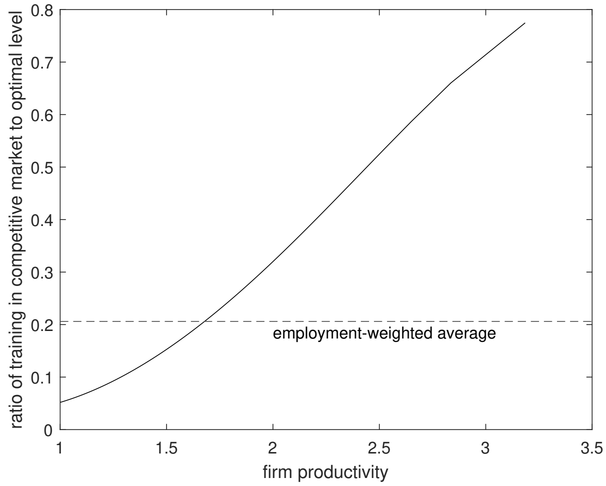
6.1 Social planner's problem and training inefficiencies

To quantify the extent of training inefficiencies within our quantitative model, we first consider the social planner's problem. We restrict the social planner to selecting the optimal training level for each firm, conditional on the vacancy and wage distributions in the competitive equilibrium. This allows us to focus on inefficiencies related to training decisions, rather than those arising from the well-known inefficiencies caused by frictional labor markets, which prevent workers from being allocated to more productive firms.[43]

Figure 6.1 presents the ratio of the training levels prevalent in the competitive equilibrium to those chosen by the social planner. This figure shows that inefficiencies are substantial across the entire productivity distribution of firms, with the employment-weighted average training

[43] In practice, we numerically compute the social marginal revenue of training for each worker in each firm within the competitive equilibrium by adding up the discounted future productivity flows for each extra unit of human capital. Given that marginal costs remain unchanged from the competitive equilibrium, we can then determine the social planner's optimal training decisions.

Figure 6.1: Social Planner and Competitive Equilibrium



intensity in the competitive equilibrium being 21% of that chosen by the social planner. Nevertheless, unproductive firms tend to provide significantly lower levels of training in the competitive equilibrium relative to the planner's problem than more productive firms. This is driven by the higher labor shares prevalent in more unproductive firms, which reduce the direct benefits of training for firms, along with the larger likelihood of workers leaving, which aggravates the hold-up problem.

In Figure C.4 we examine the relative importance of these labor share and hold-up mechanisms in driving training inefficiencies by assuming that all firms have the same labor share as the most productive firm and recomputing the optimal training levels prevalent in the competitive equilibrium. We find that the inefficiencies in the provision of training by unproductive firms decrease but still remain considerable, suggesting that the increased likelihood of workers leaving is the primary driver of the inefficiency of training provision in less productive firms. Overall, these pieces of evidence suggest that training subsidies should be more heavily targeted towards less productive firms.

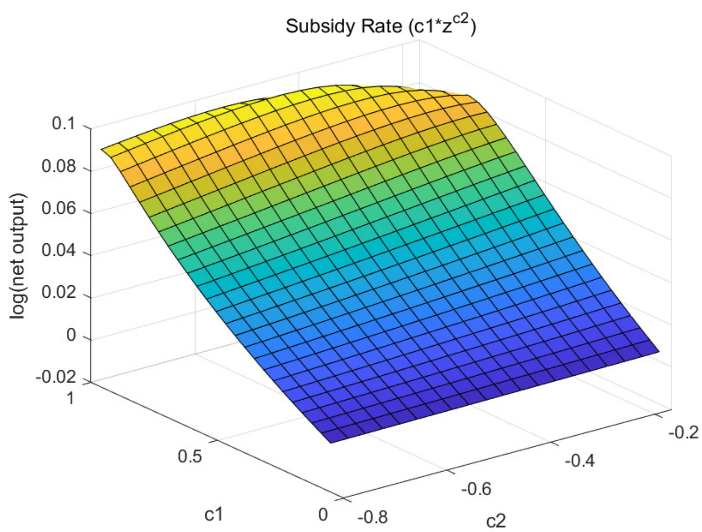
6.2 Training subsidies

In light of the training inefficiencies highlighted above, and having characterized the constrained planner's choices, we now examine the scope of policies that subsidize training to correct these inefficiencies and promote human capital and output gains. We consider three types of policies. First, given that training inefficiencies vary across firms, we consider a policy that allows subsidy rates to differ across firms of different productivity levels. Second, and given that policies cannot generally target subsidies to firms of every specific size or productivity level, we allow subsidy rates to vary across firms in different firm size brackets. Finally, we consider a policy that assumes the same subsidy rate for all firms, and is thus the most realistic and in line with training policies in place today.[44]

6.2.1 Policy targeting firms of different productivity levels

Given that training inefficiencies vary across firms, we first consider a policy that allows training subsidy rates to differ across firms of different productivity levels. For computational purposes, we define the subsidy for a firm with productivity z as $s(z) = c_1 z^{c_2}$. Figure 6.2 shows the output gains net of training and vacancy costs given the choices of c_1 and c_2 in the model. We normalize the log of net output in the baseline model to zero for ease of comparison.

Figure 6.2: Net Output Gain as Function of Subsidy Parameters



The optimal scenario generates a 10% increase in net output for the U.S., with $c_1 = 0.92$ and $c_2 = -0.5$. These estimates suggest that training subsidies should be more generous for smaller firms, which face greater inefficiencies in training provision. However, it is also intuitive to provide substantial subsidies to large firms. First, training inefficiencies persist even among these firms, as they cannot fully retain trained workers. Second, subsidizing

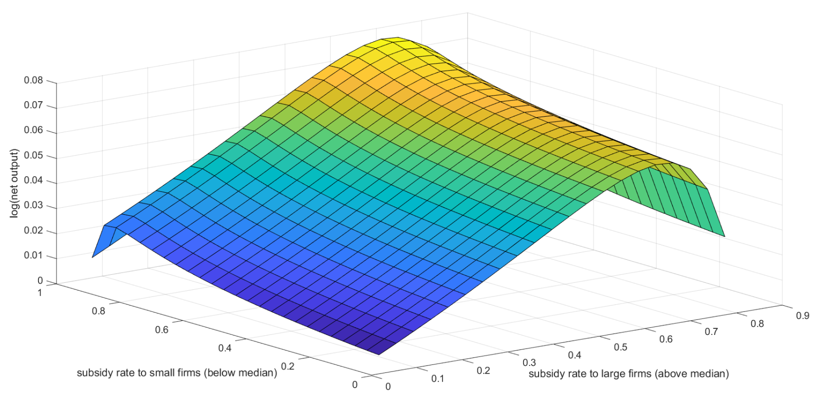
[44] While we condition on the vacancy and wage distributions of the competitive equilibrium when considering the social planner's choice and evaluating the training inefficiency shown in Figure 6.1, we allow these distributions to adjust endogenously when assessing subsidy policies, as training subsidies will prompt firms to modify their vacancy and wage posting decisions.

large firms can help limit inefficient labor reallocation toward small, unproductive firms that might result from heavily subsidizing only the latter.

6.2.2 Policy targeting firms in different size brackets

In practice, it is difficult for policymakers to tailor subsidies to each firm based on its exact size or productivity level. We therefore consider a more realistic scenario in which subsidy rates vary across firm size brackets. Specifically, we simulate a policy that assigns different training subsidy rates to firms above and below the median size. Figure 6.3 shows the resulting net output gains across different combinations of subsidy rates.

Figure 6.3: Net Output Gain from Subsidizing Small and Big Firms



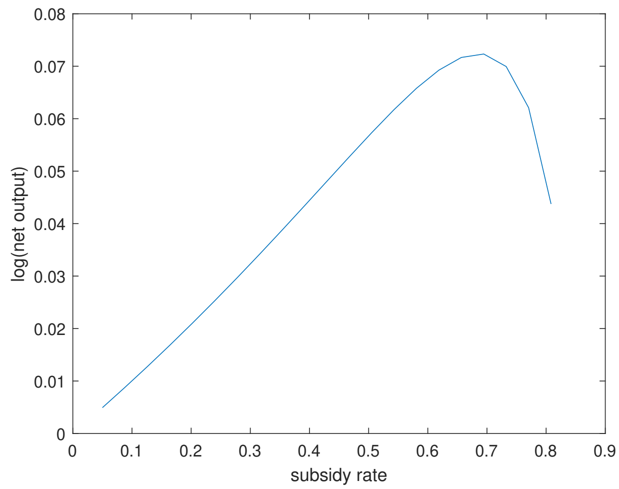
Starting from low subsidy rates for both firm size brackets, the figure shows that moderately increasing subsidies for both small and large firms raises net output. However, beyond a certain point, and especially when subsidies for large firms become too generous, net output begins to decline. This reflects diminishing returns to training at high investment levels. The optimal policy yields an 8% increase in net output for the US and implies a training subsidy rate of 88% for small firms and 65% for large firms.

6.2.3 Policy targeting all firms equally

We now consider a policy that provides a uniform subsidy rate to all firms, which is both the most realistic and most closely aligned with existing training programs. Figure 6.4 shows the net output gains across a range of uniform subsidy rates in the model.

The optimal uniform subsidy scenario generates a 7% increase in net output in the US, with a subsidy rate of 69%. This suggests that even without targeting firms by size or productivity, training subsidies can yield substantial returns and correct much of the inefficiency in training provision. Under this optimal rate, the employment-weighted average training intensity in the competitive equilibrium reaches 90% of the level chosen by the social planner, an increase of 69 percentage points relative to the scenario with no subsidies.

