PHBS WORKING PAPER SERIES

Human Capital Investment and Development: The Role of On-the-Job Training

Xiao Ma
Peking University

Alejandro Nakab
Universidad Torcuato Di Tella

Daniela Vidart
University of Connecticut

September 2021

Working Paper 20210904

Abstract

Workers in richer countries experience faster rates of wage growth over their lifetimes than workers in poorer countries. We offer an explanation for this pattern by showing that workers in richer economies receive more firm-provided training. Using cross-country enterprise and worker-level data, we document that the share of workers who receive firm-provided training increases with development, and that firm-provided training is a key determinant of workers' human capital. We then build a general equilibrium search model with firm-provided training investments. Our model suggests that firm-provided training accounts for 43% of cross-country wage growth differences and 10% of cross-country income differences.

Keywords: On-the-Job Training, human capital accumulation, lifecycle wage growth
JEL Classification: E24, J24, O11, O15, J63, J64, M53

Peking University HSBC Business School
University Town, Nanshan District
Shenzhen 518055, China

Human Capital Investment and Development: The Role of On-the-Job Training*

Xiao Ma — *Peking University*

Alejandro Nakab — *Universidad Torcuato Di Tella*

Daniela Vidart — *University of Connecticut*

This Version: July 2021; **First Version:** October 2019

Abstract

Workers in richer countries experience faster rates of wage growth over their lifetimes than workers in poorer countries. We offer an explanation for this pattern by showing that workers in richer economies receive more firm-provided training. Using cross-country enterprise and worker-level data, we document that the share of workers who receive firm-provided training increases with development, and that firm-provided training is a key determinant of workers' human capital. We then build a general equilibrium search model with firm-provided training investments. Our model suggests that firm-provided training accounts for 43% of cross-country wage growth differences and 10% of cross-country income differences.

Keywords: On-the-Job Training, Human Capital Accumulation, Lifecycle Wage Growth
JEL Codes: E24, J24, O11, O15, J63, J64, M53

*Email: xiaoma@phbs.pku.edu.cn, anakab@utdt.edu, and daniela.vidart@uconn.edu.

We are grateful to Titan Alon, Ruixue Jia, Munseob Lee, David Lagakos, Marc Muendler, Tommaso Porzio, Valerie Ramey, Natalia Ramondo and seminar participants at ITAM, University of California San Diego, Universidad de Montevideo, Universidad Torcuato Di Tella and Banco de Mexico for helpful comments. Finally, we would like to thank Kevin Donovan, Will Lu, and Todd Schoellman for sharing data. All the results and conclusions are ours and not those of Eurostat, the European Commission, the World Bank, or any of the national statistical authorities whose data have been used. All potential errors are our own.

1 Introduction

Recent papers have shown that workers in richer countries have faster rates of wage growth over their lifetimes than workers in poorer countries (Lagakos et al., 2018b; Islam et al., 2019). Several different factors can explain this pattern, including cross-country differences in human capital accumulation, labor market frictions, and long-term work contracts. These possible drivers differ in both their scope to explain cross-county differences and their implications for policy. As such, understanding the reasons behind this pattern is a first-order question. In this paper, we offer an explanation for this new stylized fact by focusing on one key source of workers' human capital accumulation: firm-provided training. To that end, we carefully measure workers' post-schooling human capital investments and explore how they differ across countries. Our results provide an explanation for why post-schooling human capital accumulation is greater for workers in more-developed economies, and thus why their lifetime wage growth is higher.

We present both empirical and quantitative evidence on the link between firm-provided training and the level of development. In the empirical portion of the paper, we rely on enterprise surveys covering more than 400,000 firms across 102 countries and worker-level surveys containing detailed information on workers' training investments for more than 600,000 people across 26 countries. These surveys allow us to construct harmonized representative measures of on-the-job training provision across countries with PPP-adjusted GDP per capita ranging from \$1,000 to \$60,000 and thus spanning a broad range of development levels. We document two novel facts.

First, we document that the share of workers who receive firm-provided training rises strongly with country-level GDP per capita. We show that a key margin explaining this positive correlation is poor countries' large share of self-employed workers who do not receive employer-provided training. However, we still find evidence of this positive correlation when we restrict our attention to firm employment. Richer countries have a larger share of firms offering training, along with a larger share of trainees within these firms and a greater share of hours in training relative to total hours worked. In addition, firms in richer countries spend more on training per participant, which potentially reflects training quality.

Second, we show that job-related firm-provided training is the main source of on-the-job human capital accumulation for workers. We find that 72% of all reported adult education corresponds to job-related training and that almost all of this training is financed by firms across all countries. This evidence suggests that firms play a substantial role in adults' human capital investments, and thus canonical models à la Ben-Porath, which do not include firm-level decisions, provide an incomplete picture of the on-the-job skill acquisition process.

To shed light on the mechanisms giving rise to the positive correlation between training and development and its consequences on workers' wage growth, we build a general equilibrium model that explicitly accounts for firm-worker decision-making regarding on-the-job training. The model features two sectors: a self-employment sector and a wage sector. The self-employment sector has no learning opportunities and no frictions. The wage sector, on the other hand, is characterized by labor market frictions and firm heterogeneity à la Burdett and Mortensen (1998). Firms rent physical capital, post vacancies and wages, and meet workers by random search following Mortensen and Pissarides (1994) and Pissarides (2000). We incorporate general training investments that follow the theoretical framework developed by Acemoglu (1997), Acemoglu and Pischke (1998), and Moen and Rosén (2004). However, we depart from this literature in the way training costs are allocated between workers and firms and by incorporating richer job turnover dynamics based on on-the-job search and contract quality. In our model, workers can be separated from firms for two reasons: an exogenous separation shock that may lead workers to unemployment, and on-the-job search as workers look for new job offers while working. When employed workers receive a new job offer, they can choose to exert efforts to break their contract, incurring costs that depend on the economy's contract quality.

We calibrate our model to match representative economies at different income levels. We focus on three main channels that vary greatly across different stages of development to explain the training gap between poor and rich countries: differences in self-employment shares, job turnover rates, and physical capital endowments. The emphasis on these channels stems both from our empirical findings and the literature. The focus on self-employment is motivated by our empirical evidence, particularly by the fact that the high prevalence of this type of work is a key mechanism driving low firm-training investments in poor economies. The focus on job turnover and thus on labor market frictions is rooted in the training literature, specifically the fundamental problem of financing training investments first identified by Becker (1964). In this problem, training investments are less likely to occur when the probability of losing the worker is higher. Finally, we also focus on physical capital differences, because capital-skill complementarities could also affect the returns to skills and shape the incentives for training (Krusell et al., 2000). These channels may play important roles in depressing training investments in developing countries, since these countries exhibit higher worker turnover (Donovan et al., 2020), higher self-employment shares (Gollin, 2002, 2008), and lower physical capital endowments (Krusell et al., 2000).

First, we find that the model explains 61% of wage growth differences across all countries.[1] We then decompose the wage growth predicted by our model across all income levels into training and job turnover components in order to quantify their relative importance. We find that training explains 70% of the cross-country differences in wage growth profiles predicted by our model. Thus, firm-provided training accounts for about 43% of cross-country wage growth differences. In addition, we find that the contribution of firm-provided training to explaining workers' lifecycle wage growth is large for every economy, but declines somewhat with income. This decline stems from the high level of job destruction in the poorest economies, which prevents workers from climbing up the job ladder. As countries' income increases, fewer workers are separated from their jobs and become unemployed, which generates larger increases in wages through job-to-job transitions.

We then conduct a factor decomposition analysis in order to explore the evolving importance of the different channels to explain the training gap at different stages of development. We find three main results. First, we perform a sectoral accounting analysis and find that a third of the aggregate training gap between the poorest and richest economies is explained by differences in the share of the self-employment sector in aggregate employment. We also find that the importance of self-employment slightly decreases with income. This is driven by the high self-employment shares in poor economies stemming from the high labor market frictions and low physical capital and aggregate productivities prevalent in the wage sector.

Second, we find that labor market frictions constitute the main driver of the differences in training investments across countries, explaining around 80% of the training gap at all income levels. The higher job separation rates prevalent in low- and medium-income economies and stemming from job destruction and job-to-job

[1] The model matches most of the cross-country differences in wage growth for countries above $10,000 of GDP per capita, but overpredicts wage growth for economies at the bottom of the world income distribution.

transitions not only could lead to higher shares of self-employment, but also depress the incentives to invest in training in the wage sector. When we decompose the importance of these labor market frictions along its two key components, we find that job destruction is the most important factor explaining the lack of training in poorer economies, while frictions in job-to-job transitions are more important in explaining the training differences between more-developed economies. Finally, we show that differences in physical capital productivity and sectoral productivity levels jointly explain the remaining 20% of the training gap. The difference in physical capital productivity mainly drives the training gap between the richest and poorest economies, while differences in sectoral productivity levels drive the training gap between middle- and high-income countries.

Third, we show that on-the-job training explains 10% of the income differences across countries in our quantitative model. Thus, the contribution of on-the-job training to cross-income differences is sizeable. Lagakos et al. (2018a) show that experience could explain around 20% of the income differences across countries.

Related Literature. This paper relates to several strands of literature. First, our theory combines insights from two related strands of the literature studying on-the-job human capital accumulation. Our model builds on the theoretical literature on general training investments, first proposed by Becker (1964), and later developed by others such as Acemoglu (1997), Acemoglu and Pischke (1998), and Moen and Rosén (2004). By embedding this firm-worker training investment dynamic into a search model, our work also relates to the literature that tries to disentangle the contributions of human capital and search dynamics to earnings (e.g., Bunzel et al., 1999; Rubinstein and Weiss, 2006; Barlevy, 2008; Yamaguchi, 2010; Burdett et al., 2011; Bowlus and Liu, 2013; Bagger et al., 2014; Gregory, 2021). These papers differ from ours along several key dimensions. First, a large contingent of these papers assume that on-the-job human capital accumulation does not follow from an optimization problem where workers face tradeoffs between work and learning and is thus simply an exogenous by-product of work.[2] Second, the focus of these papers contrasts sharply with the goal of our theory, which is to explain cross-country differences in training and income. In particular, this literature analyzes how job search and human capital accumulation

[2]Exceptions to this are Wasmer (2006) and Flinn et al. (2017), who incorporate micro-founded human capital investment decisions. However, they focus on studying the distinction between firm-specific and general training.

contribute to explaining workers' wage growth for specific developed economies. We contribute to this literature by extending this decomposition analysis for countries at all income levels.

By exploring the role of workers' training in explaining differences in GDP per worker across countries, our paper relates to a large strand of the literature that measures the importance of different factors in explaining cross-country income differences (e.g., Klenow and Rodriguez-Clare, 1997; Caselli, 2005; Hsieh and Klenow, 2010), and in particular to studies focusing on human capital.[3] Our paper focuses on one understudied source of cross-country human capital differences, namely on-the-job human capital accumulation. Thus, our work relates to the recent literature that highlights the potential importance of differences in life-cycle human capital accumulation across countries (De la Croix et al., 2018; Lagakos et al., 2018a,b; Islam et al., 2019). This literature, however, does not explain how these cross-country differences in on-the-job human capital accumulation patterns emerge. Our paper attempts to fill this gap by delving into the processes and features giving rise to the low skill acquisition prevalent among workers in poor countries by focusing on employer-provided training.

Third, our paper is related to the literature that explores the relationship between labor market dynamics and development. In particular, we incorporate insights from: (1) the literature on cross-country job turnover differences (Donovan et al., 2020); (2) the vast literature documenting cross-country capital intensity differences, as reviewed in Caselli (2005); and (3) the literature focusing on cross-country differences in self-employment shares (e.g., Gollin, 2002, 2008; Poschke, 2018). We contribute to this development literature by incorporating the interaction between these channels and firm-provided training.[4]

Our paper also relates to two recent papers focusing on a cross-country analysis of training. The first of these papers is Doepke and Gaetani (2020), who focus on the

[3]These papers focus on explaining cross-country productivity differences by quantitatively measuring the role of educational attainment (e.g., Hall and Jones, 1999; Erosa et al., 2010; Jones, 2014), school quality (e.g., Hanushek and Woessmann, 2012; Schoellman, 2012, 2016), parental influence on learning (De Philippis and Rossi, 2021), and skill specialization in secondary and post-secondary curricula (Alon, 2017; Alon and Fershtman, 2019).

[4]Moreover, through the interaction between employment distribution across firms and training, this paper relates to the misallocation literature, which studies the productivity losses stemming from the extensive existence of small unproductive firms in developing countries (e.g., Hsieh and Klenow, 2009; Restuccia and Rogerson, 2013; Bento and Restuccia, 2017; Poschke, 2018). Our paper focuses on documenting a new channel causing productivity losses: the lack of on-the-job training.

effect of employment protections on firms' and workers' incentives to invest in skills in order to study cross-country differences in on-the-job skill acquisition. The second paper is Engbom (2021), who studies how the costs of doing business affect human capital formation using a search model featuring endogenous human capital investments. Our work differs from both of these papers by focusing on different channels to explain on-the-job training differences, which include different labor market frictions, physical capital endowments, and self-employment. More importantly, we focus on explaining the trend component of training with respect to per-capita GDP, while they study different channels that vary across countries but may not directly explain the relationship between income and training. While these papers focus on developed economies, we provide evidence and quantitative analysis for countries at all stages of development.

Finally, by analyzing human capital differences at all stages of development, our paper also relates to Manuelli and Seshadri (2014). In their paper, Manuelli and Seshadri (2014) focus on worker-level decisions on human capital while abstracting from firm-level decisions. They find that the lower TFP levels prevalent in developing economies raise the costs of accumulating human capital, thus lowering households' incentives to invest in human capital after schooling. In this paper, we offer a very different explanation for this phenomenon by focusing on firm-provided training, which we also show is a key component of adults' human capital investments.

The paper is organized as follows. Section 2 introduces our data and empirical findings. Section 3 presents the theory, and in Section 4, we calibrate a quantitative version of the model. Section 5 shows evidence on the drivers behind the wage growth differences across countries and the factor decomposition of training, and we present the income accounting results. In Section 6 we conclude.

2 Empirical Evidence on On-the-Job Training

In this section, we start by describing the data sources and defining key concepts. We then document some facts about on-the-job human capital accumulation and the development process. We include further details in Appendix Section A.

2.1 Data Description

To document our cross-country facts, we rely on labor and firm surveys for more than 100 countries. For developing countries, we use the World Bank Enterprise Survey (WB-ES). For developed countries, on the other hand, we rely on the European Union Labor Force Survey (EU-LFS), the Adult Education Survey (EU-AES), and the Continuing Vocational Training (EU-CVT) enterprise survey. Our cross-country evidence encompasses developing and developed economies with per-capita GDP ranging from \$1,000 to \$60,000.

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. The ES usually 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. The ES has a set of country-specific questions reflecting each country's characteristics and a set of standardized questions that enable cross-country comparison. We rely on the two ES waves, between 2002 and 2005 and between 2006 and 2017, which have standardized questions on workers' training provisions. We use the second wave (which provides individual weights) for the main analysis.

For the EU enterprise data, we rely on the EU-CVT. This survey provides information on enterprises' investments in continuing vocational training of their staff, providing information on participation, time spent, and the costs of such training. Due to data availability, in our analysis we rely on three of the five waves of EU-CVT conducted in 2005, 2010, and 2015, which cover all EU member states and Norway.

For the European countries' worker-level data, we rely on data from the EU-LFS and EU-AES. The EU-LFS is a large household survey that provides data on labor force participation, unemployment, job characteristics, socioeconomic characteristics, and education and training of adults (ages 15+). The survey is conducted in all of the EU member countries and the three European Free Trade Association countries. Although the data collection dates back to 1983 for some countries, the data series are generally available from 1992 according to EU membership. We use the data ranging from 2009 to 2018 for all countries to ensure consistency. Finally, the EU-AES collects information on participation in education and learning activities including job-related training, among others. Thus, this survey is conducted with the specific objective of understanding adult education patterns. The AES is one of the main data sources for

EU lifelong learning statistics and it covers around 666,000 adults ages 25–64. These data were collected during 2007, 2011, and 2017 in 26, 27, and 28 EU member states, respectively.

2.2 Defining On-the-Job Training

We first carefully define training and its characteristics to ensure consistency across different data sources and to be able to provide meaningful economic interpretations through the lens of the model. We present more detailed definitions of training and other sources of human capital for comparison in Appendix Section B.

We define *training* following the definition of "Non-formal Education and Training" category from ISCED (2011),[5] which is any organized and structured learning activity outside the formal education system. This definition has two main features. First, it differentiates *training* from *schooling*, as *training* is a learning activity that happens outside the formal education system. Therefore, *training* does not consist of programs such as MBAs that may be a source of human capital for workers. Second, this activity must have a certain degree of organization and structure, which differentiates *training* from other *informal learning* activities such as reading journals, visiting museums, or learning through media in an unstructured or unplanned way.[6]

We also decompose *training* into *formal training* and *informal training* to further ensure consistency across different data sources. *Formal training* has a structured and defined curriculum and includes classroom work, seminars, and workshops, among other activities planned in advance. Formal training activities are typically separated from the active workplace and show a high degree of organization by a trainer or institution. Furthermore, this training is typically more general and not geared towards tasks, machinery, or equipment specific to certain jobs or workers. *Informal training* is less structured and more related to job-specific skills for workers. It also differs from formal training in that it is tailored to specific workers' needs and is connected to the active workplace. Thus, informal training tends to be more hands-on and

[5] 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."

[6] *Informal learning* is defined as a type of a learning activity that is not structured and is more related to workers' self-investments. This category encompasses the following activities in the data: learning by reading printed material or using computers, learning through media (television, radio, or videos), learning through guided tours in industrial sites or museums, and visiting learning centers such as libraries. These are predominantly self-directed and employers are usually not involved.

task related. It encompasses guided on-the-job training, job rotation, exchanges, and other forms of learning through colleagues and training arising from participation in learning circles.

2.3 On-the-Job Training Facts

The wide variety of data sources allows us to analyze training patterns for 102 countries in the main analysis and describe in detail the key sources giving rise to adults' human capital accumulation. In this section, we document two key facts about firm-provided training.

Fact 1 ***There exists a positive cross-country correlation between firm-provided training and income.***

We first focus on formal training, of which the data is available from and consistent across the enterprise surveys (WB-ES and EU-CVT) for 102 countries in our data set, to study the cross-country correlation between on-the-job training and income.[7] We construct country-year measures of the share of employees who receive formal training with the following formula:

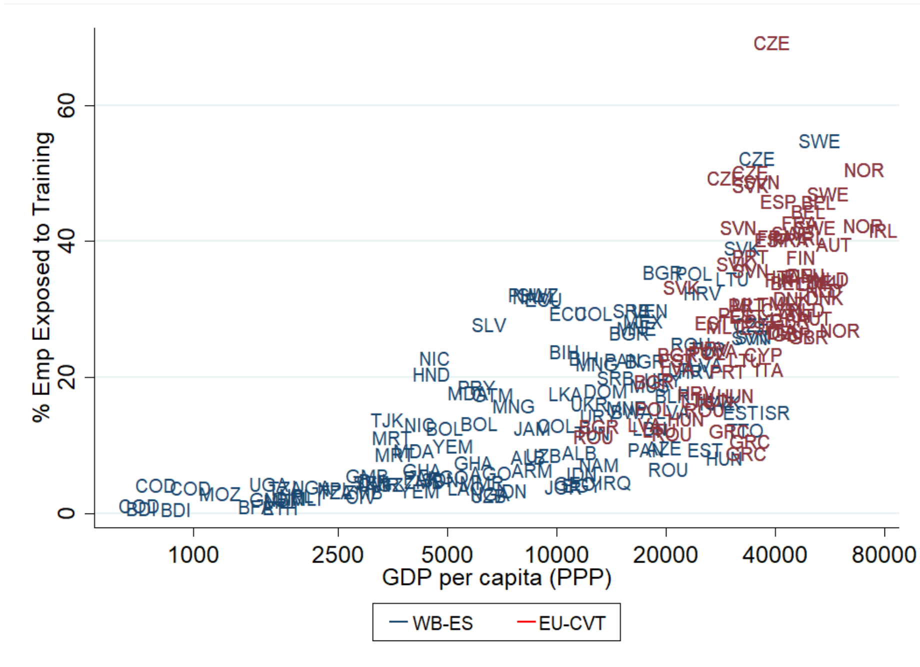
$$\%\,Trained\ Workers = \frac{Firms'\ Trained\ Workers}{All\ employees\ in\ firms} \times (100 - Self.\ Emp.\ Share)$$

The WB-ES and EU-CVT provide information about which firms provide training, along with the share of workers who receive training in those firms. We use these two measures to construct the country-year measure of the share of employees offered training. Since only firms are surveyed, we then adjust this measure by the share of self-employment for the main specification, assuming that self-employed workers do not receive training from employers.[8]

[7] Initial vocational training, employee orientation, and apprenticeships are excluded.

[8] We restrict the sample from the WB-ES to 2005–2015 for comparability with the EU-CVT. The WB-ES tends to overweight larger firms, which causes mean firm-based employment to be counterfactually large in some countries. Poschke (2018) shows the log mean employment is lower than 4 even for countries with more than $60,000 of GDP per worker for different data sources. Thus, we restrict our sample from the WB-ES to all countries with log mean employment lower than 4 to avoid countries largely overweighting large firms. We show that the pattern documented is robust to performing the analysis on the unrestricted sample in Appendix Figure C.1.

Figure 1: Share of Formally Trained Employment and Development



Notes: The share of formally trained employment follows from adjusting the share of workers who receive training from firms by the share of self-employment. Data on the share of employees trained within firms comes from the WB-ES for all developing economies and from the EU-CVT for European economies. Both surveys contain data on whether firms provided formal training in the previous fiscal year and the share of employees who participated. For the WB-ES we use the standardized wave with data from 2005 to 2017 for which we have firm weights. We restrict the sample from the WB-ES for the years between 2005 and 2015 to have the same years as the EU-CVT, and we restrict the data to countries with log mean employment in firms lower than 4. Data on GDP per capita and self-employment comes from the Penn World tables and World Bank Indicators, respectively.

Formal on-the-job training increases with development. In Figure 1, we show the results of our combined measure of on-the-job training and GDP per capita. We find that as countries become more developed, on-the-job training increases substantially. In particular, for the poorest countries in our sample, with a per-capita GDP of about \$1,000, only approximately 5% of workers are exposed to training. In contrast, this share rises to approximately 50% for the richest countries, with great variation in between. It is also noteworthy that the data from the WB-ES and the data from the EU-CVT overlap for the income range common to both, denoting both harmony between the training definitions and a consistent pattern between training and income in the two data sources.

Self-employment is a key driver of low levels of training in poor economies. We now show that the large share of self-employment prevalent in developing countries is key to explaining the low levels of on-the-job training in these countries. In Panel (a) of Figure 2, we show that the share of workers who are offered training rises with income even when unadjusted for self-employment. However, the difference between poor and rich economies is more compressed in this case, suggesting that the high share of self-employment exhibited in poor countries, and the strong correlation of this with income—evidenced in Panel (b)—are a key factor driving low training levels in poor economies.

Figure 2: Unadjusted on-the-job Training Shares and Self-Employment

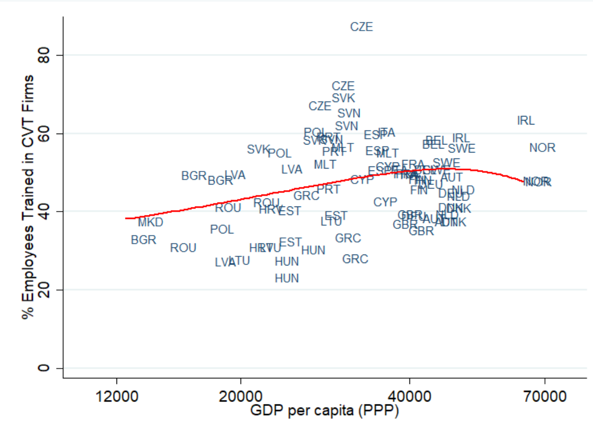
(a) % Trained workers in Firms **(b)** Self-employment

Notes: This figure shows both margins from the formal training measure: in Panel (a) we show workers who are trained by their employer as a share of total workers in firms, and Panel (b) shows the share of workers who are self-employed. Data comes from the WB-ES for all developing economies and from the EU-CVT for European economies. Data on GDP per capita and self-employment comes from the Penn World tables and World Bank Indicators, respectively.

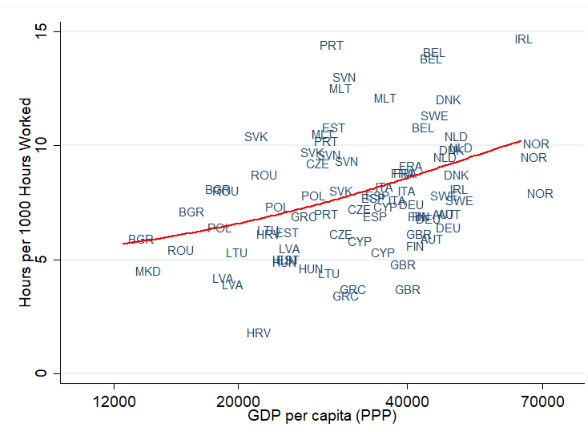
The wage sector's training increases with development in every margin. We now analyze the detailed relationship between training and income for workers employed by firms only (the wage sector) using enterprise survey data from European countries. Although we rely on fewer countries, the relatively wide survey time frame and country coverage allow for sizeable income variation. We find that the positive correlation between formal training and income is prevalent along both the extensive and intensive margins. In Figure 3, we show that richer countries exhibit both a larger share of firms offering training (extensive margin), and larger shares of trainees and more hours in training relative to total hours worked within these firms (intensive margin). In addition, richer countries exhibit a higher cost of training per participant,

Figure 3: OTJ Training Margins within the Wage Sector

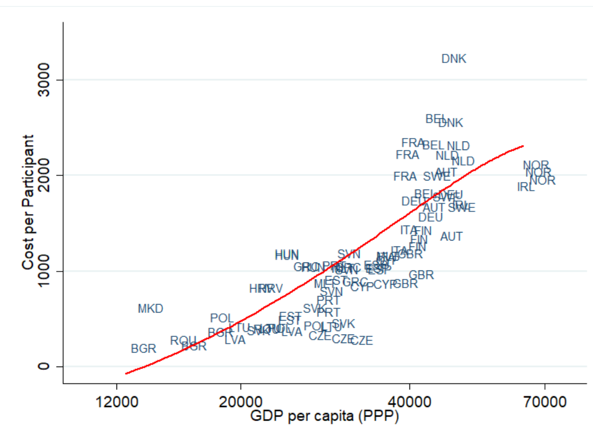
(a) % Firms Offering Training



(b) % of Participants within Firm



(c) Hours per 1000 Hrs Worked



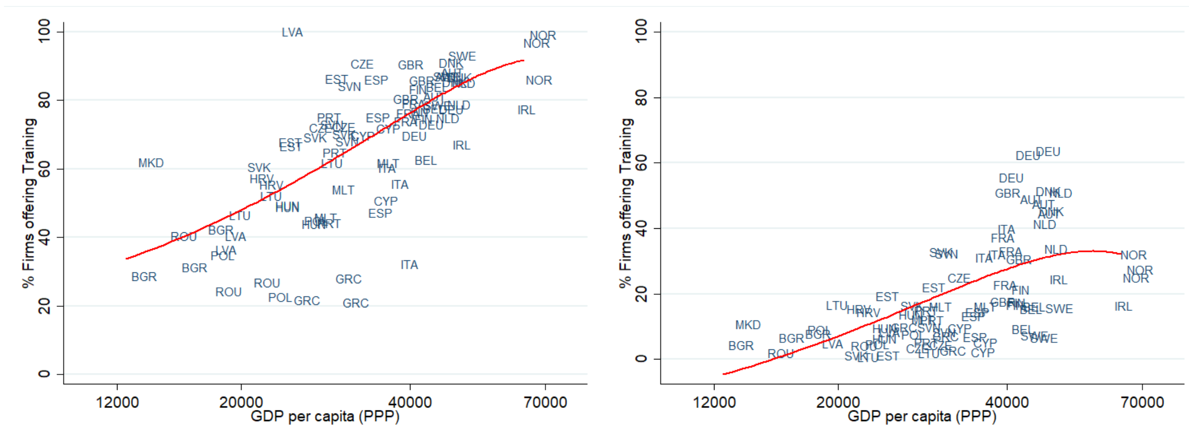
(d) Cost per Participant

Notes: This figure shows all margins of training. Panel (a) shows the share of firms that offer training, which was any type of continuing vocational training in the previous fiscal year. Panel (b) shows the share of participants within the firms who participated in training, conditional on the firm offering training at all. Panel (c) shows the hours per 1000 hours worked by all employees in the firms (those who did and who did not participate in the training). Panel (d) shows the total cost of training per participant that includes both direct and indirect training costs (wages of trainers and wage lost by not working during training). Data comes from the EU-CVT. Data on GDP per capita comes from the Penn World tables.

which potentially proxies training quality.

Both continuing and initial vocational training increase with development. Our formal training measure is based on continuing vocational training, which does not include worker orientation or initial training, and seems like the most relevant margin to explain life-cycle increases in productivity. Nevertheless, it could be the case that continuing vocational training is more prevalent in developed economies, but initial vocational training (IVT), which takes place when the worker starts the job, is more prevalent in developing economies. Although we do not have measures of the share of workers who receive initial vocational training, we do have measures on the share of firms offering IVT and CVT, which are depicted in Figure 4. In both cases, there is a positive correlation with development. As countries become richer, firms invest more in both IVT and CVT, which rules out the possibility that our results stem from a difference in the timing of human capital investments across countries.

Figure 4: Share of firms Offering CVT and IVT



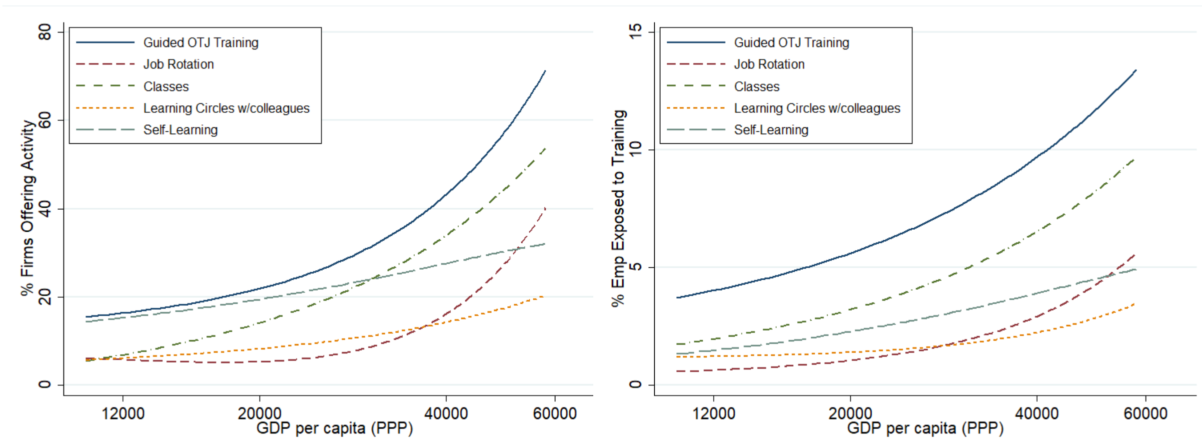
(a) Firms Offering CVT

(b) Firms Offering IVT

Notes: Panel (a) shows the share of firms that offer continuing vocational training (CVT), and Panel (b) shows the share of firms offering initial vocational training (IVT). CVT is defined as all training for workers except for initial training. IVT includes coaching workers on job-specific skills for a new job or teaching workers general knowledge about the firm as they enter a new job. Data comes from the EU-CVT. Data on GDP per capita comes from the Penn World tables.

Informal on-the-job training increases with development. For our previous results we focused on formal training in order to cover countries in all stages of development. However, using the EU-CVT we are able to show evidence on the relationship between income and informal training, which is typically connected to the active workplace and is often tailored according to the learner's individual needs. This is important because more-developed countries could be providing more formal training at the expense of informal training. For all EU countries in 2005 and 2010, for which we have detailed data, we construct measures of the share of employees trained and the share of firms offering five different types of training: guided on-the-job training; job rotation and exchanges; participation in conferences, workshops, trade fairs, and lectures; participation in learning or quality circles; and self-directed learning. In Panels (a) and (b) of Figure 5 we plot the quadratic fit of the training measures with respect to GDP per capita for the share of firms that offer each one of these activities and the share of workers who participate, respectively. We find that all informal training activities increase with development.[9]

Figure 5: Informal Training



(a) % Firms Offering Activities

(b) % Workers Offered Activities

Notes: This figure shows five types of "other forms of continuing vocational training" that are not considered as CVT: planned training through guided on-the-job training; planned training through job rotation, exchanges, secondments, or study visits; planned training through participation (instruction received) in conferences, workshops, trade fairs, and lectures; planned training through participation in learning or quality circles; and planned training through self-directed learning/e-learning.[10] Data comes from the EU-CVT. Data on GDP per capita comes from the Penn World tables.

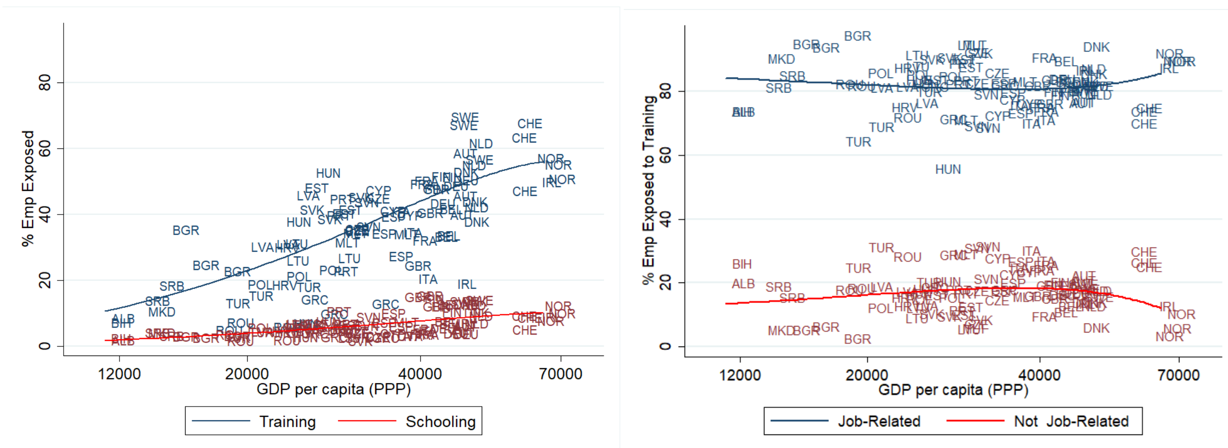
Fact 2 ***Firm-provided training is the main source of adult education.***

We have shown a strong correlation between on-the-job training and development using enterprise-level data. However, on-the-job training may represent a small fraction of adults' human capital investments, reducing the scope of this positive correlation to explain cross-country human capital differences. Thus, we now turn our attention to labor force and worker-level surveys containing detailed information

[9]It might be possible that due to a lack of resources, firms in poor countries do not offer training and workers replace this human capital source with other types of worker informal learning. However, this does not seem to be the case. Appendix Figure C.4 provides measures of all types of informal learning in the AES survey (e.g., learning from peers and by using printed material or media, among others), and we show these have at best weakly positive correlations with development.

on workers' training activities and education, which allow us to quantify the role of on-the-job training relative to other human capital sources of workers' learning. In particular, we focus on data from the EU-AES and the EU-LFS, which collect information on the characteristics of all education and training investments in European countries.

Figure 6: Adults' Human Capital Accumulation Characteristics

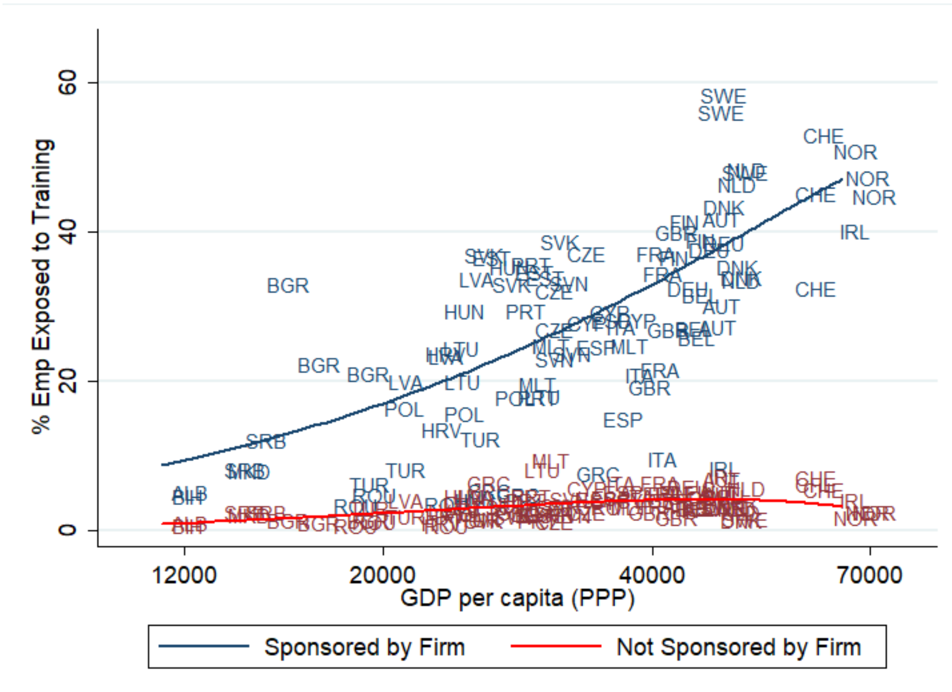


(a) Type of Education **(b)** Job Relation of Training

Notes: Panel (a) shows the difference in the share of adults who participate in any type of educational activity. "Training" refers to our definition of informal + formal training, or the category of education defined by "Non-formal Education and Training" from the International Standard Classification of Education 2011 (ISCED 2011). "Schooling" refers to "Formal Education and Training" according to the ISCED 2011. Panel (b) presents the share of job-related training in all the training reported in Panel (a) (blue line). Data comes from the EU-AES. Data on GDP per capita comes from the Penn World tables.

Most of adult education is job-related training. In Figure 6 we show how the proportion of workers exposed to different types of education varies with cross-country income. Panel (a) shows that the vast majority of adult education (around 90% of all adult education reported in the past year) is training, while less than 10% is schooling. Additionally, Panel (b) shows that around 80% of workers who report participating in some type of training (blue measures in Panel (a)) claim that this is job related, and interestingly this share is uncorrelated with cross-country income. Moreover, Appendix Table C.1 shows the same pattern in terms of the share of adults who report being involved in training in the Labor Force Survey (EU-LFS). On average, 84% of adults in European countries report that the education they receive is job related and only 16% mention personal or social reasons as the purpose of their training or education. This evidence suggests that job-related training is a primary source of adults' learning and human capital accumulation.