Figure 6.4: Net Output Gain as Function of Subsidy Parameters



In Table C.1, we review training subsidy policies across 21 US states. We find that the median policy reimburses roughly 50% of training costs. While this suggests that the US recognizes training as a key driver of productivity and acknowledges potential underinvestment by firms, current subsidy rates remain below the optimal rate of 69%.[45] Nonetheless, this gap is relatively modest, indicating that achieving the optimal subsidy rate may be feasible within the current policy framework.

6.3 Changes in labor market concentration

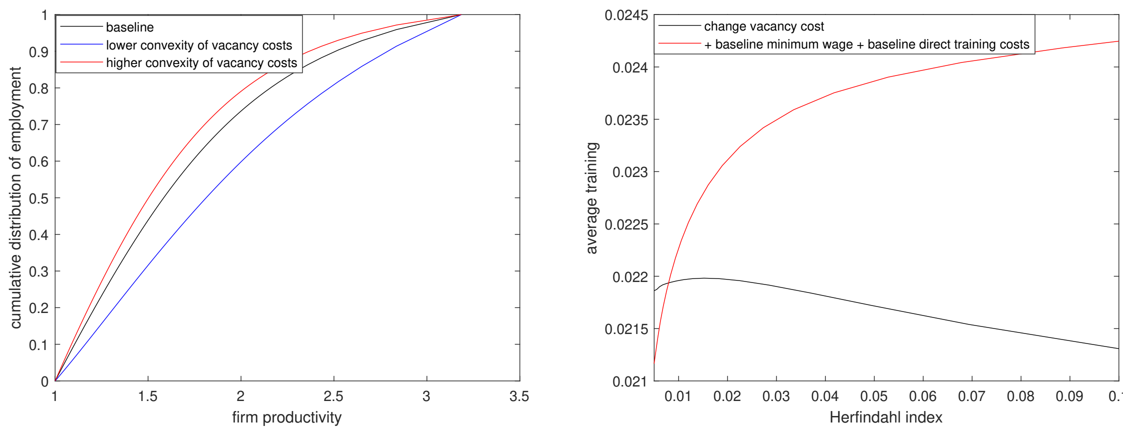
Finally, we examine the influence of labor market concentration on training dynamics within our quantitative model. To achieve this, we introduce a shock to the cost associated with posting vacancies by adjusting the convexity parameter (γ_v). Specifically, a decrease in this parameter results in enhanced employment opportunities within larger firms, while smaller firms experience a decline in employment. Panel (a) of Figure 6.5 displays the cumulative distribution of employment in our calibrated economy, showcasing the relationship between firm productivity and employment given an increase or decrease in the convexity parameter.

[45] Moreover, given the earlier findings that smaller firms face greater training inefficiencies, these firms may be particularly underserved by existing subsidy policies.

As anticipated, reducing the convexity of vacancy costs leads to a higher proportion of workers being employed by more productive firms. This, in turn, contributes to an expansion in the labor market share of these more productive firms, resulting in greater labor market concentration overall.

Figure 6.5: Concentration and training

(a) Cumulative distribution of employment (b) Average training



Panel (b) of Figure 6.5 illustrates the relationship between labor market concentration, measured by the Herfindahl index, and average training levels in the economy.[46] The black line shows that as the Herfindahl index rises, reflecting changes in the convexity of vacancy costs and indicating greater employment concentration in higher-productivity firms, the average training level first increases and then decreases. This non-monotonic pattern results from two opposing forces. First, more productive firms tend to offer higher training levels, so shifting employment toward these firms raises the economy-wide average training level. However, this reallocation also raises average wages, which increases both the minimum wage and the direct costs of training, since training costs are proportional to average wages. This reduces firms' incentives to invest in training. To isolate the role of wage compression, the red line in panel (b) plots the same relationship while holding the minimum wage and direct training costs fixed at their baseline equilibrium values. In this case, only the first effect operates, and we observe a monotonic increase in average training levels as the Herfindahl index rises.

These results suggest that an increase in the labor market share of larger firms stemming, for instance, from the rise of superstar firms as characterized by Autor et al. (2020), can have

[46] Since the Herfindahl index depends on the number of firms, we compute it consistently based on 1,000 firms across all scenarios.

important implications for on-the-job human capital formation and the dynamics of worker productivity, since these are shaped by training inefficiencies and wage dispersion across the firm productivity distribution.

7 Conclusions

In this paper, we examine how training patterns vary with firm characteristics and how these patterns relate to the distribution of training costs between firms and workers. Using data from over 100 countries, we show that smaller firms consistently offer fewer on-the-job training opportunities to their workers. Drawing on administrative firm-level data from China and Mexico, we further show that differences in labor share and productivity levels across firms are key to explaining this pattern.

We build a general equilibrium model with firm heterogeneity and training expenditures to shed light on these findings and explore four alternative training cost-sharing between firms and workers. Our analytical results suggest that only the scenarios in which firms bear a significant share of explicit training costs are consistent with the empirical patterns we observe.

We then consider a quantitative calibrated version of the model and show that a scenario in which firms pay a constant, calibrated share of explicit training costs yields training returns consistent with empirical estimates in the literature. Within this framework, we document substantial inefficiencies in training provision, particularly among smaller firms, due to higher worker turnover, which amplifies the hold-up problem. In light of these findings, we conduct two policy exercises. First, we show that the optimal training subsidy rate is higher for smaller firms, although even a uniform subsidy can increase net output by 7% in the U.S. Second, we show that an increase in the labor market share of larger firms can significantly impact on-the-job human capital formation.

Our findings have important implications for understanding the determinants of on-the-job learning and career advancement. First, we underscore the central role firms play in shaping workers' human capital and productivity trajectories, highlighting the need to integrate firm behavior into both data collection efforts and models of human capital formation. Second, our results provide direct evidence of a hold-up problem in training decisions, which may help explain why less fluid labor markets, such as those in Europe, exhibit significantly higher rates of on-the-job training compared to the US. Finally, our findings suggest that policies aiming to expand access to training and improve worker productivity should focus on targeting firms and addressing the constraints that limit their ability and willingness to invest in training. Future research could build on this by examining how the effectiveness of training subsidies varies across firms with different characteristics, 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. 7

Acemoglu, D. (1997): "Training and Innovation in an Imperfect Labour Market," *The Review of Economic Studies*, 64, 445, iSBN: 00346527. 3, 6, 7, 21

Acemoglu, D. and W. Hawkins (2014): "Search with Multi-worker Firms," *Theoretical Economics*, 9, 583–628. 19, 29

Acemoglu, D. and J.-S. Pischke (1998): "Why Do Firms Train?" *The Quarterly Journal of Economics*, 113, 79–119. 3, 7, 18, 21

——— (1999): "Beyond Becker: Training in Imperfect Labour Markets," *The Economic Journal*, 109, 112–142. 2, 7, 17, 21

Altonji, J. G. and R. A. Shakotko (1987): "Do Wages Rise with Job Seniority?" *Review of Economic Studies*, 54, 437–459. 20

Amodio, F., P. Medina, and M. Morlacco (2021): "Labor Market Power, Self-employment, and Development," *PEDL Research Papers*. 8

Arellano-Bover, J. (2020): "Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size," . 5, 7, 31, 73

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, 7, 31, 73

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. 6, 8, 16, 23, 38

Axtell, R. L. (2001): "Zipf Distribution of U.S. Firm Sizes," *Science*, 293, 1818–1820. 28, 30

Azar, J., S. T. Berry, and I. E. Marinescu (2022): "Estimating Labor Market Power," *NBER Working Paper*. 8

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. 3, 7, 18, 19, 68

Barkai, S. (2020): "Declining Labor and Capital Shares," *The Journal of Finance*, 75, 2421–2463. 8, 16

Barro, R. and J.-W. Lee (2013): "A New Data Set of Educational Attainment in the World, 1950-2010," *Journal of Development Economics*, 104. 76

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. 7

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. 7

Bassanini, A., A. Booth, G. Brunello, M. De Paola, and E. Leuven (2005): "Workplace Training in Europe," *IZA Discussion Paper Series*. 2, 7, 32, 77

Becker, G. S. (1964): *Human Capital*, Chicago: University of Chicago Press. 2, 6

Ben-Porath, Y. (1967): "The Production of Human Capital Over the Life Cycle," *Journal of Political Economy*, 75, 352–365. 3, 21

Berger, D. W., K. Herkenhoff, and S. Mongey (2022): "Labor Market Power," *American Economic Review*, 112, 1147–1193. 8

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. 7

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. 28, 29

Braga, B. (2018): "Earnings Dynamics: The Role of Education throughout a Worker's Career," *Labour Economics*, 52, 83–97, publisher: Elsevier B.V. 7

Burdett, K. and D. T. Mortensen (1998): "Wage Differentials, Employer Size, and Unemployment," *International Economic Review*, 39, 257–273. 3, 17, 27, 72

Cahuc, P., F. Postel-Vinay, and J. M. Robin (2006): "Wage Bargaining with On-the-job Search: Theory and Evidence," *Econometrica*, 74, 323–364. 19, 20

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. 7

Caselli, F. (2005): "Accounting for cross-country income differences," in *Handbook of Economic Growth*, ed. by P. Aghion and S. Durlauf, Elsevier, 679–741, arXiv: 1011.1669v3. 77

Conti, G. (2005): "Training, Productivity and Wages in Italy," *Labour Economics*, 12, 557–576. 7

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. 7

Donovan, K., W. J. Lu, and T. Schoellman (2023): "Labor Market Dynamics and Development," *The Quarterly Journal of Economics*, 138, 2287–2325. 30, 66, 67

Engbom, N. (2021): "Labor Market Fluidity and Human Capital Accumulation," *Manuscript*. 5, 7, 31, 73

Engbom, N., C. Moser, and J. Sauermann (2022): "Firm Pay Dynamics," *NBER Working Paper Series*. 7

Foster, L., J. Haltiwanger, and C. Syverson (2008): "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability," *American Economic Review*, 98, 394–425. 14

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. 32, 77

Frazis, H. J., D. E. Herz, and M. W. Horrigan (1995): "Employer-provided Training: Results from a New Survey," *Monthly Lab. Rev.*, 118, 3. 7

Friedrich, B., L. Laun, C. Meghir, and L. Pistaferri (2021): "Earnings Dynamics and Firm-Level Shocks," *NBER Working Paper Series*. 7

Gollin, D. (2002): "Getting Income Shares Right," *Journal of Political Economy*, 110, 458–474, iSBN: 00223808. 13, 51, 54

Gouin-Bonenfant, E. (2022): "Productivity Dispersion, Between-Firm Competition, and the Labor Share," *Econometrica*, Forthcoming. 8, 16

Gregory, V. (2019): "Firms as Learning Environments: Implications for Earnings Dynamics and Job Search," *Manuscript*, 1–57. 7, 19

Grullon, G., Y. Larkin, and R. Michaely (2019): "Are US Industries Becoming More Concentrated?" *Review of Finance*, 23, 697–743. 8, 16

Hall, R. E. and A. B. Krueger (2012): "Evidence on the Incidence of Wage Posting, Wage Bargaining, and On-the-Job Search," *American Economic Journal: Macroeconomics*, 4, 56–67. 19

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. 7

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. 7

Hornstein, A., P. Krusell, and G. L. Violante (2011): "Frictional Wage Dispersion in Search Models: A Quantitative Assessment," *American Economic Review*, 101, 2873–98. 27, 29

Hsieh, C. and P. Klenow (2014): "The Life Cycle of Plants in India and Mexico," *The Quarterly Journal of Economics*, 129, 1035–1084. 16

Hsieh, C.-T. and P. J. Klenow (2009): "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 124. 13, 14, 15, 16, 51, 54, 55, 63

Imai, S. and M. Keane (2004): "Intertemporal Labor Supply and Human Capital Accumulation," *International Economic Review*, 45, 601–641. 31

Islam, A., R. Jedwab, P. Romer, and D. Pereira (2019): "Returns to Experience and the Misallocation of Labor," *Working Paper*. 76

Jarosch, G. (2022): "Searching for Job Security and the Consequences of Job Loss," *NBER Working Paper Series*. 7

Kambourov, G. and L. Manovskii (2009): "Occupational Specificity of Human Capital," *International Economic Review*, 50, 63–115. 20

Karabarbounis, L. and B. Neiman (2014): "The Global Decline of the Labor Share," *The Quarterly Journal of Economics*, 129, 61–103. 8, 16

Kluve, J. (2010): "The Effectiveness of European Active Labor Market Programs," *Labour Economics*, 17, 904–918, publisher: Elsevier B.V. 7

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. 7

Lagakos, D., B. Moll, T. Porzio, N. Qian, and T. Schoellman (2018): "Life Cycle Wage Growth across Countries," *Journal of Political Economy*, 126. 30

Lazear, E. P. (2009): "Firm-Specific Human Capital: A Skill-Weights Approach," *Journal of Political Economy*, 117, 699–726. 20

Leuven, E. (2004): "A Review of the Wage Returns to Private Sector Training," *Working Paper*. 32, 77

Leuven, E. and H. Oosterbeek (2004): "Evaluating the Effect of Tax Deductions on Training," *Journal of Labor Economics*, 22, 461–488. 2, 7

Levinsohn, J. and A. Petrin (2003): "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 70, 317–341. 13, 14, 15, 16, 51, 54, 55, 64

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. 4, 7, 10, 11

Manuelli, R. E. and A. Seshadri (2014): "Human Capital and the Wealth of Nations," *American Economic Review*, 104, 2736–2762. 3, 21, 32

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. 7

Mincer, J. (1989): "Human Capital and the Labor Market: A Review of Current Research," *Educational Researcher*, 18. 28

Moen, E. and A. Rosén (2004): "Does Poaching Distort Training?" *The Review of Economic Studies*, 71, 1143–1162. 3, 6, 7, 21

Oi, W. Y. and T. L. Idson (1999): "Firm Size and Wages," *Handbook of Labor Economics*, 3rd Part, 2165–2214. 16

Olley, G. S. and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–1297. 13, 14, 15, 16, 51, 54, 55, 64

Shimer, R. (2005): "The Cyclical Behavior of Equilibrium Unemployment and Vacancies," *American Economic Review*, 95, 25–49. 29

——— (2012): "Reassessing the Ins and Outs of Unemployment," *Review of Economic Dynamics*, 15. 31

Song, J., D. Price, F. Guvenen, N. Bloom, and T. V. Wachter (2019): "Firming Up Inequality," *Quarterly Journal of Economics*. 7

Syverson, C. (2011): "What Determines Productivity?" *Journal of Economic Literature*, 49, 326–365. 16

van den Berg, G. J. (1990): "Search Behaviour, Transitions to Non-participation and the Duration of Unemployment," *The Economic Journal*, 100, 842–865. 27

What Works - Centre for Local Economic Growth (2016): "Evidence Review 1: Employment Training," Tech. rep., London School of Economics. 7

Online Appendix

A Data sources

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

The WB-ES is a collection of firm-level surveys of a representative sample of an economy's private manufacturing and service sectors[47] spanning 140 low- and middle- income countries. The survey is conducted by private contractors via face-to-face interviews on behalf of the World Bank. Typically the ES conducts 1200-1800 interviews in large economies, 360 in medium-sized economies, and 150 in small ones. The ES uses a stratified random sampling method per which firms are grouped according to firm size, business sector, and geographic region, and random sampling within those groups is representative of each stratum.[48] Since the majority of firms are small and medium-sized in the majority of economies considered, the WB-ES oversamples large firms.

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. This 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 operation procedures, or employee orientation at the beginning of an 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."

We capture enterprises that provide formal and informal training, along with the labor share, and firm size in the following way:

- Establishments offering formal training: Enterprises that had formal training programs for its permanent full-time employees over the last completed fiscal year.
- Labor share: Ratio between total compensation of employees and total sales of the enterprise.
- Firm size: Total number of employees in the enterprise.

[47] ISIC codes 15-37, 45, 50-52, 55, 60-64, and 72, ISIC Rev.3.1.

[48] The firm sizes considered are 5-19 (small), 20-99 (medium), and 100+ employees (large-sized firms). Geographic regions are selected based on which cities/regions collectively contain a majority of economic activity.

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 fair, 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 participate.

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 census 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 sales in each firm; winsorized at the 99th percentile.
- Firm size: Total number of employees in the 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 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 $2/3$ 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. Variables are deflated using the corresponding price index (base December 1998) in December of the year of reference as follows. Revenue is deflated using the three-digit industry-level producer price index.[49] Capital and investment are deflated using the new capital expenditures price index. Payroll is deflated using the employment cost index. Cost of intermediate inputs is deflated using the intermediate goods and service price index.

  To implement TFP(OP) in Stata, we use the *prodest* Stata package (OP method) for firms with positive levels of investment in each 2-digit industry using:

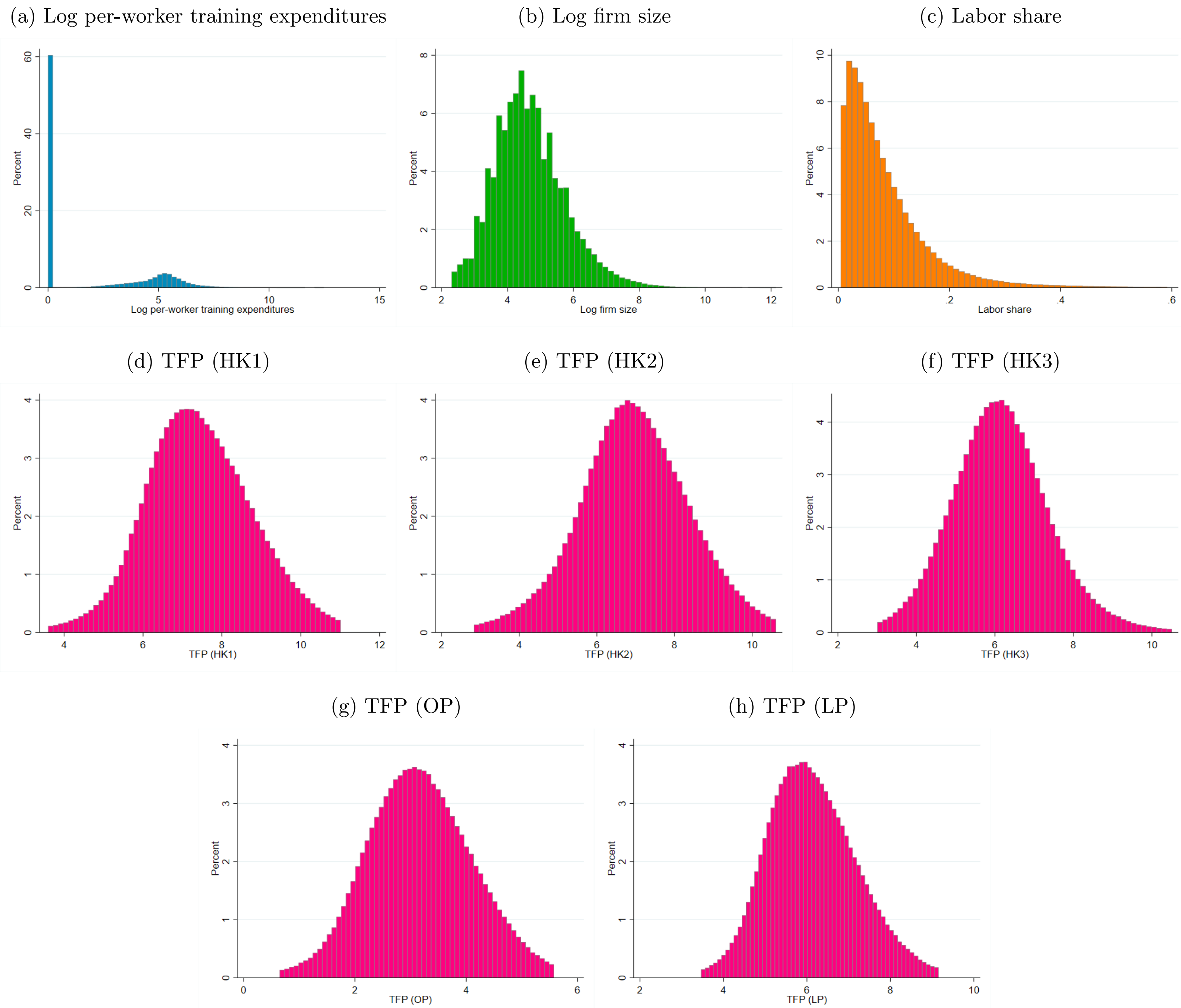
[49] In instances where this information was not available we used the two-digit industry-level producer price index instead.