Figure 7: Training Financing



Notes: The graph shows the difference in share of adults who participate in firm-sponsored and non-firm-sponsored job-related training relative to GDP per capita. Data comes from the EU-CVT. Data on GDP per capita comes from the Penn World tables.

Almost all of the job-related training is sponsored at least partially by firms. Figure 7 shows how training financing varies across European countries. In particular, we look at how the proportion of job-related training financed at least partially by the firm or completely by the worker varies with income. The graph shows that the vast majority of job-related training is sponsored by firms; less than 5% of workers for all countries receive some training directly related to their job that is entirely self-financed. Moreover, the share of adults who fully self-finance their job-related education is constant as a function of per-capita GDP, which reflects the fact that the correlation between job-related training and income is driven by firms offering more training, and not by workers themselves investing more in education. Doing some back-of-the-envelope accounting, our results show that 90% of all reported adult educational investments correspond to training, and 80% of all training is job related and firm sponsored. This means around 72% of all human capital investments are at least partially provided and financed by firms. [11]

Training Decomposition by Occupation, Industry, Educational Level and Firm Size. Finally, we exploit the rich EU-AES private data containing information on training participation, occupation, industry, education level of workers, and firm size in order to account for how much of the cross-country training differences documented above stem from differences in observables, particularly differences in the share of workers in "high training" bins. To this end, we decompose the trend in on-the-job training using shift-share accounting results. We first calculate the share of trained workers in each specific bin and calculate the aggregate level of training for each country-year survey using the richest economies' weights for those bins.[12] For instance, we calculate the share of trained workers for different occupation categories and then assume that all economies have the richest economies' share of workers in each occupation category. By comparing the slope of the original measure with the one that assumes the same occupation structure for all economies, we can calculate the share of the increase in training driven by the larger share of workers in high-training occupations. We show the main results in Table 1 and provide details on the definitions of each source, the mean level of training for each category, and robustness checks in Appendix Section D. The results suggest that richer countries have larger shares of workers in industries and occupations that require more training. Moreover, richer economies also have more highly educated workers and more employment in larger firms, which taken together partially explain the positive correlation between training and development. The most important individual factor is the occupation heterogeneity, accounting for 11% of the slope of training with respect to per-capita GDP. All these factors jointly explain 21% of the increase in training, which implies these observables drive only a small fraction of our results.

Taken jointly, the patterns shown in this section imply that employer-provided

[11] Moreover, as a robustness check, in Appendix Figure C.3, we provide historical evidence from the United States during the period 1991 to 2005 and show that job-related training is both predominantly paid for by employers in this setting and accompanies economic growth in the time series. In Appendix Table C.1 and Table C.2, we further show that firms and workers report that on-the-job training predominantly occurs during working hours, and its objective is to improve technical and job-specific skills.

[12] We use the mean share of workers in each bin for the top 10% richest economies in the EU-AES sample.

Table 1: Accounting for On-the-Job Training

	Industry	Occupation	Education	Firm Size	All
Share Accounted For (%)	8	11	5	4	21

Notes: This table reports the share of the slope of the increase in training with respect to per-capita GDP that is accounted for by the industry, occupation, education, or firm size structure. For each category we split the sample into three categories with high, medium, and low levels of training according to the mean training levels for the full sample shown in Appendix Section C. Then, we calculate the share of workers in each bin for the top 10% richest economies in the sample and use those shares as weights for all the economies. Finally, we regress each measure on per-capita GDP and calculate the "share accounted for" as 1-$\hat{\beta}/\beta$, where $\hat{\beta}$ is the coefficient of the regression using the richest economies' structure and β is the coefficient of the regression with each particular economy's structure.

training is a key determinant of on-the-job human capital accumulation and that firms play a substantial role in adults' human capital investments. This suggests that canonical human capital accumulation models à la Ben-Porath, which do not include firm-level decisions, provide an incomplete picture of workers' human capital accumulation occurring after formal education or schooling concludes. This, in addition to our first fact showing that on-the-job training increases with development, suggests that understanding firms' decisions to provide training is key to understanding cross-country human capital and income differences.

3 Model

To shed light on the mechanisms giving rise to the positive correlation between training and development and its consequences for workers' wage growth, we build a general equilibrium model that explicitly accounts for firm-worker decision-making regarding on-the-job training. The model features two sectors: a self-employment (or traditional) sector and a wage (or modern) sector. The self-employment sector has no learning opportunities and no frictions. The wage sector, on the other hand, is characterized by labor market frictions and firm heterogeneity

Workers' Preferences. The model economy is populated by a continuum of workers whose lives span two periods. Every period, the same number of workers who die are born, and we normalize the size of each generation's population to be one. All workers are born ex ante equal, but accumulate human capital through training at potentially different rates. Workers offer 1 unit of labor inelastically to the market every period. Their utility is assumed to be linear in consumption, and thus they maximize the present value of consumption:

$$\max_{\{c^Y, c^O, k^Y\}} \; c^Y + \frac{c^O}{1+\rho},$$

$$s.t. \quad Pc^Y = w^Y - \frac{k^Y}{\chi}, \quad Pc^O = \left(\frac{1-\delta_k}{\chi} + R\right) k^Y + w^O,$$

where superscripts Y and O denote young and old ages and $\rho > 0$ governs time preference. We treat the wage sector good as the numeraire, and have that P is the price of the consumption good. Young workers can invest in physical capital to save for the next period. While the Kaldor facts suggest a constant capital-to-output ratio, the recent Penn World Table shows that the capital-to-output ratio increases with development (Inklaar et al., 2019), reflecting capital deepening along the growth path (Acemoglu and Guerrieri, 2008). We follow Krusell et al. (2000) and introduce a capital-specific technological change parameter, χ, denoting that one unit of the wage sector good can be transformed into χ units of capital. We will let this parameter χ vary across countries to capture the increases in the capital-to-output ratio that occur with development in the quantitative analysis.[13]

We assume workers make sectoral choices in the beginning of the first period and thus abstract from workers' reshuffling between sectors afterward, as in Hsieh et al. (2019). This avoids the computational difficulties of tracking the employment and training histories of each worker. Several assumptions can lead to immobility between sectors, including complete asset markets or large switching costs. We will discuss workers' sectoral choices below.

Consumption Good Production. The consumption good is a composite of goods from the two different sectors: the self-employment (or traditional) sector good C_T and the wage (or modern) sector good C_M,

$$C = \left(\gamma C_T^{\sigma} + (1-\gamma) C_M^{\sigma}\right)^{\frac{1}{\sigma}}.$$

[13] In the steady state, if workers consume a non-zero amount of consumption goods in both periods, the discount rate shall be equal to the return of investing in capital, implying that $R = (\rho + \delta_k)/\chi$.

Since the wage sector good is the numeraire, the price of the consumption good is given by

$$P = \frac{1}{1-\gamma}\left(\frac{C}{C_M}\right)^{\sigma-1}.$$

Self-Employment Sector. Production in the self-employment sector is characterized by a constant-returns-to-scale function:

$$Y_T = A_T N_T,$$

where A_T and N_T denote productivity and labor, respectively, in this sector. We assume training is not provided to workers in this sector. All the goods produced by the self-employment sector are used for consumption: $Y_T = C_T$. The price of the self-employment-sector good is given by

$$P_T = \frac{\gamma}{1-\gamma}\left(\frac{C_T}{C_M}\right)^{\sigma-1}.$$

Wage Sector. This sector is characterized by frictional labor markets. There is a unit measure of firms, which are heterogeneous in productivity $z \sim G(z)$ and produce a homogeneous good, which is used for consumption and paying training and vacancy costs, as well as being transformed into physical capital. We incorporate capital-skill complementarities by assuming that physical capital complements human capital in producing the good in the wage sector. Once workers and firms are matched, worker i's production in firm j is given by

$$y_{ji} = A_M z_j \left(\mu h_i^{\epsilon} + (1-\mu) k_{ji}^{\epsilon}\right)^{\frac{1}{\epsilon}},$$

where A_M denotes the productivity in this sector, z_j is the firm-specific productivity, h_i is worker i's efficiency units of labor (human capital), and k_{ji} is the amount of capital rented by firm j to equip worker i. The elasticity of substitution between capital and human capital is $1/(1-\epsilon) < 1$, with capital being complementary to human capital. Since capital is rented at a constant rate R, the firm chooses the optimal capital level k_{ji} to maximize net revenues from worker i's production given her level of human capital,

$$\max_{k_{ji}\geq 0}\ r_{ji} = A_M z_j \left(\mu h_i^{\epsilon} + (1-\mu)k_{ji}^{\epsilon}\right)^{\frac{1}{\epsilon}} - Rk_{ji}.$$

By solving this problem, if k_{ji} yields an internal solution, we obtain:

$$k_{ji}^{*} = \left[\left(\frac{R}{(1-\mu)A_M z_j}\right)^{\frac{\epsilon}{1-\epsilon}} - (1-\mu)\right]^{-\frac{1}{\epsilon}} \mu^{\frac{1}{\epsilon}} h_i$$

and with optimal k_{ji}^{*}, the net revenues from worker i's production is

$$r_{ji} = A_M z_j h_i \mu^{\frac{1}{\epsilon}} \left(1 - (1-\mu)^{\frac{1}{1-\epsilon}} (A_M z_j)^{\frac{\epsilon}{1-\epsilon}} R^{\frac{\epsilon}{\epsilon-1}}\right)^{\frac{\epsilon-1}{\epsilon}},$$

which increases if capital rent R is lower. We note the net revenue is proportional to worker i's human capital h_i. We denote $\tilde{r}(z) = A_M z \mu^{\frac{1}{\epsilon}} \left(1 - (1-\mu)^{\frac{1}{1-\epsilon}} (A_M z)^{\frac{\epsilon}{1-\epsilon}} R^{\frac{\epsilon}{\epsilon-1}}\right)^{\frac{\epsilon-1}{\epsilon}}$ as net revenues per efficiency unit in a firm with productivity z, which facilitates the characterization of training decisions below. By integrating the production across all workers within each firm and across all firms, we get total production in the wage sector:

$$Y_M = A_M \int_{j\epsilon J} z_j \int_{i\epsilon j} \left(\mu h_i^{\epsilon} + (1-\mu)k_{ji}^{\epsilon}\right)^{\frac{1}{\epsilon}}\ di\ dj.$$

Job Search and Matching. Firms post vacancies $v(z)$ at the start of each period, with a contract stipulating the wage rate $w(z)$ and working period, which we assume to be two periods for young workers and one period for old workers. The vacancy cost is defined by $c_v \frac{v^{1+\gamma_v}}{1+\gamma_v}$ and we require vacancy costs to be convex (i.e. $\gamma_v > 0$) to ensure that firms with different productivity levels coexist. The total number of vacancies is then $V = \int v(z)dG(z)$.

There is a probability δ of exogenous destruction of workers' contracts in the beginning of the second period when they become old. These exogenously separated old workers enter the unemployment pool and look for a full-time job jointly with all newly born workers. Moreover, a portion η of the remaining old workers search on the job. Therefore, the amount of searchers is denoted by $\tilde{U} = (1 + \eta(1-\delta) + \delta)N_M$, where N_M is the share of each generation's workers in the wage sector.

For analytical tractability, we assume the matching function is $M(\tilde{U}, V) = \min\{\tilde{U}, V\}$, and that c_v is small enough such that $V > \tilde{U}$, which ensures full employment in the equilibrium. As usual, market tightness is defined by $\theta = \frac{V}{\tilde{U}}$.

Contract Enforceability and Workers' Optimal Separation Policy. If old workers who search on the job get an outside offer, they can make efforts to break their current work contract. Specifically, these workers choose a probability p of breaking their current contract, and incur the costs $c_p^{\gamma_p} \frac{p^{1+\gamma_p}}{1+\gamma_p}$ per efficiency unit. The costs of breaking the contract represent a friction in job-to-job transitions, with a lower c_p representing lower costs of leaving the firm. We assume $\gamma_p > 0$, such that the marginal cost of breaking the contract increases with probability p.[14]

In a firm with productivity z, a worker chooses the optimal leaving probability $p \in [0,1]$ when faced with an outside offer w' by solving[15]

$$\max_{p \in [0,1]} (w' - w(z))p - c_p^{\gamma_p} \frac{p^{1+\gamma_p}}{1+\gamma_p}.$$

We solve for $p(w(z), w')$ which yields a piece-wise function,

$$p(w(z), w') = \begin{cases} 0 & \text{if } w' < w(z) \\ \frac{1}{c_p} (w' - w(z))^{\frac{1}{\gamma_p}} & \text{if } 0 < w' - w(z) < c_p^{\gamma_p} \\ 1 & \text{if } w' - w(z) > c_p^{\gamma_p}. \end{cases}$$

This result is intuitive. If the new wage offer is lower than the wage at the current firm, workers do not want to switch jobs and the investment in breaking the contract is 0. On the other hand, if the new wage offer is large enough ($w' > w(z) + c_p^{\gamma_p}$), workers want to switch firms, and will therefore break the contract with a probability of 1. If the cost of breaking the contract increases, workers are less willing to switch and thus put fewer resources into breaking the contract.

[14]Consistent with previous literature (Acemoglu and Pischke, 1999), in our setting, firms cannot break the contract, as they always benefit from hiring and willingly pay for their share of training costs in the first period. Also note that if firms have long-term reputations, the absence of contractual problems for firms breaking contracts is reasonable.

[15]We solve for workers' optimal choice of leaving probability p while taking the level of training investments as given for two reasons. First, when the new offer arrives in the beginning of the second period, training has already occurred. Second, firms and workers need to internalize workers' probability of leaving the firm when deciding on the optimal level of training. Thus, they must choose training according to the optimal breaking-contract efforts conditional on each new offer.

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

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

where ζ is a constant and $0 < \gamma_s < 1$ governs the diminishing returns of training. In each period, training is decided upon jointly by firms and workers and the cost of training is paid jointly by them when training occurs. There is a constant cost c_s per unit time of training, reflecting trainers' wages and material costs.[16]

It is worth noting that we assume all training raises general human capital, so the benefits accrue even if the worker changes firms. Moreover, we assume that if s_W and s_F are optimal training levels from workers' and firms' perspectives, respectively, training s will be given by $s = \min\{s_W, s_F\}$. This assumption implies that the training level is determined by the party with lower affordability and is thus quite reasonable. For instance, if firms bear all the training costs, workers may desire large training levels, yet firms would not like to pay for them. Thus, the optimal level of training for workers and firms is determined by Proposition 1:

Proposition 1 (Firms' and Workers' Optimal Training Levels) *In a firm with productivity level z, if μ_i is the proportion of training costs borne by group i (workers or firms), then*

$$s_i(z) = \left(\frac{\zeta \gamma_s MR_i(z)}{(1+\rho)\mu_i c_s}\right)^{\frac{1}{1-\gamma_s}},$$

where, in a firm with productivity z, current wage w, new offers of wage w', a wage distribution of offers $F(w)$, and optimal investments to break contract $p(w, w')$ (denoted by $p(w')$), the marginal benefits of training for workers and firms are

$$MR_W(z) = (1-\delta)\left(\underbrace{\left(1 - \eta\int p(w')dF(w')\right)w}_{\text{if stay in current firm}} + \underbrace{\eta\int p(w')w'dF(w')}_{\text{if move to new firm}} - \underbrace{\eta\int c_p^{\gamma_p}\frac{p(w')^{1+\gamma_p}}{1+\gamma_p}dF(w')}_{\text{cost of breaking contract}}\right) + \underbrace{\delta\int w'dF(w')}_{\text{U back to a firm}}$$

[16]In principle, training also reduces trainees' production time. Since the analytical properties of the model will not be affected by these training time costs, we omit them here, but we add them in the quantitative analysis, because this is a key feature of the data.

$$MR_F(z) = \underbrace{(1-\delta)\left(1-\eta\int p(w')dF(w')\right)(\tilde{r}(z)-w)}_{\textit{future profits, from workers who stay}}.$$

Proposition 1 explains how optimal training is determined when we have different divisions of training costs. As the share of training costs paid by each group increases, the optimal level of training decreases. Moreover, taking the share of costs paid as given, workers' training levels depend on the expected wage flows if they stay in the firm or switch employers. On the other hand, firms choose the optimal level of training to maximize their net profits, which increase with firms' productivity and the probability of keeping the worker. One key difference between workers and firms is that firms cannot reap the gains from training after the trained worker leaves.

In this model, firms are willing to invest in general training. This departure from Becker (1964) follows from frictional labor markets, which allow firms to extract partial rents from training (Acemoglu and Pischke, 1999). We differ from the general training literature (e.g., Acemoglu and Pischke, 1999; Engbom, 2021) in that (1) we assume the cost shares paid by workers and firms are common across firms and (2) we add a time cost of training when we take the model to the data.[17]

For the calibrated economy, we find that firm decisions determine training investments, as firms always want lower levels of training than workers. Thus, we now focus on understanding firm-level decisions.

Proposition 2 (Labor Market Frictions and Firms' Training) *In a firm with productivity level z, the firm's optimal training $s_F(z)$*
(1) increases with costs of breaking the contract c_p;
(2) decreases with exogenous separation rate δ; and
(3) increases if capital rent R is cheaper.

The first two results in Proposition 2 indicate that a higher probability of job separation leads to lower training. These results provide the mechanisms through which better contract quality and lower job destruction generate more training investments in our model. In addition, the third result suggests that lower capital rent leads to higher training, which stems from capital-skill complementarities.

[17] We model the economy in this way because when we include training time costs (which are the main costs of training in the data), the constant cost share of training across firms can help us jointly match training patterns by firm productivity and the aggregate training levels. Nonetheless, if the share changes to maximize the joint surplus, the model would generate a decrease in training investments with firm productivity, which is counterfactual.

Additionally, it is worth noting that a wage increase has two opposing forces affecting training decisions for firms. On the one hand, the incentives to invest in training decrease when the wage increases, because firms capture a lower share of the surplus of the match. On the other hand, wage increases raise the probability of keeping the worker, which generates higher training incentives. Interestingly, a wage increase in the aggregate economy does not impact the probability of keeping workers, although the labor shares do decrease. In this case, training investments go down, which means that higher firm competition for workers translates to lower training investments.

Solving Firms' Problem. Firms choose wage $w(z)$, vacancies $v(z)$, and young workers' training $s(z)$ each period to maximize profits. Their value function can be written as

$$J(z, l^O_{-1}, w_{-1}, F_{-1}) = \max_{\{w,v,s\}} \underbrace{l^O_{-1}(1-\delta)\left(1-\eta\int p(w_{-1},w')dF(w')\right)(\tilde{r}(z)-w_{-1})}_{\text{profits from remaining workers}} - \underbrace{\frac{c_v v^{1+\gamma_v}}{1+\gamma_v}}_{\text{vacancy costs}}$$

$$+ \underbrace{\frac{v}{\theta}\frac{\tilde{r}(z)-w-\mu_F c_s s}{1+\eta(1-\delta)+\delta}}_{\text{profits from hiring young workers}} + \underbrace{\frac{v}{\theta}\frac{\eta(1-\delta)\int p(w',w)dF_{-1}(w')\bar{l}(w)+\delta\bar{l}}{1+\eta(1-\delta)+\delta}(\tilde{r}(z)-w)}_{\text{profits from hiring old workers}} + \frac{J(z,l^O,w,F)}{1+\rho}$$

$$s.t. \quad l^O = \frac{v}{\theta}\frac{1}{1+\eta(1-\delta)+\delta}(1+\zeta s^{\gamma_s}), \quad F = \Gamma(F_{-1}), \quad w \geq b\bar{w},$$

where we use the subscript -1 to denote variables that are determined in the last period. l^O_{-1} is the total supply of efficiency units by old workers before exogenous separations and $F_{-1}(w)$ is the wage distribution of job offers during the last period. The first term on the right-hand side represents the net profits generated by all the workers who remain in the firm from the previous period. The second term represents the total vacancy costs. The third term represents the profits from hiring young workers net of training costs. The fourth term represents the profits from poaching old workers who are willing to move to the current firm.[18] Note that $s(z)$ is determined according to Proposition 1, whereas $w(z)$ and $v(z)$ are determined according to FOCs. In particular, $w(z)$ is determined by a first-order differential equation, combined with

[18]On-the-job movers have average efficiency units $\bar{l}(w) = 1 + \frac{\int \zeta p(w_{-1}(z),w)s_{-1}(z)^{\gamma_s}dF_{-1}(w_{-1}(z))}{\int p(w_{-1}(z),w)dF_{-1}(w_{-1}(z))}$, whereas the average efficiency units of unemployed old workers are $\bar{l} = 1+\int \zeta s_{-1}(z)^{\gamma_s}dF_{-1}(w_{-1}(z))$.

the minimum wage $b\bar{w}$, as in Burdett and Mortensen (1998).[19]

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. Because hiring workers generates profits, firms want to post vacancies, but will stop posting eventually as the costs of additional vacancies increase.