- outcome variable: log value added
- "free" variable (variable input): log firm size and log payroll
- "state" variables: log book value of fixed capital and age
- "proxy" variable, used as an instrument for productivity: log investment, measured as net purchases of fixed capital.
- the attrition option to control for firm exit

To implement TFP(LP) in Stata, we use the *prodest* Stata package (LP method) for firms with positive purchases of intermediate inputs in each 2-digit industry using:

- outcome variable: log value added
- "free" variable (variable input): log firm size and log payroll
- "state" variables: log book value of fixed capital and age
- "proxy" variable, used as an instrument for productivity: log expenditures in intermediate goods and services
- the attrition option to control for firm exit

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



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.[50] 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 or 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 number of workers that received training in the past year and 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.
- 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 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,

[50] 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.

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. Variables are deflated using the corresponding price index (base June 2012) in December of the year of reference as follows. Revenue is deflated using the three-digit industry-level producer price index.[51] Capital and investment are deflated using the new capital expenditures price index. Payroll is deflated using the employment cost index. Cost of intermediate inputs is deflated using the intermediate goods and service price index.

  To implement TFP(OP) in Stata, we use the *prodest* Stata package (OP method) for firms with positive levels of investment in each 2-digit industry using:

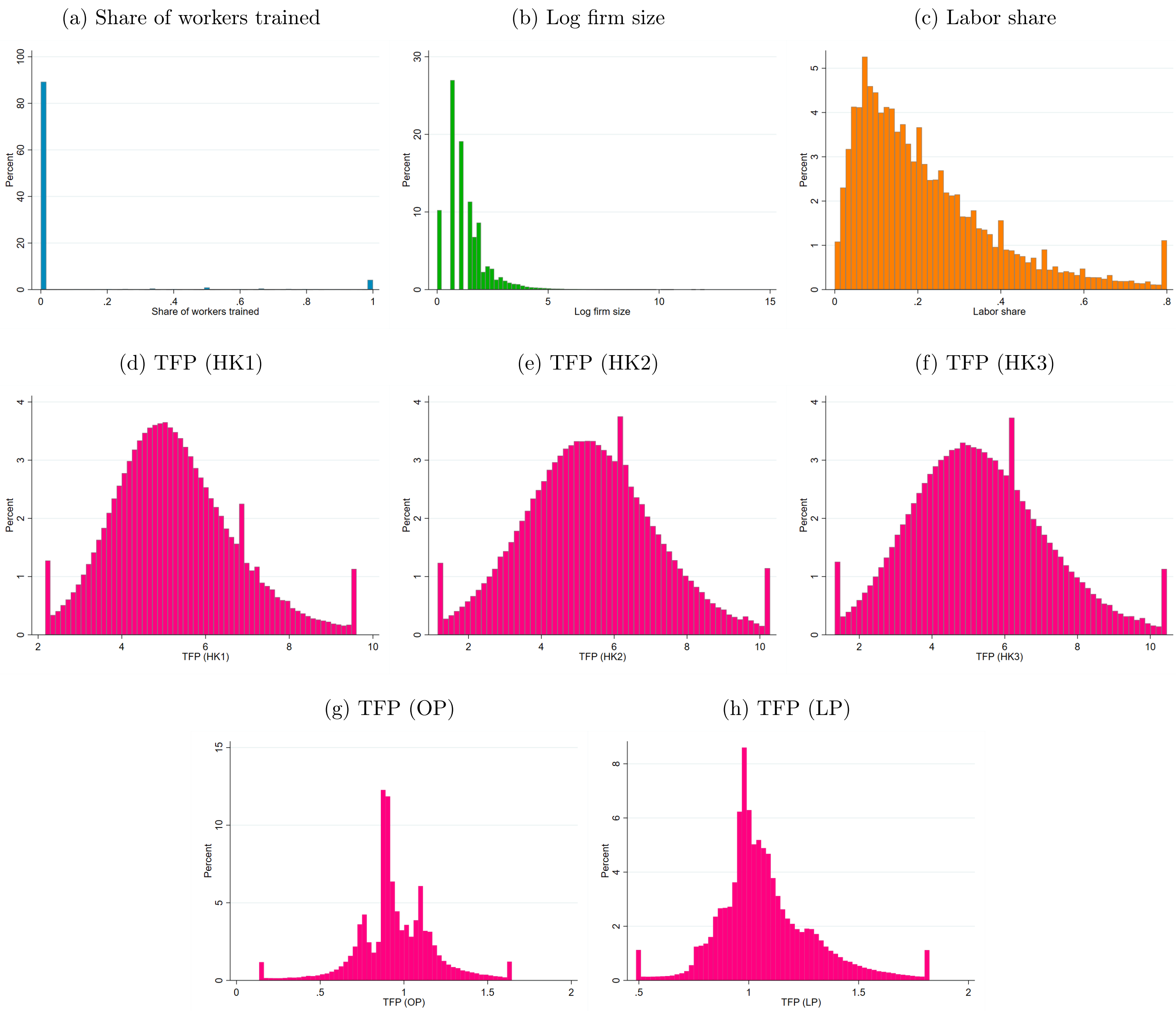
  - outcome variable: log revenue
  - "free" variable (variable input): log firm size and log payroll
  - "state" variables: log book value of fixed capital and age
  - "proxy" variable, used as an instrument for productivity: log investment, measured as net purchases of fixed capital.
  - the attrition option to control for firm exit

  To implement TFP(LP) in Stata, we use the *prodest* Stata package (LP method) for firms with positive purchases of intermediate inputs in each 2-digit industry using:

[51] In instances where this information was not available we used the two-digit industry-level producer price index instead.

- outcome variable: log revenue
- “free” variable (variable input): log firm size and log payroll
- “state” variables: log book value of fixed capital and age
- “proxy” variable, used as an instrument for productivity: log expenditures in intermediate goods and services
- the attrition option to control for firm exit

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



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 US 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.7 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.[52]

We leverage information collected in this survey on both whether workers engaged in training in the last year, and whether this training was desired or not, to further support our findings regarding the importance of firms in training decisions.

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

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

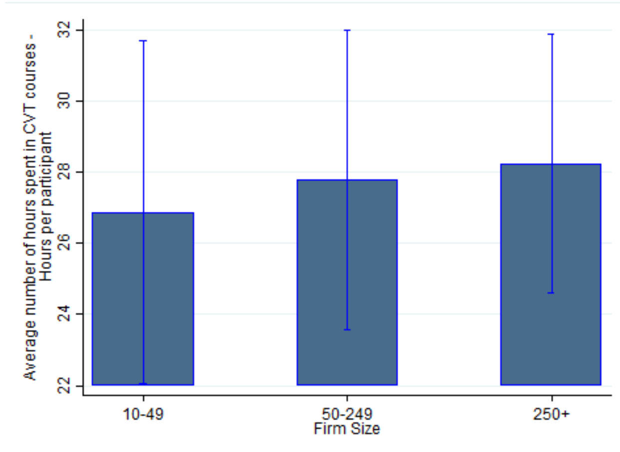
Notes: This table shows different specifications in which we regress a dummy variable indicating if the firm offers formal training to its workers on firm size. Data come from the WB-ES. Industry FE correspond to two-digit industries. Robust standard errors in parentheses. *p <0.05, **p <0.01, ***p 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

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

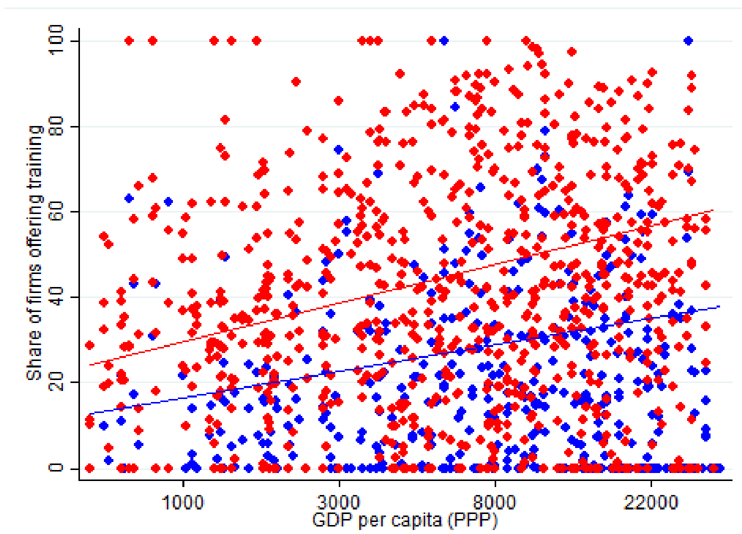
Notes: This table presents the average shares of firms reporting formal training for different purposes by firm size in the EU-CVT in 2010.

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 hours spent in training for each group. Data comes from the EU-CVT.

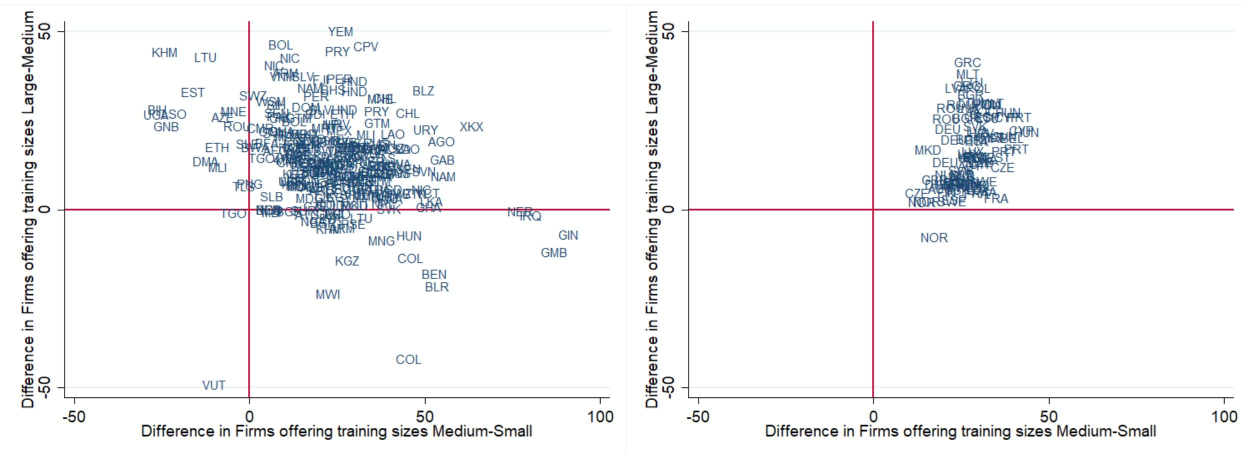
Figure B.2: Share of firms offering formal training by relative firm size



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 come from the Penn World Tables.

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. 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				
	(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. Industry FE correspond to two-digit industries. Robust standard errors in parentheses. *p <0.05, **p <0.01, ***p <0.001.

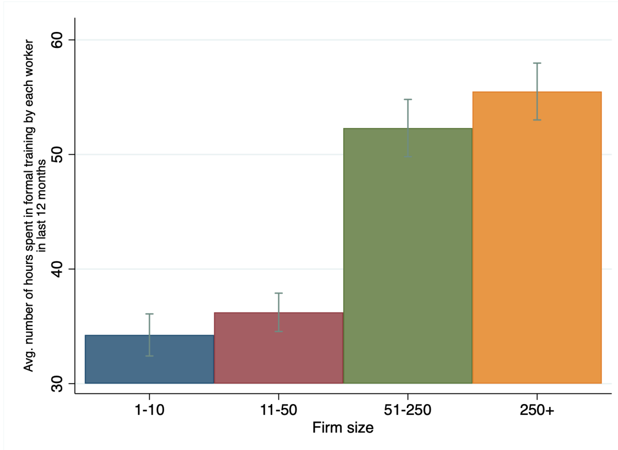
B.2 Results using worker-level data (PIAAC)

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. All regressions are weighted using observations' weights provided in the surveys. 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.

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	-0.762***	-0.013	-1.056***	-0.053	-1.733***	-0.669***	0.016***	-0.050***	-0.049***
	(0.042)	(0.066)	(0.042)	(0.066)	(0.039)	(0.068)	(0.002)	(0.001)	(0.001)
TFP (HK1)	0.111***	0.083***					0.015***		
	(0.002)	(0.004)					(0.0002)		
TFP (HK2)			0.081***	0.077***				0.003***	
			(0.002)	(0.004)				(0.0001)	
TFP (HK3)					0.034***	0.073***			0.003***
					(0.003)	(0.004)			(0.0001)
Log firm size	0.264***	-0.034***	0.276***	-0.046***	0.297***	-0.071***	0.053***	0.061***	0.061***
	(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
Firm FE		Y		Y		Y			
Observations	772,501	658,304	768,216	654,538	773,619	659,894	1,561,690	1,558,774	1,561,672
R-squared	0.061	0.708	0.060	0.709	0.060	0.709	0.208	0.206	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. 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	-0.929***	0.040	-0.019***	-0.039***
	(0.044)	(0.070)	(0.043)	(0.067)	(0.005)	(0.002)
TFP (OP)	-0.001	0.056***			0.024***	
	(0.004)	(0.006)			(0.006)	
TFP (LP)			0.119***	0.110***		0.031***
			(0.004)	(0.006)		(0.002)
Log firm size	0.308***	-0.052***	0.238***	-0.073***	0.080***	0.062***
	(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
Firm FE		Y		Y		
Observations	771,645	657,216	771,646	657,829	122,381	719,736
R-squared	0.060	0.709	0.058	0.707	0.249	0.147

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. 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.

B.4 Additional theoretical results

B.4.1 Proof of Proposition 1

We derive the marginal value of additional human capital for firms and workers by considering their respective optimization problems. A young worker's present value of income in firm z is:

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

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

where $h' = 1 + \zeta s^{\gamma_s}$ 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:

$$\underbrace{(1-\beta(z))z}_{\text{current profits}} - \underbrace{\mu_F\,(z+c_s\bar{w})}_{\text{firm's per-unit training costs}} \times \underbrace{s}_{\text{training level}} +$$
$$\frac{1}{1+\rho} \times \underbrace{(1-\delta)\left[1-\eta\bar{F}(w(z))\right](1-\beta(z))\,z}_{\text{future profits, from workers who stay}} \times \underbrace{h'}_{\text{next-period human capital}}.$$

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 $MR_F(z)$.

B.4.2 Proof of Proposition 2

1. Firms determine training level. When the firms determine the training level, according to equation (3), the training level is given by (after some reorganization):

$$s(z) = \left(\frac{\zeta\gamma_s MR_F(z)}{(1+r)\mu_F(c_s\bar{w}+z)}\right)^{\frac{1}{1-\gamma_s}} = \left(\frac{\zeta\gamma_s(1-\delta)}{(1+r)\mu_F}\left[1-\eta\bar{F}(w(z))\right](1-\beta(z))\frac{z}{c_s\bar{w}+z}\right)^{\frac{1}{\gamma_s}}.$$

Effect of productivity on training. On the right-hand side, the first term $\zeta\gamma_s(1-\delta)/(1+r)\mu_F$ is a constant. Holding the labor share $\beta(z)$ constant across firms, wages $w(z)=\beta(z)z$ increase with productivity z, and therefore the second term $\left[1-\eta\bar{F}(w(z))\right]$ increases with productivity z. Moreover, if the labor share $\beta(z)$ is constant, the third term $(1-\beta(z))$ remains unchanged with productivity z. Finally, the fourth term $\frac{z}{c_s\bar{w}+z}$ increases with productivity z. Thus, we have that $s(z)$ increases with firm productivity z.

Effect of the labor share on training. Holding firm productivity z constant, we find that the first and fourth terms remain unchanged. The second term $\left[1-\eta\bar{F}(w(z))\right]$ increases with the labor share $\beta(z)$, as $w(z)=\beta(z)z$ increases with the labor share $\beta(z)$. The third term $(1-\beta(z))$ decreases with the labor share $\beta(z)$. If the on-the-job search intensity η is small, the change in the second term is small, and thus the third term dominates, suggesting a negative relationship between the training level $s(z)$ and the labor share $\beta(z)$.

2. Workers determine training level. According to equation (4), when workers determine the training level, this is given by:

$$s(z)=\left(\frac{\zeta\gamma_s\left[(1-\delta)\left[1-\eta\bar{F}(w(z))\right]\beta(z)z+(1-\delta)\eta\int\mathbf{1}_{\{w(u)>w(z)\}}\beta(u)udF(w(u))+\delta\int\beta(u)uf(w(u))w'(u)du\right]}{(1+r)\mu_W(c_s\bar{w}+z)}\right)^{\frac{1}{1-\gamma_s}}.$$

Effect of productivity on training. Since productivity z will affect both the numerator and denominator of the right-hand term, productivity z has an ambiguous impact on the optimal training level z.

Effect of the labor share on training. Since the numerator increases with $\beta(z)$ and the denominator remains unchanged with $\beta(z)$, the labor share $\beta(z)$ has a positive impact on the optimal training level z.