Solving Workers' Sectoral Choices. If there is a non-zero measure of workers in both sectors, workers must be indifferent in terms of expected utility between going to the self-employment sector and the wage sector in the beginning of the first period:[20]

$$P_T A_T + \frac{P_T A_T}{1+\rho} = \int_z \left(w(z) - \mu_W c_s s(z) + \frac{1+\zeta s(z)^{\gamma_s}}{1+\rho} MR_W(z) \right) dF(w(z)).$$

The left-hand side of the equation represents the present discounted value of working as self-employed, while the right-hand side shows the expected discounted labor income from working in the wage sector for both periods.

Equilibrium. We now define the model's general equilibrium in the steady state.

Definition 1 *The general equilibrium for this economy is*
(1) workers' decisions over consumption $\{c_i^Y, c_i^O\}$*, savings* k_i^Y*, and sector to work in;*
(2) workers' decisions over optimal breaking contract probability $\{p(w, w')\}$*;*
(3) firms' decisions over physical capital, wages, and vacancy posting $\{k(z), w(z), v(z)\}$*;*
(4) the joint decision of human capital accumulation $\{s_F(z), s_W(z)\}$*;*

[19] 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. Because under these assumptions unemployment benefits do not affect any other equilibrium outcomes, we abstract from unemployment benefits in the model.

[20] In the steady-state equilibrium, labor and goods markets are cleared in both the self-employment and wage sectors in each period. We abstract from workers' reshuffling between the wage and self-employment sectors, because this requires tracking employment and training histories of each worker and is thus computationally intractable. Several assumptions can lead to immobility between sectors, including complete asset markets, family networks, or large switching costs.

(5) aggregate prices $\{P_T, P\}$; and

(6) perceived law of motion for firms' wage distribution $\Gamma(F_{-1}(w))$,

such that:

(i) given prices, wage distribution, and human capital accumulation, (1) solves the households' consumption, saving, and sectoral choices problems, and (2) solves the workers' optimal breaking contract probability problem;

(ii) given prices, workers' leaving rates and wage distribution, human capital accumulation, market tightness, and the perceived law of motion, (3) solves the firm's problem.

(iii) given prices, wage distribution, and workers' leaving rates, (4) solves the optimal training problem for firms and workers;

(iv) perceptions are correct;

(v) workers' total savings equal the total amount of capital demanded by firms; and

(vi) the wage-sector output equals consumption, capital investments, and vacancy and training costs, and the self-employment-sector output equals consumption.

Self-employment and Training. Propositions 1 and 2 provide the mechanisms through which higher costs of breaking contracts, lower job destruction, and higher physical capital stocks generate more training investments in our model. We now consider how training is affected by changes in the share of workers in the self-employment sector. Since the self-employment sector features no learning, any change that lowers the returns of working in the wage sector relative to the self-employment sector generates a decrease in training in the aggregate conditional on the wage sector's investments. For instance, if δ increases, the expected return of working in the wage sector decreases, because it is more difficult for workers to move up the job ladder. This increases the economy's self-employment share and pushes aggregate human capital downward.

4 Quantitative Model

In this section, we extend our two-period analytical model for the quantitative analysis and take our model to the data. Thus, we add some features to closely replicate key aspects of the labor market and economic environment:

Workers. On the worker side, we assume workers live for T periods with discount rate ρ, since job search models are usually calibrated using high-frequency labor flows data. We still normalize each generation's population to be 1, and hence the total population in the economy is T. We assume human capital from training depreciates at rate d every period, and overall human capital remains above a lower bound, which we assume is the level of human capital agents are born with (basic cognitive and physical skills). As in our analytical model, we abstract from workers' reshuffling between the wage and self-employment sectors.

Firms. Training costs are assumed to be proportional to the wage sector average wage $c_s\bar{w}$, while each unit time of training also causes a δ_s decrease in efficiency units for production. This explicitly avoids that training differences are driven by different training costs due to income levels and also captures that training costs (e.g., trainers' wages) could increase with the countrywide income. Similarly, we scale vacancy costs with income levels $c_v\bar{w}\frac{v^{1+\gamma_v}}{1+\gamma_v}$. Moreover, we parameterize the firm's productivity to be Pareto-distributed $G(z) = 1 - z^{-\kappa}$.

Exogenous Job-to-Job Moves. We assume that the moving probability p has some lower bound $\underline{p} > 0$. This aims to capture that a portion of job-to-job flows are associated with wage losses (see Table 4 in Haltiwanger et al. (2018)). The economic intuition is that some job-to-job moves reflect idiosyncratic shocks related to family, health, or geographic reasons.

Labor Market. In the quantitative model, we use a widely used matching function $M(\tilde{U}, V) = c_M\tilde{U}^{\psi}V^{1-\psi}$ for the wage sector. This matching function yields positive unemployment and meaningful elasticities of matches with regard to searchers $\tilde{U}$ and vacancies V.

Conditions for Simulations. The optimal conditions for the quantitative model provide the same intuition as in our analytical model and are presented in Appendix Section E. We show the optimal levels of training conditional on firm productivity and age of workers and the determination of firms' wages and vacancies.

4.1 Model Parametrization and Quantification

We proceed to calibrate the model in two steps. First, we calibrate the model to the United States as our baseline economy. For this, we draw on 15 moments describing labor market dynamics and training investments to identify model parameters. Then, we perform a second calibration for representative economies at 10 different income levels to understand how training investments change with development. For this purpose, we jointly recalibrate the parameters δ, c_p, A_M, A_T, and χ to match self-employment, job destruction rate, job-to-job transition, income levels, and capital-to-output ratios for each representative economy.

4.1.1 Calibrating the Model to the United States

Pre-assigned parameters. We first directly set some parameters following the literature. Our model has quarterly frequency. Thus, 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 whole lifetime is set to $T = 160$ quarters. The ratio of the lowest wage to the average wage is calibrated to be $b = 0.6$ 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 in the matching function to be $\psi = 0.7$, as estimated by Shimer (2005). We use $\frac{1}{1-\sigma} = 3$ for the elasticity of substitution between the self-employment and wage sectors in the aggregate production function as in Feng et al. (2018). We set the on-the-job search intensity to be 0.4 following Faberman et al. (2017), who find that the average number of offers per month received by employed workers is around 40% of that for unemployed people in the US data. We set the elasticity of substitution between physical and human capital to be $\frac{1}{1-\epsilon} = 0.67$ (Krusell et al., 2000). Finally, according to the Penn World Table, the relative price of consumption to capital formation and annual capital depreciation rate were 1.1 and 0.04 respectively between 1994 and 2007 in the United States. Thus, we set the capital-specific technology for the United States to be $\chi = 1.1$ and the quarterly depreciation rate to be $\delta_k = 0.01$.

We calibrate two other parameters using other countries' data, given that there is no estimate for the United States. First, to generate nontrivial wage dispersion, we need firms' hiring costs to be convex in the amount of vacancies. There are relatively few estimates on the convexity in vacancy costs γ_v. Dix-Carneiro et al. (2019) find

γ_v ranges from 0.8 to 2.3 for Brazilian firms, whereas Blatter et al. (2016) find a relatively low convexity value (0.2) for Swiss firms. We use $\gamma_v = 1$ in our baseline calibration. Second, we calibrate losses in production hours per unit of training time to be $\delta_s = 0.7$, by taking the average from European countries' labor force surveys. We summarize the information on pre-assigned parameters in Table 2.

Table 2: Pre-Assigned Parameters

Parameter	Model	Source
ρ - Discount rate	0.01	Annualized interest rate of 0.04
T - Number of periods	160	40 years of work
η - on-the-job search intensity	0.4	Faberman et al. (2017)
b - Ratio of lowest wage to average wage	0.6	Hornstein et al. (2011)
ψ - Elasticity of matching function to searchers	0.7	Shimer (2005)
$\frac{1}{1-\sigma}$ - Elasticity of substitution	3	Feng et al. (2018)
$\frac{1}{1-\epsilon}$ - Elasticity between physical and human capital	0.67	Krusell et al. (2000)
χ - Capital-specific technology (US)	1.1	Penn World Table
δ_k - Capital depreciation rate	0.01	Penn World Table
γ_v - Convexity of vacancy costs	1	Dix-Carneiro et al. (2019) 0.8–2.3
δ_s - Loss in production hours per unit time of training	0.7	EU-LFS 2004 Training Module

Parameters to estimate. The remaining parameters to estimate are the constant in the matching function, c_M; 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 the function of leaving costs, c_p; the constant in training returns, ζ; the convexity in training returns, γ_s; the self-employment-sector share in the aggregate production function, γ; the convexity in the function of leaving costs, γ_p; the shape parameter of the Pareto productivity distribution, κ; exogenous separation rates, δ; the lower bound of leaving probability, $\underline{p}$; the share of training costs paid by firms, μ_F; the depreciation rate of human capital, d; and the human capital share in production, μ. Finally, since the relative ratio of A_T and A_M has the same effect as the self-employment-sector share γ in the aggregate production, we normalize the US aggregate productivity to be $A_M = A_T$ and choose A_M such that the output per worker is 1.

Targeted moments and fit. To calibrate those remaining parameters, we target the following moments listed in Table 3: average unemployment rates from 1994 to

2007 as computed by Hornstein et al. (2011); the ratio of the number of vacancies to the number of unemployed people from FRED for 2000 to 2007 (data available after 2000); the share of self-employment in total employment for 1994 to 2007 from the World Bank; the ratio of capital to annual real GDP for 1994 to 2007 from the Penn World Table; the Pareto parameter of firm employment distribution as estimated by Axtell (2001); workers' average wage growth after job-to-job transitions and the share of job-to-job transitions from high to low wage firms as computed by Haltiwanger et al. (2018); 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. We compute the last two moments using the 1995 Survey of Employer-provided Training implemented by the Bureau of Labor Statistics (BLS), which has both employers' and employees' information. We add the percent wage growth of 20 and 40 years' experience, as estimated by Lagakos et al. (2018b), to calibrate training returns. Finally, we add three more moments—job-to-job and job-to-unemployment probabilities and training intensity—which we explain next.

For job transition dynamics, we rely on two moments: the share of employed people remaining in the same firm and the share of employed people remaining employed after a quarter. We rely on data from Donovan et al. (2020), who provide these two probabilities for many countries. Their study is the first to delineate the relationship between these probabilities and development. Given that the purpose of our paper is to provide a comparison across countries at different development stages, we directly predict the two probabilities of interest from Donovan et al. (2020) using the per-capita GDP of the United States. Although these predicted values are a little higher than actual US values, we choose to use the predicted values in order to be consistent with our calibration in the second step for representative economies at different income levels.

Finally, it is important to note that the available data does not provide a direct measure of overall firm-provided training for all countries. For instance, we do not have measures of informal training for most of the economies we consider in this paper. Thus, we first take the average hours of formal training per worker from the data.[21] We then impute overall training intensity for every economy, relying on

[21] We multiply shares of workers exposed to formal training by hours spent on formal training per participant, which are predicted using the relationship between hours of formal training per participant and GDP per capita from the EU-CVT data.

Table 3: Moments in the Model vs Data

Moments	Data	Model
1. Moments: labor market		
1.1 Unemployment rate (%)	6.5	6.1
1.2 Ratio of #Vacancies to #Unemployed	0.55	0.62
1.3 Self-employment sector employment share (%)	6.0	5.5
1.4 Pareto parameter of firm size distribution	1.06	1.10
1.5 % workers remaining in same firm after one quarter	0.95	0.95
1.6 % workers remaining employed after one quarter	0.97	0.97
1.7 Workers' avg wage growth after job-to-job transition	0.13	0.13
1.8 % job-to-job transition from high to low wage firms	0.22	0.21
2. Moments: training intensity and value		
2.1 Average training intensity (% time)	2.2	2.3
2.2 Ratio of training intensity in firms with 100–499 employees to that in firms with 50–99 employees	1.2	1.2
2.3 Ratio of training costs to wage costs of training	0.24	0.26
2.4 Percent wage increase of 20 years' experience (%)	88	86
2.5 Percent wage increase of 40 years' experience (%)	89	89
3. Other moments		
3.1 Ratio of capital to annual real income	3.3	3.3
3.2 Output per worker (normalization)	1	1

The table reports the targeted moments in the data and in the model.

two assumptions according to the Survey of Employer-provided Training (US-SEPT) survey: the average worker spends two hours in informal training for each hour spent on formal training and there are 50% more workers participating in informal training than in formal training.[22]

Table 3 shows that the model almost exactly matches all the moments related to training. Moreover, the model almost exactly or very closely matches all the moments reflecting labor market dynamics.

Calibrated Parameters. We report the calibrated parameters in Table 4. Our parameters are reasonable compared with the literature. Our parameter γ_s can be interpreted as the diminishing returns of human capital investments (in terms of effective hours) in producing new human capital. Its calibrated value $\gamma_s = 0.28$ is in the ballpark of the estimates in the literature. For instance, Imai and Keane (2004)

[22]In the United States, 60% of workers receive formal training and 90% receive informal training.

find this parameter to be 0.22, while Manuelli and Seshadri (2014) estimate it to be 0.48. Moreover, training a young worker for the full quarter (480 working hours) increases her hourly wage by 7%, which lies within the range of empirical studies on US training returns as reviewed by Leuven (2004) and Bassanini et al. (2005).[23] Our calibrated quarterly depreciation rate of human capital from training $d = 0.02$ is similar to the annual depreciation rate of 0.06–0.08 of training returns estimated by Blundell et al. (2021) using British labor surveys. For more evidence on the calibrated parameters and model dynamics, please see Appendix Section F, where we show how the different moments help identify the model's parameters by calculating the elasticity of the moments to each parameter.

Table 4: Calibrated Parameters Values

c_M	c_s	c_v	γ_s	γ	γ_p	κ	ζ	$\underline{p}$	μ_F	δ	c_p	d	μ	A_M
0.50	0.21	0.79	0.28	0.45	8.12	6.21	0.07	0.11	0.94	0.02	2.53	0.02	0.40	0.26

Notes: This table lists the parameters that were determined using the method of moments and their values in the quantitative analysis.

4.1.2 Cross-Country Calibration

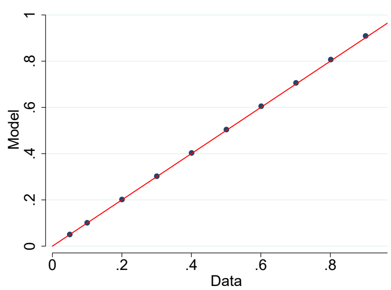
In this section, we calibrate the model for representative economies at 10 different income levels aside from that of the United States. To do this, we re-calibrate a few parameters associated with our mechanisms of interest, namely self-employment, job turnover (captured jointly by job-to-job transitions and job destruction rates), and physical capital endowments. The focus on these mechanisms is motivated by our empirical evidence, the literature, and particularly the fact that the size and nature of these channels radically change with development.[24] We keep most of the baseline parameters at the US levels and re-calibrate δ, c_p, A_m, A_T, and χ to match income

[23] For example, using the National Longitudinal Surveys (NLSY), Veum (1995) finds that increasing one hour of formal training improves hourly wages by 0.01%. Also using NLSY data, Frazis and Loewenstein (2005) find that 60 hours of formal training increases wages by 3% to 5%. Our calibration implies 4% wage growth for 60 hours of training in one quarter. The comparison with Veum (1995) and Frazis and Loewenstein (2005) is imperfect, however, because it is unclear whether training in their data happened within one quarter or in multiple quarters.

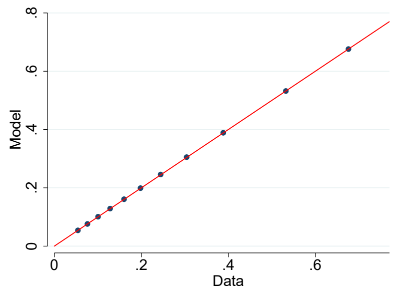
[24] Gollin (2002, 2008) find that developing countries exhibit higher shares of self-employment, while Donovan et al. (2020) find that job turnover rates are higher in these economies. The Penn World Table indicates that the capital-to-output ratio increases with development (Inklaar et al., 2019), which could facilitate human capital accumulation.

levels, self-employment, the share of workers who stay in the same firm from quarter to quarter, the share of workers who stay employed from quarter to quarter, and the capital-to-output ratios in each representative economy.[25]

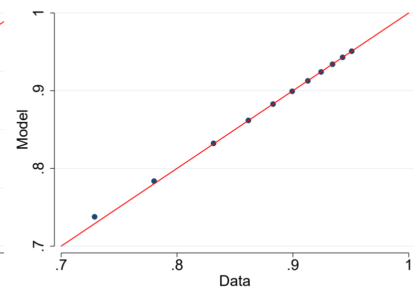
Figure 8: Cross Country Targeted Moments



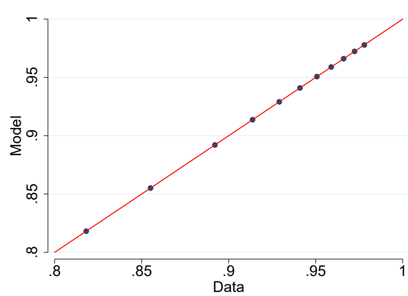
(a) GDP per capita



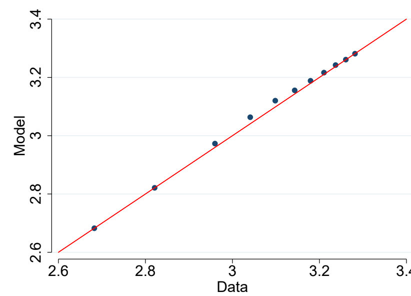
(b) Self-employment Share



(c) % Emp in Same Firm



(d) % Emp still Employed



(e) Capital/Output Ratio

Notes: This figure shows the targeted moments in the model (vertical axis) and in the data (horizontal axis). We consider 10 representative economies at income levels of \$2,500, \$5,000, \$10,000, \$15,000, \$20,000, \$25,000, \$30,000, \$35,000, \$40,000, and \$45,000 for GDP per capita (\$50,000 is the US level).

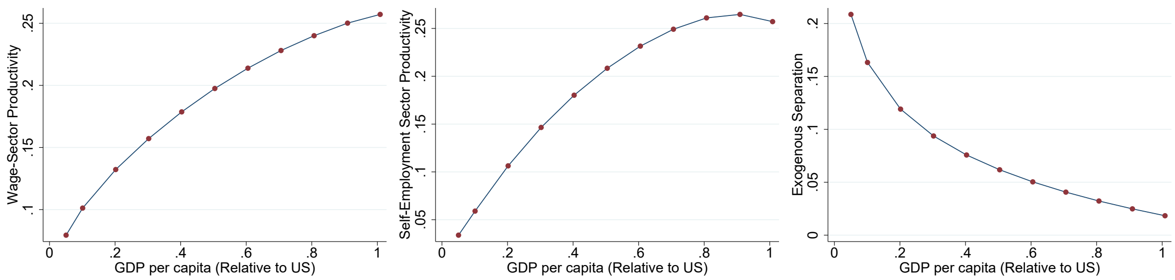
We first show how the model fits the targeted moments in each of our representative economies in Figure 8. The x-axis captures the empirical estimates of each moment while the y-axis captures the model estimates of each moment. We also plot the 45-degree line to aid the comparison between the model and data. Overall, our model matches the targeted data moments very well.

In Figure 9 we show how the values of the re-calibrated parameters change across our representative economies. In Panel (a) we plot the level of productivity in the wage sector and in Panel (b) we plot the level of productivity in the self-employment sector. As expected, both productivity levels increase with GDP per capita. In

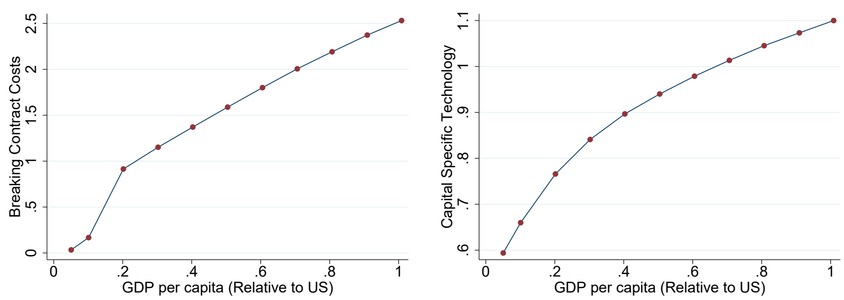
[25]According to the Penn World Table, the slope of capital-to-output ratios on log GDP per capita is 0.2 for the period 1994–2007. We use this slope to compute the capital-to-output ratio relative to the US level for each representative economy.

Panels (c) and (d) we plot the parameters shaping the labor market dynamics. The job destruction rate δ decreases with income, while the cost of breaking the contract c_p increases with income. Finally, in Panel (e) we plot the physical-capital-specific productivity, which increases with income and thus generates larger capital-output ratios in rich economies.

Figure 9: Cross Country Calibrated Parameters



(a) Wage Sector TFP **(b)** Self-Emp Sector TFP **(c)** Exogenous Separation



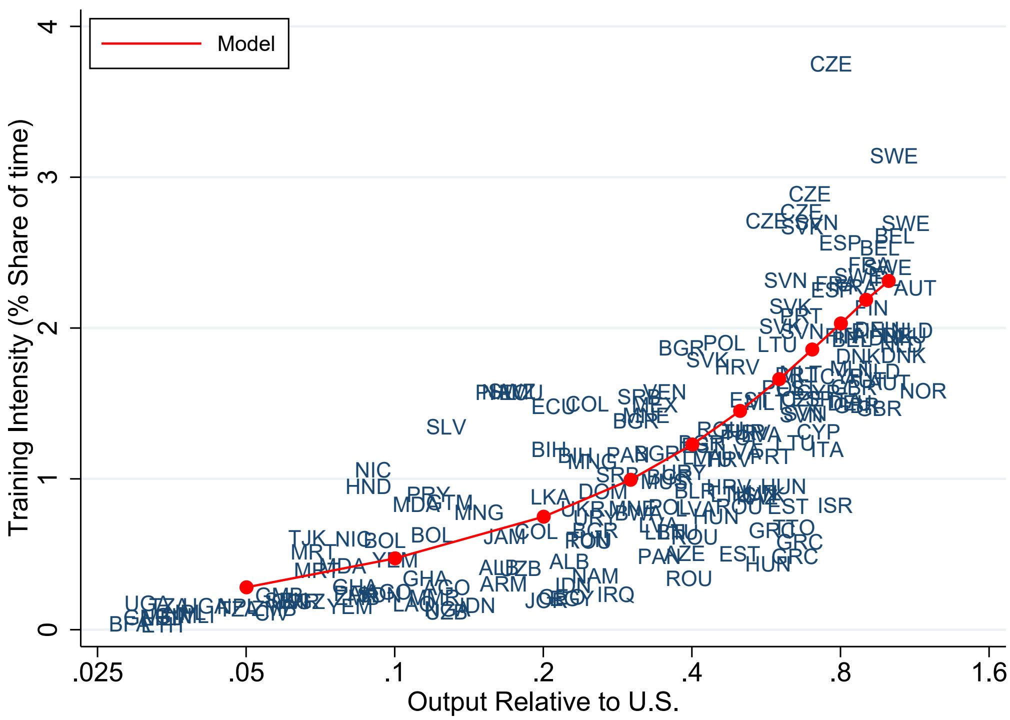
(d) Break Contract Costs **(e)** Capital Productivity

Notes: This figure shows how the values of the re-calibrated parameters change across our representative economies. We consider 10 representative economies at income levels of \$2,500, \$5,000, \$10,000, \$15,000, \$20,000, \$25,000, \$30,000, \$35,000, \$40,000, and \$45,000 for GDP per capita (\$50,000 is the US level). Panel (a) shows the wage sector productivity, A_M. Panel (b) shows the productivity in the self-employment sector, A_T. Panel (c) shows the exogenous separation rate, δ. Panel (d) shows the breaking contract costs, c_p. Panel (e) shows the physical capital productivity, χ.

4.1.3 Non-targeted Moments and Model Validation

We first turn our attention to the main non-targeted moments we want to analyze: training levels in our representative economies and the elasticity of training with respect to income. We plot the training intensity from the data and model as a function of GDP per capita in Figure 10. The model matches the levels and elasticity of training data with respect to per-capita GDP well.[26]

Figure 10: Training in Data and Model



Notes: This graph shows the cross-country training intensity (measured in the share of time that an average worker spends in training) as a function of US output. The red line shows the outcome of the model and in blue we plot each country-year observation from the data.

Now, in Table 5, we compare several untargeted moments in the model to the data for the United States and across countries. We first focus on the US moments in Panel 1. Our model implies a similar aggregate labor share as in the BLS data for the period 1994–2007. The slope of the labor share with respect to firm market shares is within the range of estimates of Autor et al. (2020), which captures how concentration affects wage returns relative to firm revenues.

[26]These results do not imply that our channels explain all training differences across countries, but rather that we capture how training varies with income. From the model, we can see that factors that affect separation rates, the probability of hiring, firm-worker matching quality, or the vacancy costs will affect the contracts and training investments. In Appendix Section H we rely on labor market institutions indexes constructed by Botero et al. (2004) to understand how the cost of firing workers, labor market institutions such as the minimum wage, and unemployment benefits correlate with our measure of training. We find that these measures increase the explanatory power over training, but they only account for a small part of the trend component between training and income.

Table 5: **Non-Targeted Moments in the Model vs Data**

Untargeted Moments	Data	Model
(1) The US		
1.1 Employees' Labor Share (%)	55	48
1.2 Slope of labor share on firm market share	$[-2.37, -0.35]$	-0.94
(2) Across countries		
2.1 Slope of labor shares on log GDPPC (adj 1)	-0.02	-0.02
2.2 Slope of labor shares on log GDPPC (adj 2)	0.02	0.07
2.3 Slope of std firm size on log GDPPC	0.18	0.13
2.4 Slope of log relative price of capital formation to consumption on log GDPPC	-0.15	-0.23

Notes: This table reports some non-targeted moments in the data and in the model. The aggregate labor share is calculated using BLS data for the period 1994–2007. The slope of the labor share on firm market shares comes from Autor et al. (2020). The measures of the labor share and income to calculate moments 2.1 and 2.2 come from Gollin (2002). The first measure (adjustment 1) assumes the self-employment sector labor share is 1 while the second measure (adjustment 2) assumes that labor share in the self-employment sector is identical to its counterpart in the wage sector. The slope of the standard deviation of employment with respect to income comes from Poschke (2018). We use the Penn World Table to compute the slope of log relative price of capital formation to consumption on log GDP per capita in the period 1994–2007.

Panel 2 focuses on cross-country non-targeted moments. First, we compare the relationship between different measures of the labor share and income from Gollin (2002). The first measure (adjustment 1) assumes that the labor share in the self-employment sector is equal to 1, while the second measure (adjustment 2) assumes that the labor share in the self-employment sector is identical to its counterpart in the wage sector. Our model captures that the labor share in the wage sector increases with income, whereas the large share of self-employment may induce a high labor share in poor countries if the labor share for self-employment is 1. Second, we compare the correlation between firm size and income in our model and the data. The literature shows that rich countries have larger firms and a firm size distribution that is skewed to the right, and that such concentration of production impacts aggregate productivity. Because it is easier for productive firms to accumulate labor in rich countries due to fewer exogenous job separations, our model also generates an increase in the number of large firms with income. We provide one informative moment of the firm size distribution, the slope of the standard deviation of firm-level employment with respect to income, which is 0.13 in our model and thus slightly lower than the slope of 0.18 found by Poschke (2018). Third, we check our modeling of the capital-specific technology by comparing how the change in the relative price of capital to consumption with respect to income differs between our model and the data. In the model, the price of capital formation relative to consumption is mainly driven by the inverse of capital-specific technology $1/\chi$.[27] We find that both the model and the data predict a decline in the relative price of capital to consumption with respect to income levels, yielding quantitatively similar elasticities.

5 Wage Growth, Training, and Income Differences

In the following section we aim to answer three main questions: (1) What portion of wage growth differences across countries can on-the-job training account for? (2) Why do developed economies invest more in training?; and (3) What portion of income differences across countries can on-the-job training account for?

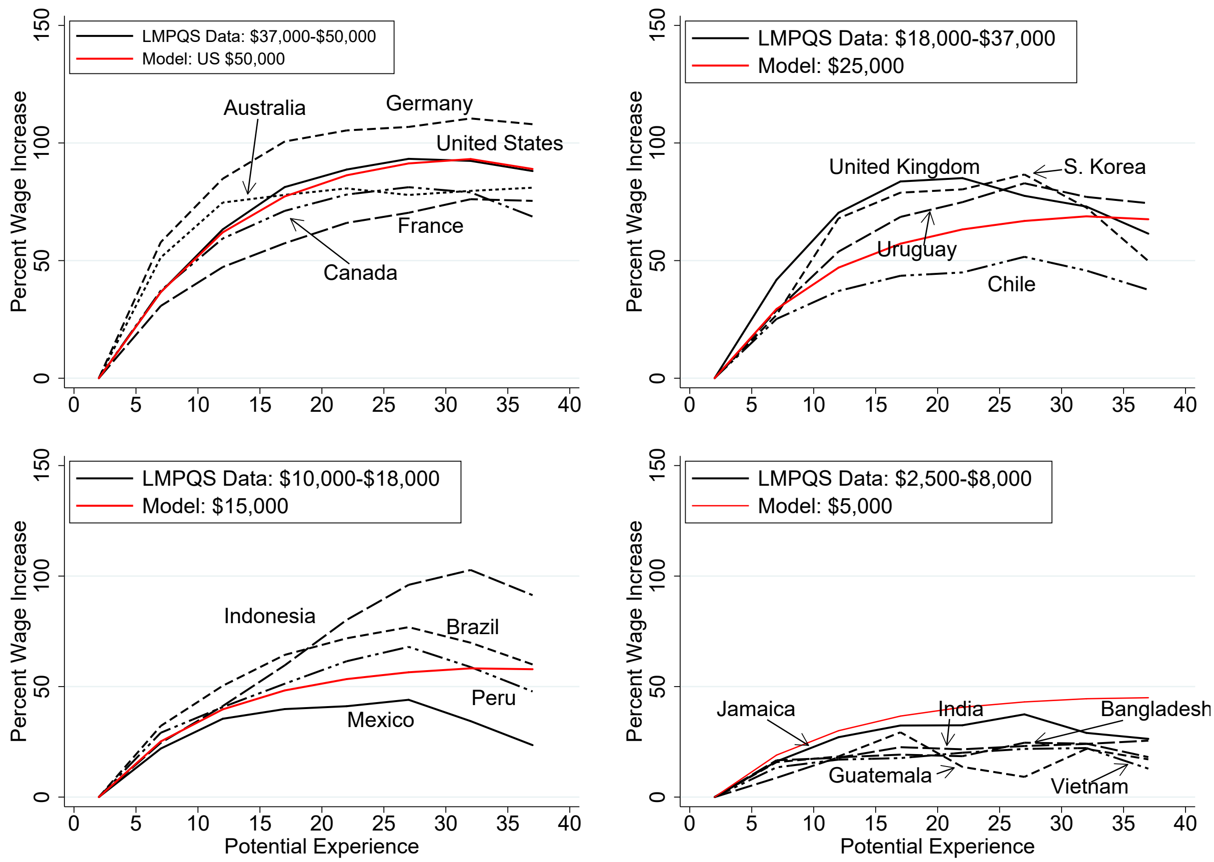
5.1 Cross-Country Wage Growth Differences

We first analyze how much our model, and specifically training, contribute to explaining the differences in workers' lifecycle wage growth between developed and developing economies. Figure 11 plots the experience-wage profiles of the 18 economies studied in Lagakos et al. (2018b) (LMPQS henceforth), which span all income levels, and the corresponding model predictions. Each panel shows the profiles from countries within a particular income range and the model's profile for an economy within that same range. Our model matches the wage growth profiles well at all income levels except for those at the bottom of the world income distribution. The calibrated economy at \$5,000 of GDP per capita has a steeper experience-wage profile than its counterparts in the data. This suggests that other factors that we do not include in our model may play an important role in explaining the low wage growth in these low-income economies.

We now turn our attention to quantifying what portion of the cross-country difference in returns to experience our model can account for. To this end, in Figure 12 we plot the cross-country returns to 20 years of experience found by LMPQS and

[27]The price of capital formation relative to consumption is given by $1/(\chi P)$ in the model, where P is the price of the consumption good. We find that $1/\chi$ strongly declines with income and mainly drives the negative relationship between the price of capital formation relative to consumption and income.

Figure 11: Cross-country Experience-Wage Profiles: LMPQS vs Model



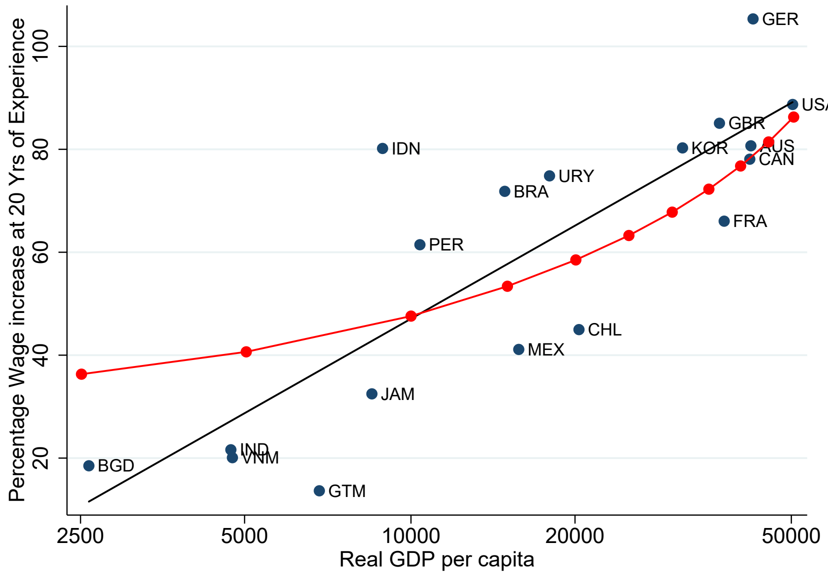
Notes: This figure replicates Figure 2 from Lagakos et al. (2018b) (LMPQS) and adds the wage-experience profiles from our model in red. Along the y-axis we plot the percent increase in wages at each potential experience bin, and along the x-axis we plot potential experience in years.

the corresponding model predictions as a function of per-capita GDP. As expected, the model matches the wage growth for middle- and high-income countries very well, and overestimates the wage growth for workers in the poorest economies. We then regress these returns on log per-capita GDP, and find a slope of 0.26 in LMPQS and a slope of 0.16 in our model. This implies that our model captures 61% of the cross-country differences in returns to experience.[28] LMPQS find that occupation and schooling differences capture around 20% and 30% of the differences in wage growth across countries, respectively. This suggests that our channel captures most of the cross-country differences in wage growth that are not explained by these two factors.

We then decompose the wage growth predicted by our model across all income

[28]Nonetheless, the model captures all of the difference for the economies above $10,000.

Figure 12: Cross-Country Experience-Wage Profiles: LMPQS vs Model



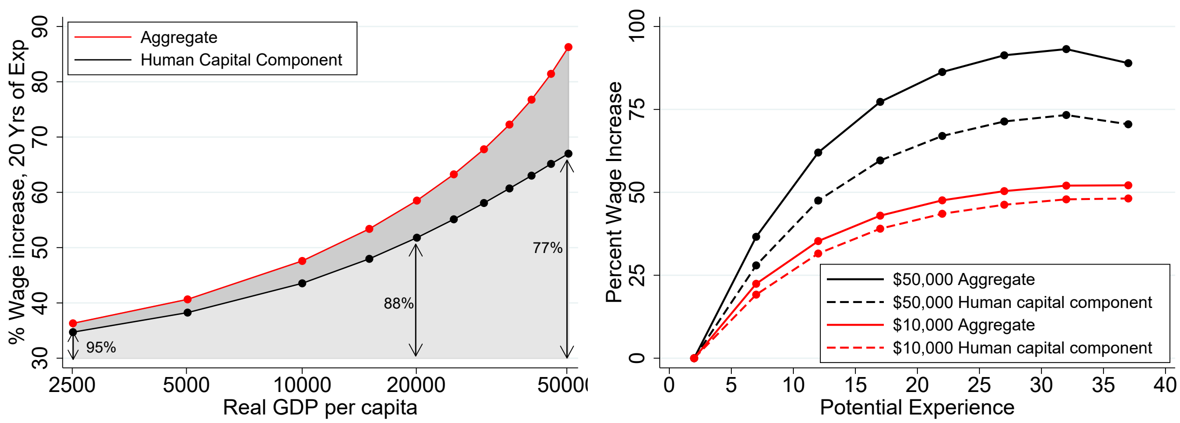
Notes: This figure replicates Figure 3 from Lagakos et al. (2018b) (LMPQS) and adds the returns to experience from our model in red. The slope in the LMPQS data is 26 while the slope of regressing the model's returns on log per-capita GDP is 16.

levels into human capital (or training) and job turnover components in order to quantify their relative importance.[29] In Panel (a) of Figure 13, we present the model-predicted returns to 20 years of experience as a function of per-capita GDP in the aggregate model and with just the human capital component. We find that the contribution of human capital to wage growth is large for every economy, though it decreases slightly with income. This stems from the high level of job destruction prevalent in the poorest economies, which prevents workers from climbing up the job ladder. As income increases, fewer workers are separated from their jobs and become unemployed, which generates larger increases in wages over the life cycle through job-to-job transitions. In Panel (b) we plot this decomposition for the full model-predicted wage-experience profiles for two economies with GDP per capita levels of $50,000 and $10,000, and find a similar pattern. We find that the human capital

[29]There is a growing literature that focuses on distinguishing the degree to which human capital accumulation and job search contribute to explaining workers' earnings dynamics in developed economies (e.g., Bunzel et al., 1999; Rubinstein and Weiss, 2006; Barlevy, 2008; Yamaguchi, 2010; Burdett et al., 2011; Bowlus and Liu, 2013; Bagger et al., 2014; Gregory, 2021). We contribute to this literature by calculating what portion of wage growth is driven by firm-provided training and labor market dynamics at all income levels.

accumulation component explains 70% of the differences in workers' wage growth between these two economies, while job turnover explains the remaining 30%.[30]

Figure 13: Cross-country Experience-Wage Profiles Composition



(a) Decomposition and Income

(b) Profile Decomposition

Notes: This figure shows the decomposition of model-predicted wage growth into human capital (or training) and job turnover components. Panel (a) plots the returns to 20 years of experience as a function of per-capita GDP, and Panel(b) plots wage-experience profiles. Wage growth stemming from human capital is calculated using the average increase in human capital for workers at each level of potential experience. The residual wage growth stems from job turnover.