3. Joint match value is maximized. According to equation (5), when the joint match value of firms and workers is maximized, the training level is given by:

$$s(z)=\left(\frac{\zeta\gamma_s\left[(1-\delta)\left[1-\eta\bar{F}(w(z))\right]z+(1-\delta)\eta\int\mathbf{1}_{\{w(u)>w(z)\}}\beta(u)udF(w(u))+\delta\int\beta(u)uf(w(u))w'(u)du\right]}{(1+r)(c_s\bar{w}+z)}\right)^{\frac{1}{1-\gamma_s}}.$$

Effect of productivity on training. As productivity z affects both the numerator and denominator of the right-hand term, productivity z has an ambiguous impact on the optimal training level z.

Effect of the labor share on training. The derivative of the numerator with regard to $\beta(z)$ is given by $\zeta\gamma_s(1-\delta)\eta f(w(z))(1-\beta(z))z^2>0$. As the denominator remains unchanged with $\beta(z)$, the labor share $\beta(z)$ thus has a positive impact on the optimal training level z.

B.5 Further evidence on importance of firms for training decisions

B.5.1 Importance of job turnover on training

First, motivated by the fact that the possibility of workers leaving the firm after being trained depresses the incentives firms have to provide training, we explore the role of job turnover in driving training investments by regressing the share of formally trained workers in each country-year in the EU-CVT on the predicted probability of staying in the same firm after a quarter.

Although ideally we would exploit the cross-country job turnover rates built by Donovan

Table B.8: Correlation between job turnover and training

Dep. variable:	Proportion of workers exposed to formal training
	(1)
$\log(GDP^{ppp}_{pc})$	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 firm after a quarter, PPP per capita GDP, 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, Government Effectiveness, 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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

et al. (2023) to do this, the timing of these measures does not match that of our training data. However, since job turnover is linked to the ease of contract termination and thus institutional quality, we predict job turnover rates for our country-years of interest by regressing the job turnover measures of Donovan et al. (2023) on the following institutional measures from the World Bank Worldwide Governance Indicators: Voice and Accountability, Political Stability, Government Effectiveness, Regulatory Quality, Rule of Law, and Control of Corruption.

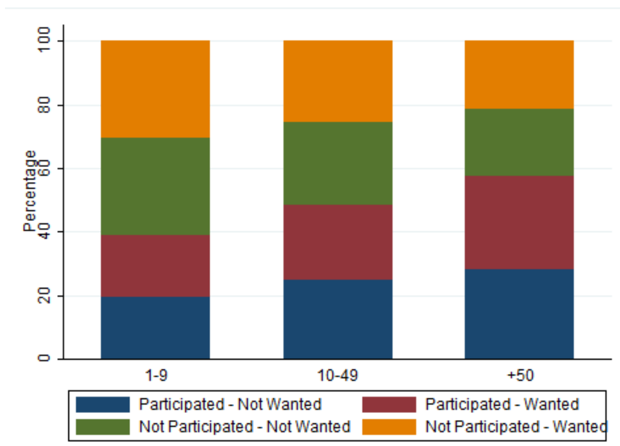
We present the results in Table B.8. We find that higher predicted job turnover rates are associated with lower levels of training even after controlling for country income. This is consistent with firms playing a key role in deciding and paying for training investments, since job turnover depresses the incentives for firms to provide training, but not for workers.

B.5.2 Workers receiving unwanted training

Second, we use worker-level data from the Adult Education Surveys conducted in the EU in 2011 and 2016 (EU-AES) containing information on both whether workers engaged in training in the last year, and whether this training was desired or not, in order to document that a sizeable share of workers receive training even when not wanted.[53]

In Figure B.5, we plot the share of workers reporting whether they participated in training in the last 12 months or not, and whether this training was (or would have been) desired by firm size. Consistent with our main empirical findings, we find that workers in larger firms report higher rates of training participation (red and blue segments). However, the split between whether this training was desired or not is pretty even across all firm sizes, with roughly half of the workers who participated in training reporting this was not wanted. This further suggests that firms are playing a pivotal role in driving training decisions.

Figure B.5: Proportion of workers by training participation, training desirability, and firm size



Notes: We plot the proportion of workers categorized by their training status and their desire for training, segmented by firm size. The categories include: (i) workers who were trained but did not want to be trained, (ii) workers who were trained and wanted to be trained, (iii) workers who were not trained and did not want to be trained, and (iv) workers who were not trained but wanted to be trained. Data comes from European Union Adult Education Surveys (EU-AES) conducted in 2011 and 2016.

C Additional information about quantitative model

C.1 Conditions for simulations

Workers' value With linear utility, workers' utility is determined by the discounted income flows that are earned with current human capital and potential 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

[53]We provide further information about the EU-AES in Appendix A.7.

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 employees ($a < J$), we can solve for their respective values through backward induction:

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

$h' = e(h,z) = 1 + (1-d)(h-1) + \zeta(s^a(h,z))^{\gamma_s}$ denotes the next-period's human capital, where d captures human capital depreciation above workers' innate abilities (normalized to 1), and $s^a(h,z)$ is the optimal training level as described below. To simplify our notation, we use $e(h,z)$ 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:

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

where $h' = e(h,z)$ 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,

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

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

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

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

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.

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

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

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 varies depending on the cost-sharing scenario.

- When the firm bears the full cost of training ($\mu_W(z) = 0,\ \mu_F(z) = 1$), the training level is $s^a(h,z) = s^a_F(h,z)$.
- When the worker bears the full cost of training ($\mu_W(z) = 1,\ \mu_F(z) = 0$), the training level is $s^a(h,z) = s^a_W(h,z)$.
- In the case of joint internal efficiency, where the cost share is based on the relative benefit received by each party ($\mu_W(z) = \frac{\partial W^a(h,z)/\partial h'}{\partial W^a(h,z)/\partial h' + \partial F^a(h,z)/\partial h'},\ \mu_F(z) = \frac{\partial F^a(h,z)/\partial h'}{\partial W^a(h,z)/\partial h' + \partial F^a(h,z)/\partial h'}$), the training level is $s^a(h,z) = \left(\frac{\zeta\gamma_s}{(z+c_s\bar{w})}\left(\frac{\partial F^a(h,z)}{\partial h'} + \frac{\partial W^a(h,z)}{\partial h'}\right)\right)^{1/(1-\gamma_s)}$.
- When workers and firms pay a constant share of training costs ($\mu_W(z) = \mu_W,\ \mu_F(z) = \mu_F$), the training level is $s^a(h,z) = \min\{s^a_W(h,z), s^a_F(h,z)\}$.

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

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

The optimality condition for vacancies is given by:

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

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).

C.2 Model dynamics across cost-sharing scenarios

In this section, we further consider the different training cost-sharing scenarios in the quantitative model and their predictions regarding the level of training in different firms.

C.2.1 Marginal returns and costs of training

In Figure C.1 we plot the marginal returns to training for workers and firms across different firm productivity levels. It is worth noting that these returns do not depend on the division of training costs we capture in our four scenarios. We find that the marginal returns to training are lower for firms than workers at all levels of productivity. This reflects the hold-up problem discussed in Section 4, per which firms underinvest in training due to the possibility of workers leaving the firm after being trained. In addition, we find that these marginal returns increase for both parties as productivity increases due to the complementarity between productivity and human capital acquisition in the production functions, but faster so for firms. This stems from two sources. First, as firms become more productive, the probability of losing the worker is lower which alleviates the hold-up problem. Second, as suggested by Panel (b), the marginal profit per efficiency unit $(z - w(z))$, which shapes the marginal revenue from training received by firms, rises faster than the wage per efficiency unit $w(z)$, which shapes the marginal revenue from training received by workers.

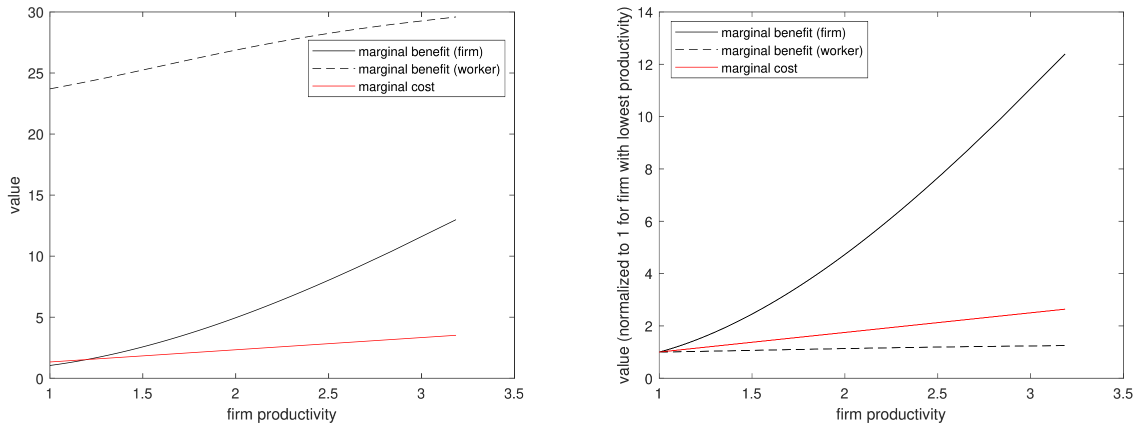
We also find that the marginal costs of training are higher in more productive firms. This is because in our baseline model, the majority (around 80%) of training costs are the opportunity cost arising form the loss of production time, and this opportunity cost increases with firm productivity.