5.2 Training Decomposition

In this section, we analyze how much of cross-country training differences can be explained by each of our channels. We aim to understand what drives the lack of training in developing economies and the role played by each channel at different stages of development.

We first perform a sectoral accounting analysis that seeks to understand how the sectoral allocation of employment between the self-employment and wage sectors shapes training differences in the aggregate. Denoting S as the training investment in the wage sector and M as the wage-sector employment share, the difference in training

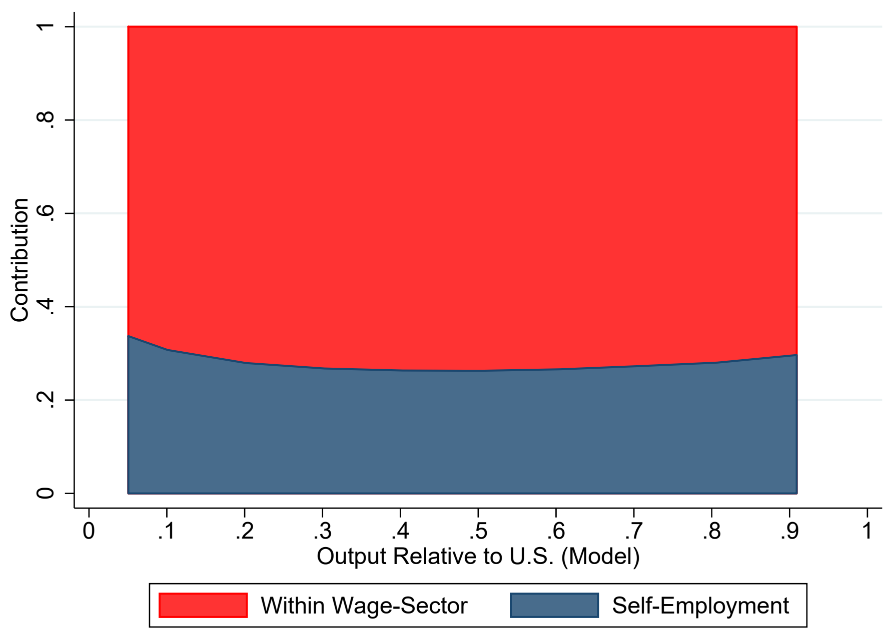
[30] We calculate these numbers by (1) subtracting the 20-year wage increase in the economy with GDP per capita level of $10,000 from that of the country with the $50,000 GDP per capita level in both the full and human capital component models and then (2) taking the ratio between these two subtractions to obtain how much of the differences between these two economies' rates of wage growth human capital can explain.

between the baseline economy and the United States can be simply decomposed as

$$log(S_{US}M_{US}/S_{base}M_{base}) = log(S_{US}M_{US}/S_{US}M_{base}) + log(S_{US}M_{base}/S_{base}M_{base}).$$

The first term reflects the training increase due to the change in the share of self-employment in total employment, while the second term represents the increase in training in the wage sector, conditional on the sectoral allocation. We plot these two components in Figure 14 and find that around 35% of cross-country training differences are explained by differences in the share of self-employment in aggregate employment. We also find that the importance of self-employment slightly decreases with income, in line with our finding that the poorest economies have very high self-employment shares and thus few workers exposed to training.

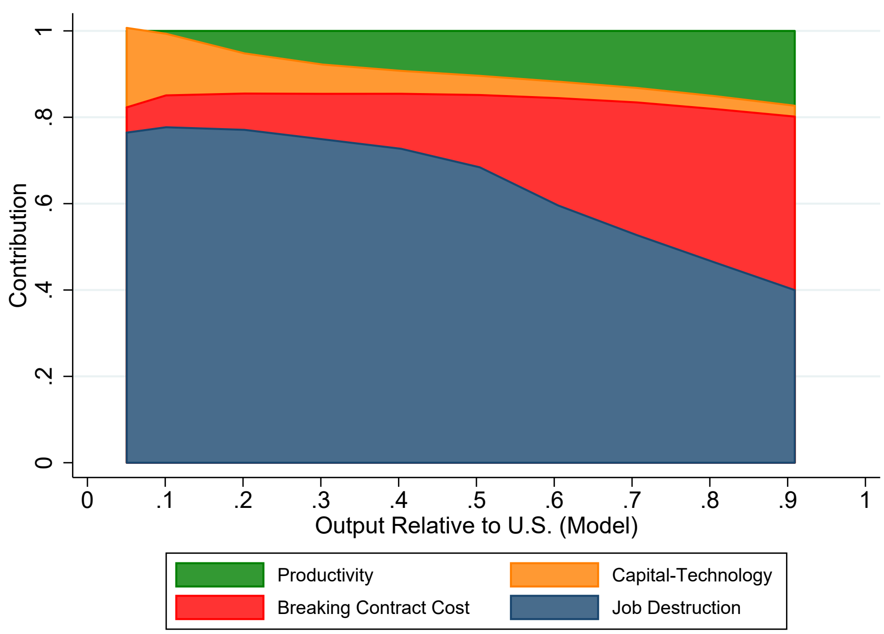
Figure 14: Training Decomposition by Sectoral Component



Notes: This figure shows how (1) changing the self-employment share while keeping training in the wage sector fixed (blue) and (2) changing the wage sector training level while keeping the self-employment share fixed (red) contribute to explaining the difference in training between each economy and the United States.

We now explore what portion of the differences in training investments across countries can be explained by differences in labor market frictions, physical capital productivity, and sectoral productivity. To do this, we perform a factor decomposition analysis where we change the values of the parameters governing these channels one at a time. We specifically focus on the five parameters that vary across countries: δ, which shapes job destruction; c_p, which shapes job-to-job transitions; χ, which shapes physical capital intensity; and A_T and A_M, which denote self-employment and wage-sector productivity levels, respectively, and shape income and self-employment shares. For each economy, we simulate the model when changing the value of one of the country-specific parameters (c_p, δ, χ, and jointly A_T and A_M) to match its value in the US economy, and compute the respective change in training. Using this, we then calculate how much of the training gap between each economy and the United States is explained by each channel. We plot the results of this exercise in Figure 15.[31]

Figure 15: Share of Training Gap Covered by each Parameter Change



Notes: This figure depicts the contribution of each channel to explaining the training gap between the economies at each income level and the United States. The green area represents the contribution from changing A_M and A_T simultaneously, the orange area represents the contribution from changing χ, the red area represents the contribution from changing c_p, and the blue area represents the contribution from changing δ.

Most of the difference in training investments across countries is driven by differences in labor market frictions. Differences in the cost of breaking contracts and job destruction jointly explain around 80% of the training differences at all income levels.

[31]Since there may be interactions between the different channels, we normalize the contribution of each factor using the sum of the individual contributions.

The higher job separation rates prevalent in low- and medium-income economies that stem from job destruction and job-to-job transitions not only could lead to higher shares of self-employment, but also depress the incentives to invest in training in the wage sector. We also find that the contribution of each of these two components changes with income. In particular, the contribution of job destruction tends to decrease with income, whereas the importance of the cost of breaking contracts increases with income. Thus, for poor economies the most important channel in terms of explaining the lack of training is job destruction, but as income increases, the difference in training stems largely from frictions in job-to-job transitions.

We also find that differences in physical capital productivity and sectoral productivity levels jointly explain the remaining 20% of the training gap. Our results show that the difference in capital-specific technology is an important factor to explain the lack of training in the poorest economies. Its contribution decreases with income, rendering it almost irrelevant to explain differences in training between high-income economies. In contrast, we find that differences in sectoral TFP levels play a larger role in determining training differences as income increases.[32]

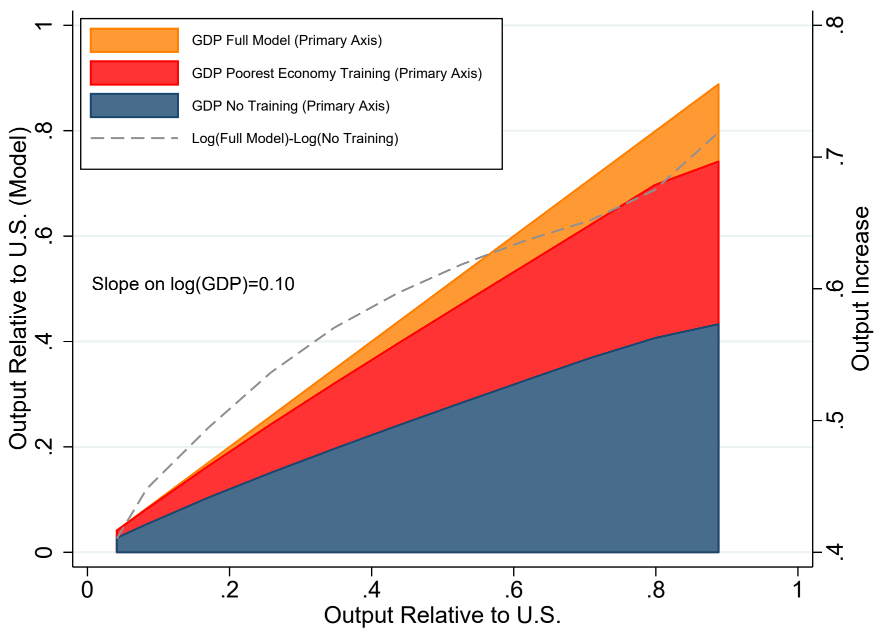
There are four main takeaways from the training decomposition analysis we perform. First, sectoral accounting suggests that around one-third of the training gap between poor and rich economies is explained by differences in self-employment shares. Second, labor market frictions are key to explaining training investments. High job separation rates and low contract quality make job turnover more likely and thus depress the incentives to invest in training in low- and medium-income economies. Third, when we decompose the contribution of these labor market frictions along its two key components, we find that job destruction is the main driver of the lack of training in the poorest economies, while differences in job-to-job transitions are more important in explaining the training differences between more-developed economies. Fourth, differences in physical capital are important in explaining the lack of training in low-income countries but not in medium-income countries.

[32]To further understand the model's dynamics, we go one step further and decompose the change in training stemming from each of the parameter changes mentioned above into partial and general equilibrium effects. The partial equilibrium effect captures the change in the firms' training level that occurs when parameter values change while the wage and employment distributions are kept fixed. The general equilibrium effect captures the change in training that occurs due to changes in the wage and employment distributions. Please see Appendix G for details.

5.3 Explaining Cross-Country Income Differences

We now focus on income differences explained by on-the-job learning. Using our calibrated representative economies, we simulate the model with different assumptions on training investments and plot the resulting per-capita GDP from each model along the primary y-axis of Figure 16. In orange we plot the original model, in blue we plot the model with no training, and in red we plot the case where all economies have the same training investments as the poorest economy.[33] Output is the lowest when there is no training. Output increases when we add the poorest economy level of training to the model with no training, and increases even more when we endogenize training, which reflects the fact that training boosts productivity in the aggregate. The heterogeneous increase in output with respect to income shows that adding training improves output more in developed economies than in developing economies.

Figure 16: Income Increase due to Training



Notes: This figure shows the model-predicted per-capita GDP for each model variation along the primary y-axis; and the percentage increase in output from training along the secondary y-axis. Both of these are plotted as functions of GDP per capita. The percentage increase in output from training is calculated as the log change in output from the model with no training (increasing c_s to an extremely large value) to the full model. The slope of 0.10 represents the share of the increase in GDP per capita explained by training in the model. Each observation comes from using the calibrated version of the model for each country. Data on GDP per capita comes from the Penn World Table.

[33] For this last case we use the training level for each firm and each age type of workers, and we assume that all economies have that exact same worker-firm training pattern.

Using this information, we now quantify the share of income differences across countries that can be explained directly by training in our model. To do this, we plot the difference between the log(per-capita GDP) in the full model and its counterpart in the model with no training along the secondary y-axis. This difference represents the percentage increase in output when we move from the model with no training to the full model. The slope of this percentage increase indicates the share of income differences explained directly by training in our model. Thus, our quantitative model suggests that on-the-job training explains 10% of income differences across countries. The contribution of on-the-job training to cross-income differences is thus sizeable, given that Lagakos et al. (2018a) show experience could explain around 20% of the income differences across countries.

6 Conclusion

Human capital accumulation plays a key role in economic growth and development. Recent research has highlighted the potential importance of on-the-job human capital accumulation in explaining workers' wage profiles. In this paper, we study one key source of on-the-job human capital accumulation: firm-provided training. We exploit rich enterprise- and worker-level data sources to show that firm-provided training increases with development and that this firm-provided training is the most important source of human capital investments in workers' careers. Then, we build a general equilibrium model with firm heterogeneity and training investments to shed light on the mechanisms giving rise to these facts.

Our results have several implications for understanding economic growth and conducting policy. First, our data and model suggest that self-employment is key to explaining the lack of on-the-job training in the poorest economies. Thus, our theory suggests that the reallocation of workers away from self-employment triggers human capital gains that add to the productivity gains identified by the literature and stemming from the movement towards higher productivity work. Second, we examine the evolving importance of different channels to explain the training gap at different stages of development, and find that the high level of job destruction is the most important factor preventing training investments in poor economies, while frictions in job-to-job transitions are more important in explaining training differences between developed economies. These results imply that in order to increase training and productivity, policies that improve the match quality between firms and workers may be desirable in developing economies, whereas policies that improve labor contracts may be more beneficial in richer countries.

Finally, it is important to note that the importance of on-the-job training could be even larger if this type of learning has complementarities with other sources of human capital, such as schooling or co-worker spillovers. A fruitful area for future research would be to study how different human capital accumulation sources interact with each other and how these interactions matter for policy making by countries at different stages of development.

References

Acemoglu, D. (1997). Training and Innovation in an Imperfect Labour Market. The Review of Economic Studies, 64(3):445–464.

Acemoglu, D. and Guerrieri, V. (2008). Capital Deepening and Nonbalanced Economic Growth. Journal of Political Economy, 116(3):467–498.

Acemoglu, D. and Pischke, J.-S. (1998). Why Do Firms Train? The Quarterly Journal of Economics, 113(1):79–119.

Acemoglu, D. and Pischke, J.-S. (1999). Beyond Becker: Training in Imperfect Labour Markets. The Economic Journal, 109:112–142.

Alon, T. (2017). Earning more by doing less: Human capital specialization and the college wage premium.

Alon, T. and Fershtman, D. (2019). Gradual specialization and the allocation of talent.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Van Reenen, J. (2020). The fall of the labor share and the rise of superstar firms. The Quarterly Journal of Economics, 135(2):645–709.

Axtell, R. L. (2001). Zipf Distribution of U.S. Firm Sizes. Science, 293(5536):1818–1820.

Bagger, J., Fontaine, F., Postel-Vinay, F., and Robin, J.-M. (2014). Tenure, experience, human capital, and wages: A tractable equilibrium search model of wage dynamics. American Economic Review, 104(6):1551–96.

Barlevy, G. (2008). Identification of search models using record statistics. The Review of Economic Studies, 75(1):29–64.

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

Becker, G. (1964). Human Capital. The University of Chicago Press, Chicago.

Bento, P. and Restuccia, D. (2017). Misallocation , Establishment Size , and Productivity Misallocation, Establishment Size , and Productivity. American Economic Journal: Macroeconomics, 9(3):267–303.

Blatter, M., Muehlemann, S., Schenker, S., and Wolter, S. C. (2016). Hiring costs for skilled workers and the supply of firm-provided training. Oxford Economic Papers, 68(1):238–257.

Blundell, R. W., Costa Dias, M., Goll, D., and Meghir, C. (2021). Wages, Experience and Training of Women. Journal of Labor Economics, 39(1):275–315.

Botero, J. C., Djankov, S., Porta, R. L., Lopez-de Silanes, F., and Shleifer, A. (2004). The regulation of labor. The Quarterly Journal of Economics, 119(4):1339–1382.

Bowlus, A. J. and Liu, H. (2013). The contributions of search and human capital to earnings growth over the life cycle. European Economic Review, 64:305–331.

Bunzel, H., Christensen, B. J., Kiefer, N. M., and Korsholm, L. (1999). Equilibrium search with human capital accumulation, volume 99. Centre for Labour Market and Social Research.

Burdett, K., Carrillo-Tudela, C., and Coles, M. G. (2011). Human capital accumulation and labor market equilibrium. International Economic Review, 52(3):657–677.

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

Caselli, F. (2005). Accounting for cross-country income differences. In Aghion, P. and Durlauf, S. N., editors, Handbook of Economic Growth, volume 1, chapter 9, pages 679–741. Elsevier.

De la Croix, D., Doepke, M., and Mokyr, J. (2018). Clans, guilds, and markets: Apprenticeship institutions and growth in the preindustrial economy. The Quarterly Journal of Economics, 133(1):1–70.

De Philippis, M. and Rossi, F. (2021). Parents, schools and human capital differences across countries. Journal of the European Economic Association, 19(2):1364–1406.

Dix-Carneiro, R., Goldberg, P., Meghir, C., and Ulyssea, G. (2019). Trade and Informality in the Presence of Labor Market Frictions and Regulations .

Doepke, M. and Gaetani, R. (2020). Why didn't the college premium rise everywhere? employment protection and on-the-job investment in skills. NBER Working Paper Series.

Donovan, K., Lu, J., and Schoellman, T. (2020). Labor Market Flows and Development.

Engbom, N. (2021). Labor market fluidity and human capital accumulation.

Erosa, A., Koreshkova, T., and Restuccia, D. (2010). How important is human capital? a quantitative theory assessment of world income inequality. The Review of Economic Studies, 77(4):1421–1449.

Faberman, R. J., Mueller, A. I., Sahin, A., and Topa, G. (2017). Job Search Behavior among the Employed and Non-Employed. NBER Working Paper Series.

Feng, Y., Lagakos, D., and Rauch, J. (2018). Unemployment and Development. NBER Working Paper Series.

Flinn, C., Gemici, A., and Laufer, S. (2017). Search, matching and training. Review of Economic Dynamics, 25:260–297.

Frazis, H. and Loewenstein, M. A. (2005). Reexamining the Returns to Training: Functional Form, Magnitude, and Interpretation. The Journal of Human Resources, 40(2):453–476.

Gollin, D. (2002). Getting income shares right. Journal of Political Economy, 110(2):458–474.

Gollin, D. (2008). Nobody's business but my own: Self-employment and small enterprise in economic development. Journal of Monetary Economics, 55(2):219–233.

Gregory, V. (2021). Firms as learning environments: Implications for earnings dynamics and job search.

Hall, R. E. and Jones, C. I. (1999). Why do some countries produce so much more output per worker than others? The Quarterly Journal of Economics, 114(1):83–116.

Haltiwanger, J., Hyatt, H., Kahn, L. B., and McEntarfer, E. (2018). Cyclical job ladders by firm size and firm wage. American Economic Journal: Macroeconomics, 10(2):52–85.

Hanushek, E. A. and Woessmann, L. (2012). Do better schools lead to more growth? cognitive skills, economic outcomes, and causation. Journal of Economic Growth, 17(4):267–321.