C.2.2 Training levels in our four scenarios

Figure C.2 shows the optimal level of training in each of our four cost-sharing scenarios. When the firm pays all of the training costs, training levels increase with productivity since firms determine the training level, and firms' incentives to train rise with productivity. The same is true in the calibrated cost-sharing scenario, since here training is also fully determined

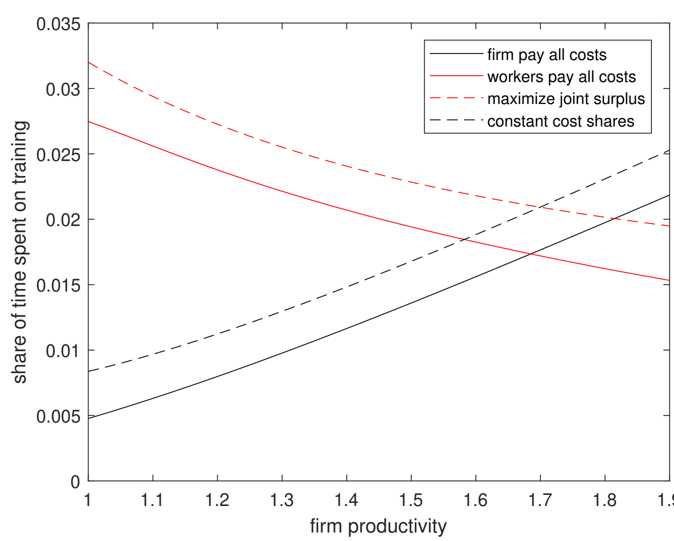
Figure C.1: Marginal benefits and costs of training

(a) Absolute values (b) Normalized values



by the firm. This matches key evidence in the literature showing that workers in more productive firms exhibit faster rates of skill acquisition (Engbom (2021), Arellano-Bover (2020), Arellano-Bover and Saltiel (2023)), and follows from the joint effects of productivity and the labor share documented in the analytical model, particularly since the labor share is lower in higher productivity firms.

Figure C.2: Training levels in four scenarios

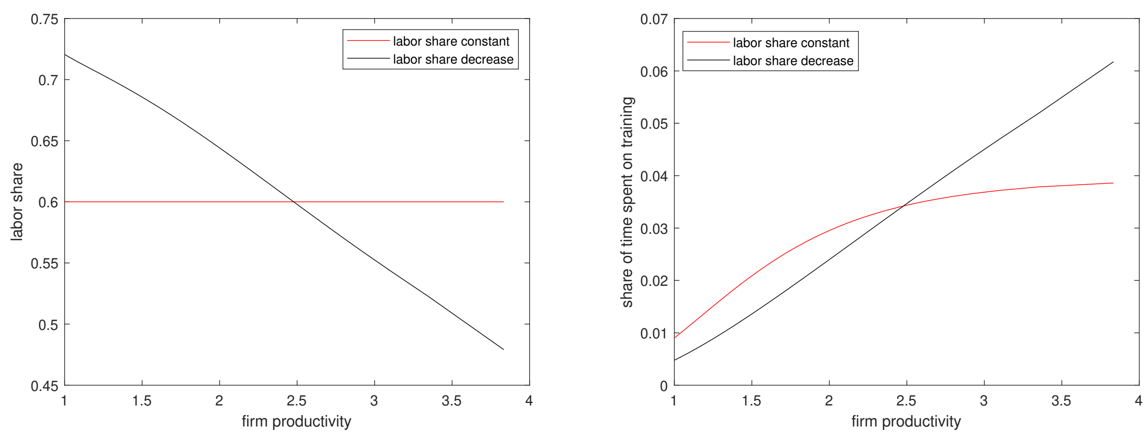


The cases of joint internal efficiency and workers paying all training costs are different. When workers pay all explicit training costs, training levels will be lower in more productive firms since (1) workers lose a significant portion of compensated time when training, and this opportunity cost is higher in more productive firms; (2) workers' returns from training change slowly with the current firm's productivity since they also incorporate the benefits after leaving the firm; and (3) the labor share is lower in more productive firms. In the case of joint internal efficiency, even though the allocation of match value between firms and workers is irrelevant, the training level also decreases with firm productivity. This is because the first and second reasons mentioned above drive the net benefits from training to decline with firm productivity for workers, thus discouraging training investments.

C.3 Additional results

C.3.1 Importance of labor share to training levels

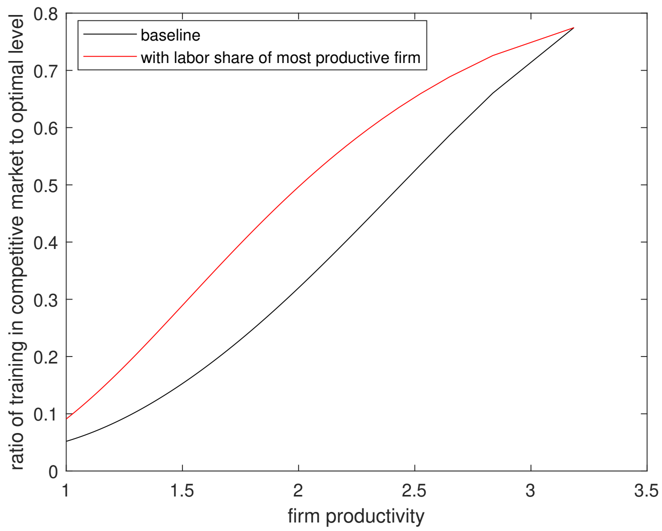
Figure C.3: Two cases: constant and decreasing labor shares



Notes: The decreasing labor share results stem from our baseline case.

C.3.2 Importance of labor share to training inefficiencies

Figure C.4: Social planner and competitive equilibrium



C.4 Training subsidies in the United States

Table C.1: Training subsidies in the United States

State	Year	Subsidy or incentive to employer
Alabama	2014 - present	75% of training costs reimbursed
Arizona	2015 - 2020	50-75% of training costs reimbursed
Colorado	2018 - present	60% of training costs reimbursed
Florida	1993 - present	50-75% of training costs reimbursed
Georgia	1994 - present	50% of training costs tax deductible
Hawaii	1991 - present	50% tuition costs reimbursed
Illinois	1992 - present	50% of training costs reimbursed
Kentucky	1984 - present	50% of training costs reimbursed
Maryland	1989 - present	50% of training costs reimbursed
Massachusetts	2008 - present	50% of training costs reimbursed
Mississippi	2013 - present	50% of training costs reimbursed
Montana	2005 - present	Funding of $5,000 for training
Nebraska	2005 - present	Funding of $800-4,000 for training
New Hampshire	2007 - present	50% of training costs reimbursed
New Jersey	1992 - present	50% of training costs reimbursed
New Mexico	1972 - present	50-75% of training costs reimbursed
Pennsylvania	1999 - present	Funding of $600-1,200 per trainee
Rhode Island	2006 - present	50% of training costs reimbursed
Washington	1983 - present	50% of training costs reimbursed
Wisconsin	2012 - present	50% of training costs reimbursed
Wyoming	1997 - present	Funding of $1,000 per trainee

C.5 Extended model with two education groups

C.5.1 Model extension

In the baseline model, we assumed workers were homogeneous ex ante. As an extension, we now differentiate between two education groups: college-educated workers and non-college workers. These groups have distinct initial human capital levels and search in separate labor markets, each characterized by different wage rates and vacancies posted by firms.[54] Given that high-education workers typically experience higher returns to experience than low-education workers (e.g., Islam et al., 2019), we assume that these two groups have distinct human capital return parameters from training, denoted as ζ_c for college-educated workers and ζ_{nc} for non-college workers.

All simulation conditions are the same as those in Appendix C.1, except that the conditions are now separated by education group and are therefore denoted with a subscript for each group. For example, the firm's optimal training level, $s^a_{F,g}(h,z)$, is given by

$$s^a_{F,g}(h,z) = \left(\frac{\zeta_g \gamma_s}{\mu_{F,g}(z)\,(z + c_s \bar{w}_g)} \frac{\partial F^a_g(h,z)}{\partial h'} \right)^{1/(1-\gamma_s)},$$

where $\frac{\partial F^a_g(h,z)}{\partial h'}$ captures the firm's return of an extra efficiency unit of human capital in the next period for workers of group $g \in \{c, nc\}$. The worker's optimal training level, $s^a_{W,g}(h,z)$, is given by

$$s^a_{W,g}(h,z) = \left(\frac{\zeta_g \gamma_s}{\mu_{W,g}(z)\,(z + c_s \bar{w}_g)} \frac{\partial W^a_g(h,z)}{\partial h'} \right)^{1/(1-\gamma_s)},$$

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

C.5.2 Calibration

We calibrate the model to align with the data moments, assuming that 53% of each cohort of workers is college-educated, based on the estimate by Barro and Lee (2013) for the US in 2005. We set the initial human capital for college-educated workers to 1. For non-college-educated workers, the initial human capital is set to 0.8 to match the education gap between

[54] We assume that firms determine the number of vacancies to post separately in each of the two labor markets. Since we lack separate labor market tightness data for these groups, we treat the parameters for vacancy-cost functions as identical across both markets.

the two groups.[55] By introducing group-specific human capital increments from training, $\{\zeta_c, \zeta_{nc}\}$, in place of the general human capital increment ζ, we add one additional data moment—the relative ratio of returns to 40 years of experience between college-educated and non-college-educated workers. All other data moments remain the same as those in Table 5.2.[56]