Hornstein, A., Krusell, P., and Violante, G. L. (2011). Frictional wage dispersion in search models: A quantitative assessment. American Economic Review, 101(7):2873–98.

Hsieh, C.-T., Hurst, E., Jones, C. I., and Klenow, P. J. (2019). The Allocation of Talent and US Economic Growth. Econometrica, 87(5):1439–1474.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing TFP in China and India. The Quarterly Journal of Economics, 124(4):1403–1448.

Hsieh, C.-T. and Klenow, P. J. (2010). Development accounting. American Economic Journal: Macroeconomics, 2(1):207–23.

Imai, S. and Keane, M. (2004). Intertemporal labor supply and human capital accumulation. International Economic Review, 45(2):601–641.

Inklaar, R., Woltjer, P., and Albarrán, D. G. (2019). The Composition of Capital and Cross-country Productivity Comparisons. International Productivity Monitor, Centre for the Study of Living Standards, 36:34–52.

Islam, A., Jedwab, R., Romer, P., and Pereira, D. (2019). Returns to experience and the misallocation of labor.

Jones, B. F. (2014). The human capital stock: a generalized approach. American Economic Review, 104(11):3752–77.

Klenow, P. J. and Rodriguez-Clare, A. (1997). The neoclassical revival in growth economics: Has it gone too far? NBER macroeconomics annual, 12:73–103.

Krusell, P., Ohanian, L. E., Ríos-Rull, J.-V., and Violante, G. L. (2000). Capital-skill complementarity and inequality: A macroeconomic analysis. Econometrica, 68(5):1029–1053.

Lagakos, D., Moll, B., Porzio, T., Qian, N., and Schoellman, T. (2018a). Life-cycle human capital accumulation across countries: lessons from us immigrants. Journal of Human Capital, 12(2):305–342.

Lagakos, D., Moll, B., Porzio, T., Qian, N., and Schoellman, T. (2018b). Life cycle wage growth across countries. Journal of Political Economy, 126(2):797–849.

Leuven, E. (2004). A review of the wage returns to private sector training.

Manuelli, R. E. and Seshadri, A. (2014). Human capital and the wealth of nations. American Economic Review, 104(9):2736–762.

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

Mortensen, D. and Pissarides, C. (1994). Job creation and job destruction in the theory of unemployment. The Review of Economic Studies, 61(3):397–415.

Pissarides, C. (2000). Equilibrium Unemployment Theory (2nd ed.). MIT Press.

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

Restuccia, D. and Rogerson, R. (2013). Misallocation and productivity. Review of Economic Dynamics, 16(1):1–10.

Rubinstein, Y. and Weiss, Y. (2006). Post schooling wage growth: Investment, search and learning. In Hanushek, E. and Welch, F., editors, Handbook of the Economics of Education, volume 1, chapter 1, pages 1–67. Elsevier.

Schoellman, T. (2012). Education quality and development accounting. The Review of Economic Studies, 79(1):388–417.

Schoellman, T. (2016). Early childhood human capital and development. American Economic Journal: Macroeconomics, 8(3):145–74.

Shimer, R. (2005). The Cyclical Behavior of Equilibrium Unemployment and Vacancies. American Economic Review, 95(1):25–49.

Veum, J. R. (1995). Sources of Training and Their Impact on Wages. Industrial and Labor Relations Review, 48(4):812–826.

Wasmer, E. (2006). General versus specific skills in labor markets with search frictions and firing costs. American Economic Review, 96(3):811–831.

Yamaguchi, S. (2010). Job search, bargaining, and wage dynamics. Journal of Labor Economics, 28(3):595–631.

Appendix

A Data Sources

For the main analysis, we rely on enterprise- and worker-level surveys in developed and developing economies. For developing countries, we use the World Bank Enterprise Survey (WB-ES). For developed countries, on the other hand, we rely on data from the European Union. Specifically, we use the European Union Labor Force Survey (EU-LFS), the Adult Education Survey (EU-AES), and the Continuing Vocational Training (EU-CVT) enterprise survey.

In addition, we rely on two secondary US data sources for empirical robustness checks, calibration, and model validation. First, we provide historical evidence on firm-provided training based on the National Household Education Surveys (NHES) Program, which consists of data on educational activities. Furthermore, to calibrate the model to the benchmark US economy, we rely on the 1995 US Survey of Employer-provided Training (US-SEPT), which was conducted during personal visits to more than 1,000 private establishments. Finally, we rely on data from Botero et al. (2004) to proxy labor market indicators for robustness checks, and data from Donovan et al. (2020) for measures of job destruction and job-to-job transitions to test cross-country correlations.

B Detailed Definitions of Sources of Education

We first carefully define training and its characteristics to ensure consistency across different data sources. We separate the sources of skill acquisition into four categories that allow for data comparability and also for meaningful economic interpretations through the lens of the model. We present a summary of these learning sources in Table B.1. The categories rank from the most structured and planned type of learning (schooling) to the least structured (informal learning). For expositional purposes, and because we focus on firm-sponsored investments, we also consider a secondary distinguishing quality within each source, which is the financing source for the educational investment (firm vs. worker sponsored). We present detailed information on each source of learning below.

Table B.1: Human Capital Sources and Examples

How Structured ↓		**Firm Sponsored**	Non-Firm Sponsored
	1. Schooling	MBA paid by firm	MBA self-financed
	2. Formal Training	Firm-organized presentation	Pre-employment training (license/certification)
	3. Informal Training	Guided o-t-j Training Job Rotation	-
	4. Informal Learning	-	Self-learning (e.g. Reading Journals)

Notes: Our definition of schooling reflects "Formal Education and Training," according to ISCED 2011, while both Formal Training and Informal Training are categories within "Non-Formal Education and Training" from ISCED 2011. The definitions of Formal and Informal Training follow the definitions in the WB-ES and EU-CVT. The different sources of human capital are ordered along two key features: (1) how structured the learning is and (2) the financing source for the educational investment (firm vs. worker sponsored).

Schooling: According to ISCED 2011, formal education and training is defined as "education that is institutionalized, intentional and planned through public organizations and recognized private bodies and in their totality constitutes the formal education system of a country. Formal education programs are thus recognized as such by the relevant national education authorities or equivalent authorities, e.g. any other institution in cooperation with the national or sub-national education authorities. Formal education consists mostly of initial education. Vocational education, special needs education and some parts of adult education are often recognized as being part of the formal education system."

Training: According to ISCED 2011, non-formal education and training is defined as "any organized and sustained learning activities outside the formal education system. Non-formal education is an addition, alternative and/or complement to formal education. Non-formal education may therefore take place both within and outside educational institutions and cater to people of all ages. Depending on national contexts, it may cover educational programs to impart adult literacy, life-skills, work-skills, and general culture. Note that within non-formal education we can have formal training or informal training depending on its level of organization."

We rely on definitions for *formal training* and *informal training* from the EU-CVT survey manuals. Continuing vocational training (*formal training*) refers to education or training activities that are planned in advance, organized, or supported with the specific goal of learning and financed at least partially by the enterprise. These activities aim to generate "the acquisition of new competences or the development and improvement of existing ones" for firms' employees. Persons currently engaging in an apprenticeship or training contract should not be considered as taking part in CVT. Random learning and initial vocational training are explicitly excluded and measured separately. These courses are typically separated from the active workplace (for example, they take place in a classroom or at a training institution), show a high degree of organization by a trainer, and the content is designed for a group of learners (e.g., a curriculum exists).

As defined by the EU-CVT survey, "Other forms of CVT" that we refer to as *informal training* are geared towards learning and are typically connected to the active work and the active workplace, but they can also include participation (instruction) in conferences, trade fairs, etc. 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 in the survey:

1. Guided-on-the job training: "It is characterised by planned periods of training, instruction or practical experience in the workplace using the normal tools of work, either at the immediate place of work or in the work situation. The training is organised (or initiated) by the employer. A tutor or instructor is present. It is an individual-based activity, i.e. it takes place in small groups only (up to five participants)."

2. Job rotation, exchanges, secondments, or study visits: "Job rotation within the enterprise and exchanges with other enterprises as well as secondments and study visits are other forms of CVT only if these measures are planned in advance with the primary intention of developing the skills of the workers involved. Transfers of workers from one job to another which are not part of a planned developmental programme should be excluded."

3. Learning or quality circles: "Learning circles are groups of persons employed who come together on a regular basis with the primary aim of learning more about the requirements of the work organisation, work procedures and workplaces. Quality circles are working groups, having the objective of solving production and workplace-based problems through discussion. They are counted

as other forms of CVT only if the primary aim of the persons employed who participate is learning."

4. Self-directed learning/e-learning: "Individual engages in a planned learning initiative where he or she manages the settings of the learning initiative/activity in terms of time schedule and location. Self-directed learning means planned individual learning activities using one or more learning media. Learning can take place in private, public or job-related settings. Self-directed learning might be arranged using open and distance learning methods, video/audio tapes, correspondence, computer based methods (including internet, e-learning) or by means of a Learning Resources Centre. It has to be part of a planned initiative. Simply surfing the internet in an unstructured way should be excluded. Self-directed learning in connection with CVT courses should not be included here."

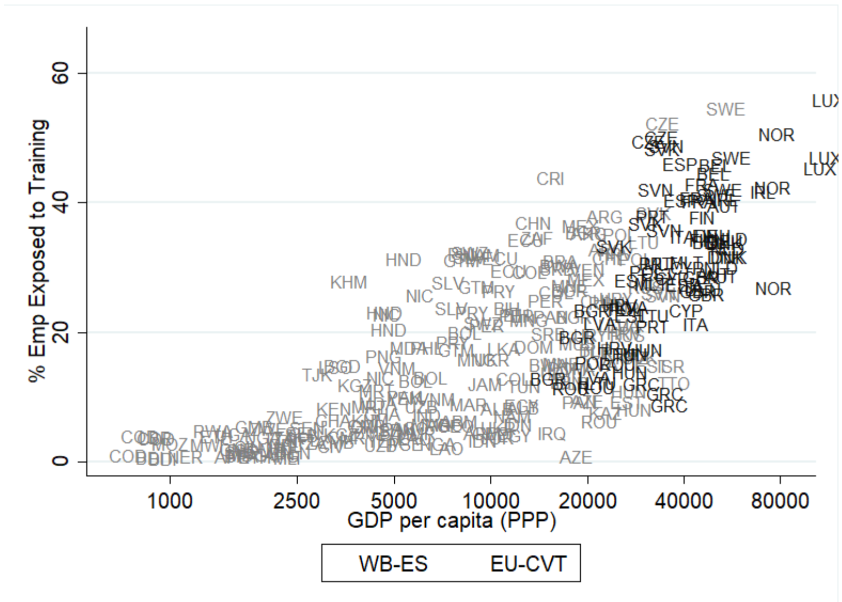
5. Participation in conferences, workshops, trade fairs, and lectures: "Participation (instruction received) in conferences, workshops, trade fairs and lectures are considered as training actions only when they are planned in advance and if the primary intention of the person employed for participating is training/learning."

Initial vocational training is defined as a formal education program (or a component thereof) where working time alternates between periods of education and training at the workplace and in educational institutions or training centers. This program consists of learning activities for workers who are new at their jobs.

Informal learning: Informal learning is defined as "intentional learning which is less organized and less structured than the previous types. It may include for example learning events (activities) that occur in the family, in the workplace, and in the daily life of every person, on a self-directed, family-directed or socially directed basis. Categories used for informal training are: learning from peers, colleagues; learning by using printed material, learning by using computers, learning through media (television, radio or videos); learning through guided tours as museums; learning by visiting learning centers as libraries."

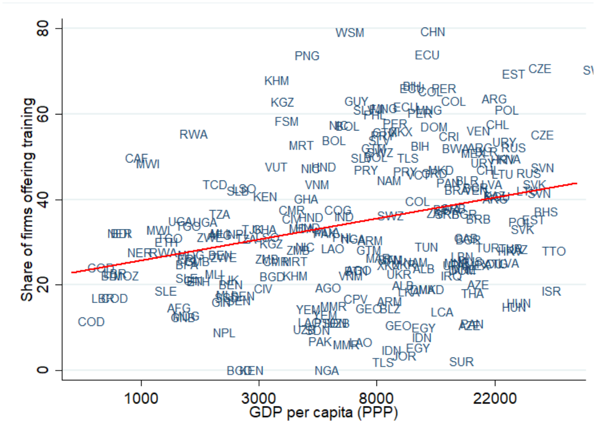
C Empirical Results: Additional Tables and Graphs

Figure C.1: Share of Formally Trained Employment and Development (Full Sample)

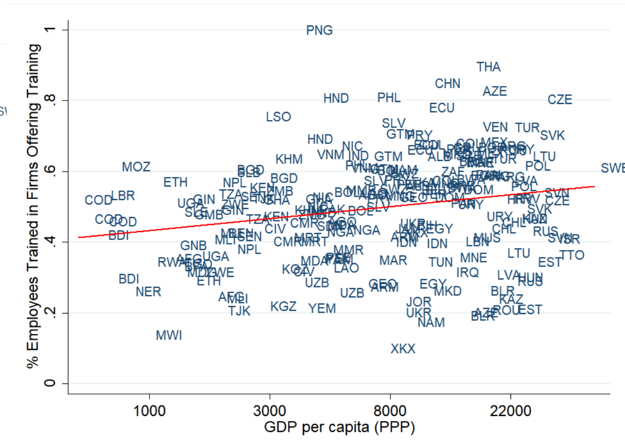


Notes: The share of formally trained employment follows from adjusting the share of workers who receive training from firms by the share of self-employment. Data on the share of employees trained within firms comes from the WB-ES for all developing economies and from the EU-CVT for European economies. Both surveys contain data on whether firms provided formal training in the previous fiscal year and the share of employees who participated. For the World Bank Enterprise Survey, we use the standardized wave with data from 2005–2017 for which we have firm weights and we plot all countries with no restrictions. Data on GDP per capita and self-employment comes from the Penn World tables and World Bank Indicators, respectively.

Figure C.2: Intensive and Extensive Margins

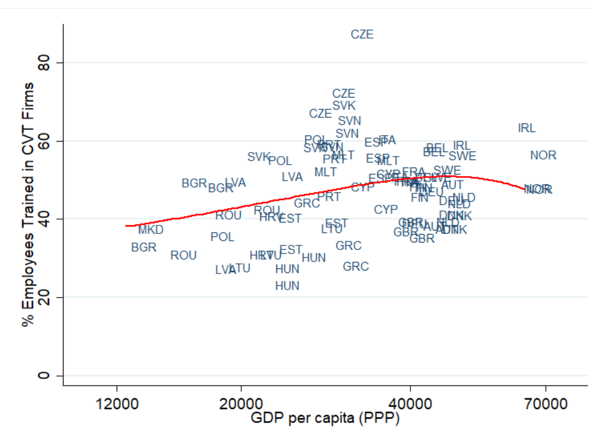


(a) %Firms Offering Training (WB-ES)



(b) %Trained Workers per Firm (WB-ES)

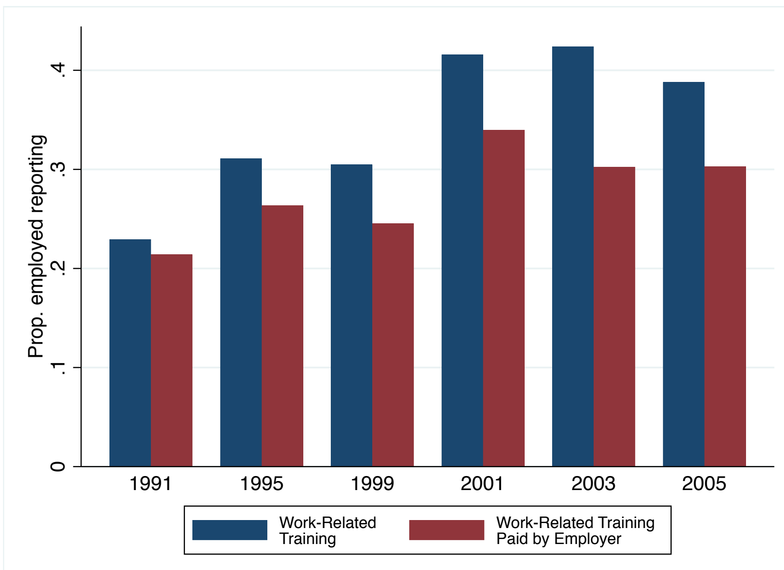
(c) %Firms Offering Training (EU-CVT)



(d) %Trained Workers per Firm (EU-CVT)

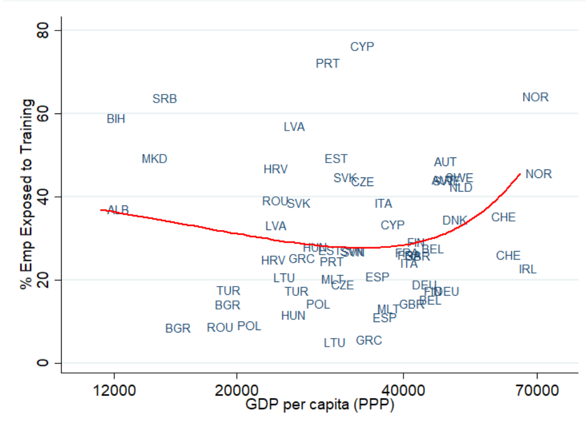
Notes: These figures show the intensive and extensive margins of training in the WB-ES and the EU-CVT. Panel (a) shows the share of firms offering training in the WB-ES and Panel (b) shows the share of participants per firm in the manufacturing and service sector weighted by the WB-ES-provided weights. Panels (c) and (d) show the counterparts for the EU-CVT. For the World Bank Enterprise Survey, we use the standardized wave with data from 2005–2017 for which we have firm weights and we plot all countries with no restrictions. Data on GDP per capita comes from the Penn World Table.

Figure C.3: Share of Workers Reporting Training by Year in the US

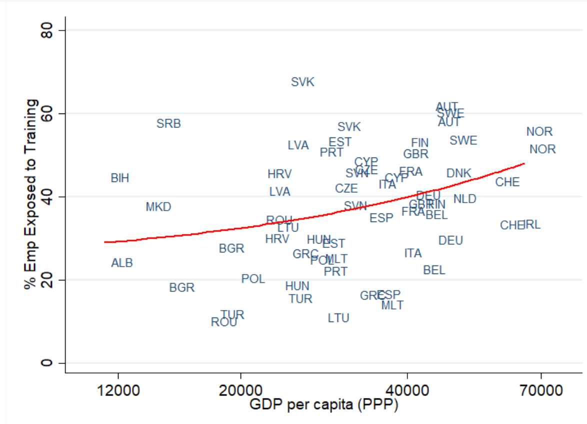


Notes: This figure shows US workers' participation rate in work-related training (training, workshops, seminars, courses, or classes for work-related reasons) in the past 12 months and work-related training sponsored by employer (training paid at least partially by employer). We use all years with data on these variables and exclude the 2016 survey from the analysis presented here due to definitional changes. Data comes from the National Household Education Survey (NHES).