Table C.2: **Calibrated parameters**

Parameter	Firms pay	Workers pay	Maximize match value	Constant cost shares
c_M - Constant in matching function	0.92	1.06	1.05	0.98
η - On-the-job search intensity	0.20	0.38	0.35	0.24
c_s - Ratio of training costs per time to wage	0.27	0.24	0.24	0.27
c_v - Constant in vacancy function	1.13	0.45	0.47	0.85
ζ_c - Constant in training function, college-educated	0.20	0.02	0.02	0.06
ζ_{nc} - Constant in training function, non-college	0.15	0.01	0.01	0.04
γ_s - Convexity of training function	0.49	0.30	0.38	0.26
κ - Parameter of Pareto productivity dist	5.35	14.38	10.97	5.52
δ - Exogenous separation rate	0.07	0.07	0.07	0.07
$\mu_F(z)$ - Share of training costs paid by firm	1	0	Value share	0.32

Table C.2 shows that the calibrated parameter values in this model extension closely resemble those in the baseline calibration presented in Table 5.3. Regarding the new parameters, we find that the human capital increment from training is higher for college-educated workers (ζ_c) compared to non-college workers (ζ_{nc}), indicating faster learning for higher-skill workers. Specifically, when the share of explicit training costs borne by firms is constant and calibrated (column 4), training a young college-educated worker for the full quarter (480 working hours) increases their hourly wage by 6%, while training a young non-college worker for the same period increases their hourly wage by 4%. These estimates align with empirical evidence on US training returns, as reviewed by Leuven (2004) and Bassanini et al. (2005).[57]

In the three other scenarios—where firms pay all explicit training costs, workers pay all explicit training costs, or match value is maximized—the training returns are either too

[55] Specifically, the education gap between the two groups is approximately 4.5 years in 2005. Given that the literature typically estimates the return to a year of schooling for college-educated workers at around 6.8% (Caselli, 2005), we calculate the initial human capital of non-college workers as $1/(1 + 4.5 \times 6.8\%) \approx 0.8$.

[56] Since we now separate labor markets by education level, we aggregate the results from the two labor markets to compute the corresponding data moment for the overall labor market in Table 5.2.

[57] For example, our estimates imply a 2–4% wage increase for 60 hours of training in a quarter, which aligns with Frazis and Loewenstein (2005), who find that 60 hours of formal training increases wages by 3–5%.

low or too high. As documented in Table C.3, this results in wage increases from 40 years of experience that are either too low or too high in these scenarios compared to the data estimates. In contrast, the scenario where firms pay a constant share of explicit training costs aligns well with the data, consistent with the baseline model.

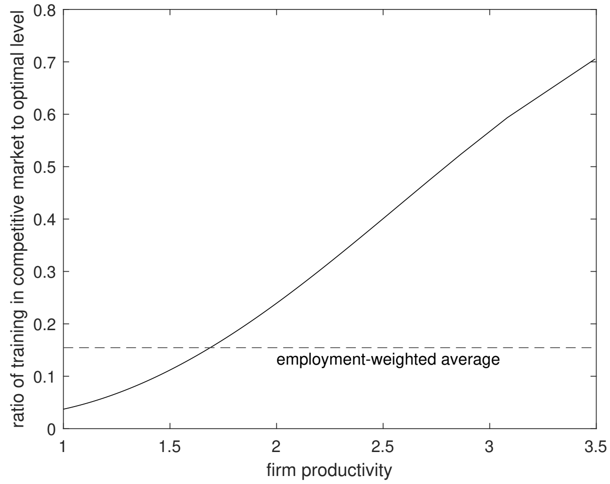
Table C.3: **Moments in the model vs data**

Moments	Data	Model: Firms pay	Model: Workers pay	Model: Maximize match value	Model: Constant cost shares
Panel (a): Targeted moments					
Moments: labor market					
Unemployment rate (%)	6.5	6.7	6.5	6.5	6.5
Ratio of #Vacancies to #Unemployed	0.56	0.55	0.55	0.54	0.55
Pareto parameter of firm size distribution	1.06	1.14	1.12	1.11	1.07
Share of employed people remaining in the same firm after one quarter	0.88	0.89	0.87	0.87	0.88
Share of employed people remaining employed after one quarter	0.94	0.95	0.94	0.94	0.94
Moments: training intensity					
Average training intensity (% time)	2.20	1.99	2.20	2.21	2.18
Ratio of training costs to wage costs of training	0.23	0.24	0.24	0.24	0.24
Moments: training across firms					
Ratio of training intensity in firms with 100-499 employees to that with 50–99 employees	1.19	1.26	0.97	0.96	1.20
Percent wage increase of 40 years' experience (%)	89	-	-	-	90
Relative ratio of experience returns (high-education group versus low-education group)	1.30	1.29	1.30	1.30	1.32
Panel (b): Non-targeted moments					
Percent wage increase of 40 years' experience (%)	89	115	26	21	-

C.5.3 Social planner's problem and training inefficiencies

In Figure C.5, we plot the ratio of training levels in the competitive equilibrium to those chosen by the social planner in this extended model. We find that inefficiencies remain substantial across the entire productivity distribution of firms, with the employment-weighted average training intensity in the competitive equilibrium being only 15% of the level chosen by the social planner.

Figure C.5: Social Planner and Competitive Equilibrium



C.5.4 Training subsidies

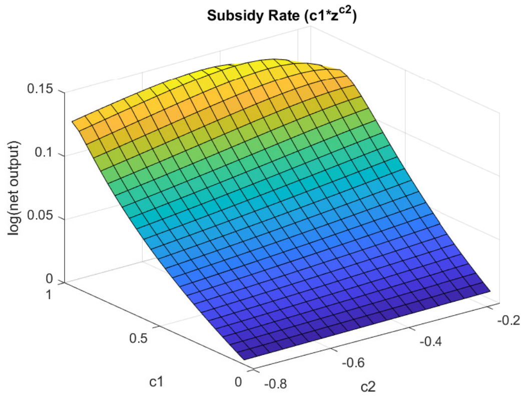
Panels (a)–(c) in Figure C.6 replicate the subsidy exercises from Section 6.2. The results closely align with the baseline findings. When the policy allows varying training subsidy rates across firms with different productivity levels, with the subsidy taking the form $s(z) = c_1 z^{c_2}$, the optimal case shows a 14% increase in net output for the US, with $c_1 = 0.92$ and $c_2 = -0.4$ (similar to the baseline values of $c_1 = 0.92$ and $c_2 = -0.5$). When the policy includes differential subsidy rates for large and small firms, the optimal policy results in a 12% increase in net output, with small firms receiving a 92% subsidy rate and large firms receiving a 69% subsidy rate (compared to 88% and 65%, respectively, in the baseline model). Lastly, when a uniform subsidy rate is applied across all firms, the optimal case produces a 10% increase in net output, with a 73% subsidy rate for all firms (up from 69% in the baseline model).

C.5.5 Changes in labor market concentration

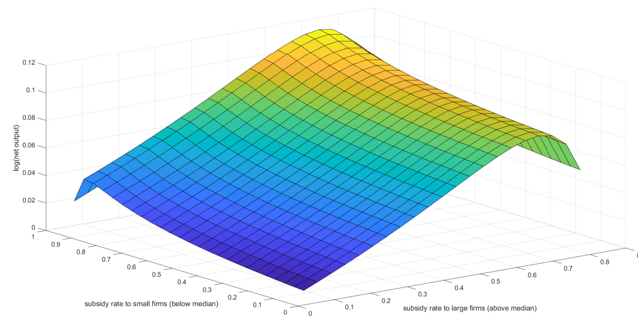
Figure C.7 replicates the labor market concentration exercises from Section 6.3. Consistent with the baseline findings, as the Herfindahl index rises, indicating greater concentration of employment in higher-productivity firms, the average training level in the economy initially increases and then decreases. When the minimum wage and direct training costs from the calibrated economy are held constant, an increase in the Herfindahl index leads to a rise in the average training level across the economy.

Figure C.6: Net Output Gain From Subsidies

(a) Policy targeting firms of different productivity levels



(b) Policy targeting firms in different size brackets



(c) Policy targeting all firms equally

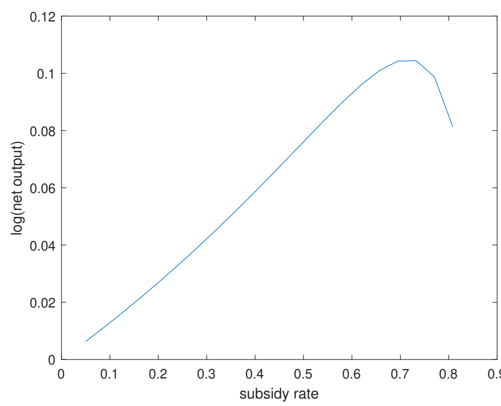
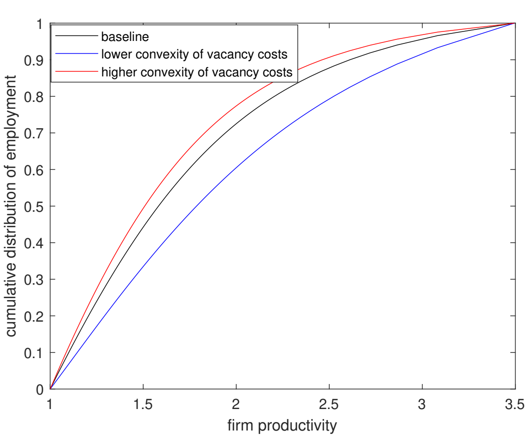


Figure C.7: Concentration and training

(a) Cumulative distribution of employment



(b) Average training