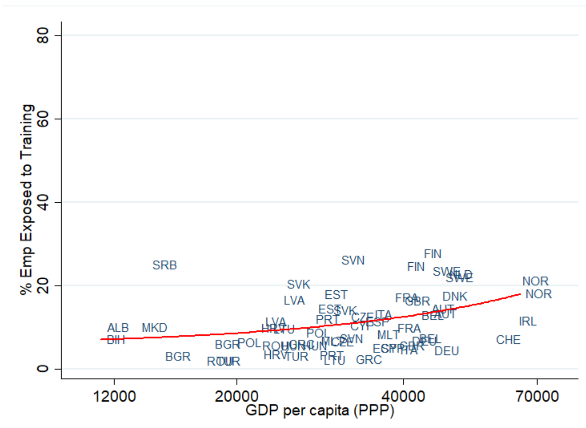
Figure C.4: Informal Learning



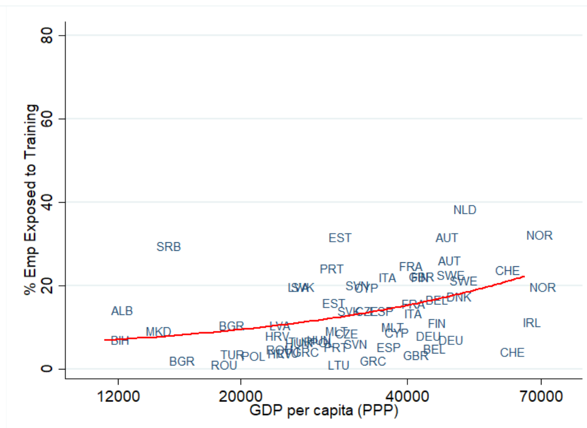
(a) Through Peers



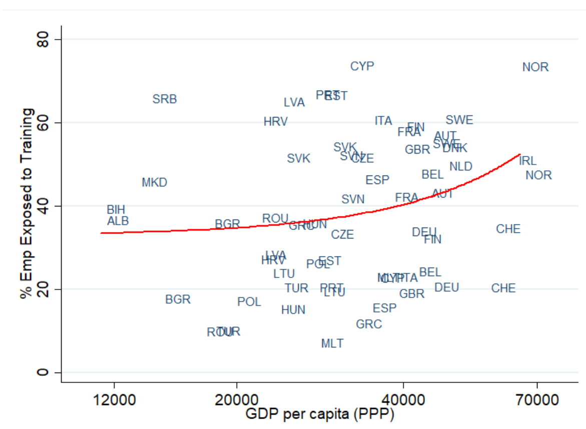
(b) Using Printed Material



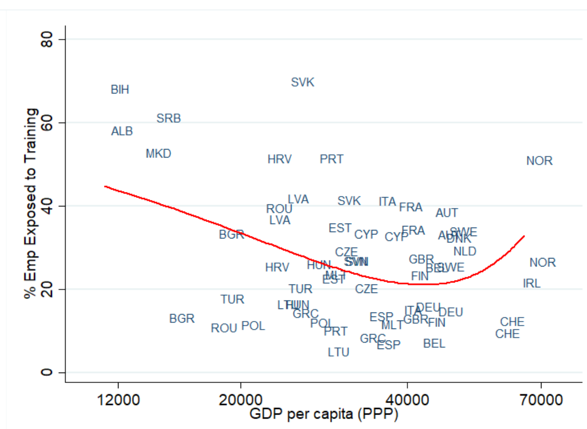
(c) Attending Learning Centers



(d) Tours on Relevant Sites



(e) Using Computer



(f) Using Media

Notes: These figures show participation rates in different types of informal learning: learning through peers (Panel (a)), using printed material (Panel (b)), attending learning centers (Panel (c)), tours of learning sites (Panel (d)), using computers (Panel (e)), and using media (Panel (f)). Data comes from the EU-AES. Data on GDP per capita comes from the Penn World Table.

Table C.1: European Union Labor Force Survey (EU-LFS)

	Hours	During Working Hours		Reason	
	Employed Population	During paid hours	Outside paid hours	Job related	Personal Social
European Union - 25 (2004–2006)	66	69.3	30.7	83.9	15.9
Germany	74			90.8	9.1
France	85	87.4	12.6	93.3	6.7
United Kingdom	35	70.3	29.7	79.2	20.8
Italy	58	56.5	43.5	83.5	15.1
Spain	102	38	62	61.7	38.3
Poland	40	59.4	40.6	91.3	8.7
Romania	80	34	66	80.3	19.7
Netherlands	76	54.8	45.2	86.1	13.9
Belgium	69	68.7	31.3	82.5	17.5
Greece	80	40.4	59.6	72.7	27.3

Notes: This table shows the portion of workers reporting different timing and reasons for their training. This data comes from Eurostat, past series, LFS ad Hoc Module 2003. We show the outcomes from the most populous European countries ranked in descending order.

Table C.2: Training Purpose (EU-CVT)

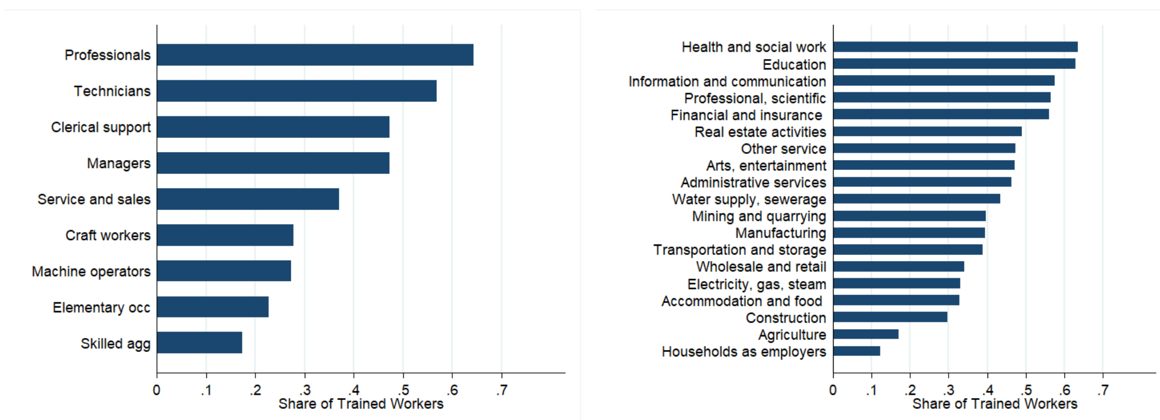
	Average By firm Size in 2010				Average By firm Size in 2015			
	All	10–49	50–249	250+	All	10–49	50–249	250+
General IT	27.3	23.7	34.5	54.7	12.8	12	15.2	15.8
Professional IT	16.9	14.5	21	37.5	10.2	9.8	11.6	11.2
Management	32	26.2	43.7	74.3	23.4	19.9	30.9	49.2
Team working	32.5	29	38.3	61.6	19.6	19.1	20.5	22.6
Customer handling	38.5	35.4	44.1	62.7	25.6	25	26.5	31.3
Problem solving	30.1	28.5	31.2	50	13.5	13.3	14.1	13.8
Office administration	26.9	24.3	32.3	45.1	13.4	13.6	14	8
Foreign language	15.3	11	24	46.9	7.9	5.9	13.2	17.5
Technical or job-specific	69	67.2	73.2	81.2	64.6	63.1	68.5	71.9
Oral or written communication	14.7	12.7	16.9	36.5	3.5	3.3	4	4.4
Numeracy and/or literacy	7	6.7	6.5	14.7	1.2	1.3	1.2	1.1
Other skills and competences	11	11.2	10.4	10.3	19.9	20.3	18	19.8

Notes: This table shows the share of enterprises providing CVT courses by type of skill targeted and firm size. A particular course may cover more than one category.[34]

D Training Composition and Firm Features

For the decomposition of the trend in on-the-job training using shift-share accounting, we split the occupations and industries into three bins (high, medium and low training) according to the aggregate level of training in the full sample as shown in Figure D.1. The occupation bins are Professionals, Technicians, and Managers (high training); Clerical Support, Service, and Sales, and Craft Workers (medium training); and Machine Operators, Elementary Occupations, and Skilled Agriculture (low training). For industry we use the ranking in Figure D.1: from Health and Social Work to Real State (high training), from Other Services to Manufacturing (medium training), and from Wholesale and Retail to Households as Employers (low training). For education and firm size we use the definitions in Figure D.2.

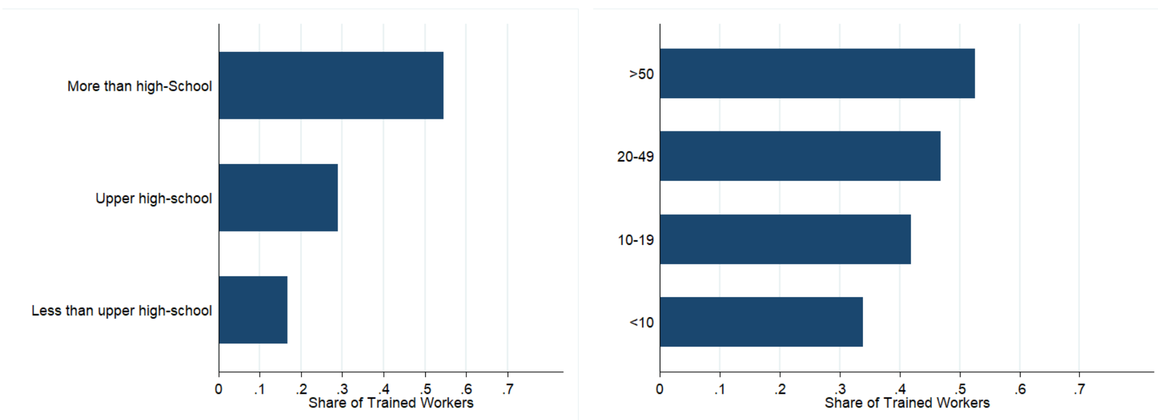
Figure D.1: Training from EU-AES by Industry and Occupation



(a) By Occupation **(b)** By Industry

Notes: These figures show employees' participation rate in training by occupation and industry categories. Data comes from the EU-AES.

Figure D.2: Training from EU-AES by Workers' Education and Firm Size



(a) By Education Level

(b) By Firm Size

Notes: These figures show employees' participation rate in training by education and firm size. Data comes from the EU-AES.

E Quantitative Model: Conditions for Simulations

Workers' Expected Utility. With linear utility and $\rho = r$, workers' utility is determined by the discounted income flows. Thus, workers' utility comes from (future) income flows produced with current human capital and potential income flows from human capital accumulation. For a worker of age a in a firm with productivity z, we denote $J_{c,a}(z)$ as the expected value of income flows per efficiency unit of human capital from training, and $J_{h,a}(z)$ as the expected value of income flows from basic skills that do not depreciate, as well as human capital investments. With a slight abuse of notation, we use $J_{c,a}(u)$ and $J_{h,a}(u)$ for unemployed workers. With $\theta = \frac{V}{\tilde{U}}$ denoted as market tightness as usual, 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 workers' lifetimes ($a = T$), workers have no incentive to accumulate human capital. Thus, we can obtain

$$J_{c,T}(z) = w(z);\ J_{h,T}(z) = w(z);\ J_{c,T}(u) = 0;\ J_{h,T}(u) = 0.$$

For younger workers ($a < T$), we can obtain their values by backward induction:

$$J_{c,a}(z) = w(z) + \frac{1-d}{1+\rho}\delta\left[\theta q(\theta)\int J_{c,a+1}(z')dF(w(z')) + (1-\theta q(\theta))J_{c,a+1}(u)\right]$$

$$+\frac{1-d}{1+\rho}(1-\delta)\left[J_{c,a+1}(z) + \eta\theta q(\theta)\int p_{a+1}(z,z')(J_{c,a+1}(z') - J_{c,a+1}(z)) - c_p^{\gamma_p}\frac{p_{a+1}(z,z')^{1+\gamma_p}}{1+\gamma_p}dF(w(z'))\right]$$

$$J_{h,a}(z) = w(z) + \frac{J_{c,a}(z) - w(z)}{1-d}\zeta s_a(z)^{\gamma_s} - \mu_W\left(c_s\bar{w} + \delta_s\tilde{r}(z)\right)s_a(z)$$

$$+\frac{\delta}{1+\rho}\left[\theta q(\theta)\int J_{h,a+1}(z')dF(w(z')) + (1-\theta q(\theta))J_{h,a+1}(u)\right]$$

$$+\frac{1-\delta}{1+\rho}\left[J_{h,a+1}(z) + \eta\theta q(\theta)\int p_{a+1}(z,z')(J_{h,a+1}(z') - J_{h,a+1}(z)) - c_p^{\gamma_p}\frac{p_{a+1}(z,z')^{1+\gamma_p}}{1+\gamma_p}dF(w(z'))\right]$$

$$J_{c,a}(u) = \frac{1-d}{1+\rho}\left[\theta q(\theta)\int J_{c,a+1}(z)dF(w(z)) + (1-\theta q(\theta))J_{c,a+1}(u)\right]$$

$$J_{h,a}(u) = \frac{1}{1+\rho}\left[\theta q(\theta)\int J_{h,a+1}(z)dF(w(z)) + (1-\theta q(\theta))J_{h,a+1}(u)\right]$$

$p_a(z, z')$ is the leaving probability conditional on getting an offer from a firm with productivity z', obtained by evaluating leaving probability for an average worker of age a in firm z:[35]

$$\max_{p \in [\underline{p}, 1]} \left[(J_{c,a}(z') - J_{c,a}(z))\bar{h}_a(z) + J_{h,a}(z') - J_{h,a}(z) \right] p - c_p^{\gamma_p} \frac{p^{1+\gamma_p}}{1+\gamma_p} (\bar{h}_a(z) + 1),$$

where $\bar{h}_a(z)$ is the average human capital from training of age a workers in firm z, as shown below.

Employment Distribution. Let N_m be the size of workers who enter the wage sector at each generation. Then, in the beginning of each period, the size of searchers in the wage sector is

$$\tilde{U} = \sum_{a=1}^{T} (u_a + (N_m - u_a)\eta),$$

which is the sum of the unemployed and on-the-job searchers across different age groups. The unemployed population (before job search and matching) for the youngest cohort is $u_1 = N_M$ and proceeds as $u_{a+1} = \delta N_M + (1 - \theta q(\theta))(1-\delta)u_a \ \forall\ 1 \leq a \leq T-1$.

We define the measure of employment $m_a(z)$ for workers of age a in firms with productivity z. Hence, the employment distribution across firms for the youngest cohort is simply $m_1(z) = \frac{\theta q(\theta) f(w(z)) w'(z)}{g(z)} u_1$ after search and matching processes. For older cohorts, their measure of employment is given by

$$m_{a+1}(z) = \underbrace{(1-\delta)\left[1 - \eta\theta q(\theta)\int p_{a+1}(z, z')dF(w(z'))\right] m_a(z)}_{\text{stayers}}$$
$$+ \underbrace{u_{a+1}\frac{\theta q(\theta) f(w(z)) w'(z)}{g(z)}}_{\text{hires from unemployed}} + \underbrace{(1-\delta)\eta \frac{\theta q(\theta) f(w(z)) w'(z)}{g(z)} \int m_a(y) p_{a+1}(y, z) dG(y)}_{\text{hires from job-to-job moves}}.$$

[35] For computational tractability, we do not use different values of leaving probability for individual workers of age a in firm z. This simplification is reasonable, given that leaving costs increase proportionally with current human capital and the present value of income flows from current human capital is larger than benefits from future human capital accumulation in most cases of our simulation.

Training. Firms' optimal training is determined by

$$\mu_F \left(\delta_s \tilde{r}(z) + c_s \bar{w}\right) = \zeta \gamma_s s_{F,a}(z)^{\gamma_s - 1} (\tilde{r}(z) - w(z)) \Psi(z, 1, a)$$

where $\Psi(z,t,a) = \sum_{\tau=t}^{T-a} (1-d)^{\tau-1}(1-\delta)^{\tau} \prod_{k=1}^{\tau} \left(\frac{1-\eta\theta q(\theta)\int p_{a+k}(z,z')dF(w(z'))}{1+\rho} \right)$. And workers' optimal training is determined by

$$\mu_W \left(\delta_s \tilde{r}(z) + c_s \bar{w}\right) = \zeta \gamma_s s_{W,a}(z)^{\gamma_s - 1} \frac{J_{c,a}(z) - w(z)}{1-d},$$

where $\frac{J_{c,a}(z)-w(z)}{1-d}$ is workers' return for an extra efficiency unit of human capital in the next period. The optimal training is $s_a(z) = \min(s_{F,a}(z), s_{W,a}(z))$. In comparison with our analytical model, the optimal training level now depends on the present value of all future returns, adjusted for the depreciation rate of training and workers' separation rates (for firms). Notably, optimal training decreases with workers' age, as training young workers produces longer-lasting returns than training old workers. Also note that training does not depend on workers' training and employment histories, which enables us to track the dynamics of average human capital for a firm's labor force.

Evolution of Human Capital. We define $\overline{h}_a(z)$ as the average human capital from training of age-a workers in firms with productivity z. The human capital from training of the youngest cohort is $\overline{h}_1(z) = 0$. We obtain the dynamics of human capital as

$$\begin{aligned}
\overline{h}_{a+1}(z) = & \underbrace{\frac{m_a(z)}{m_{a+1}(z)}(1-\delta)\left[1 - \eta\theta q(\theta)\int p_{a+1}(z,z')dF(w(z'))\right]\left(\overline{h}_a(z)(1-d) + \zeta s_a(z)^{\gamma_s}\right)}_{\text{stayers}} \\
& + \underbrace{\frac{\theta q(\theta) f(w(z)) w'(z)}{m_{a+1}(z) g(z)}}_{\text{new meets/employment}} \left[\underbrace{\eta(1-\delta)\int p_{a+1}(y,z)(\overline{h}_a(y)(1-d) + \zeta s_a(y)^{\gamma_s}) m_a(y) dG(y)}_{\text{meet on-the-job searchers}} \right] \\
& + \underbrace{\frac{\theta q(\theta) f(w(z)) w'(z)}{m_{a+1}(z) g(z)}}_{\text{new meets/employment}} \underbrace{u_{a+1}\overline{h}^u_{a+1}}_{\text{meet unemployed}},
\end{aligned}$$

where $\overline{h}^u_{a+1} = \frac{(1-\theta q(\theta))u_a \overline{h}^u_a(1-d)+\delta\int(\overline{h}_a(z)(1-d)+\zeta s_a(z)^{\gamma_s})m_a(z)dG(z)}{\delta N_M+(1-\theta q(\theta))(1-\delta)u_a}$ refers to the average human capital from training of unemployed people with $\overline{h}^u_1 = 0$.

Vacancies and Wage Determination. We now focus on the conditions for vacancies and wages. The condition for firms' optimal level of vacancies and wages is given by

$$c_v v(z)^{\gamma_v} = \underbrace{\sum_{a=1}^{T} \frac{q(\theta)(\tilde{r}(z)-w(z))}{\sum_a u_a + \eta(N_M - u_a)} \left[\eta(1-\delta)\int p_a(y,z)\overline{h}^s_a(y)m_{a-1}(y)dG(y) + u_a\overline{h}^u_a\right] \frac{\Psi(\phi,0,a)}{(1-d)^{-1}}}_{\text{benefits of new hires' human capital from training}}$$

$$+ \sum_{a=1}^{T} \frac{q(\theta)\left[\eta(1-\delta)\int p(y,z)m_{a-1}(y)dG(y) + u_a\right]}{\sum_a u_a + \eta(N_M - u_a)}$$

$$\times \underbrace{\sum_{t=0}^{T-a} D(z,t,a)\left[\tilde{r}(z) - w(z) + \zeta s_{a+t}(z)^{\gamma_s}(\tilde{r}(z)-w(z))\Psi(z,1,a+t) - \mu_F c_s(z)s_{a+t}(z)\right]}_{\text{benefits of new hires' basic skills and future training}}.$$

We define $D(z,t,a) = \prod_{k=1}^{t}\left(\frac{1-\eta(1-\delta)\theta q(\theta)\int p_{a+k}(z,z')dF(w(z'))-\delta}{1+\rho}\right)$ with $D(z,0,a) = 1$. $\bar{h}^s_a(y) = \overline{h}_{a-1}(y)(1-d) + \zeta s_{a-1}(y)^{\gamma_s}$, and $c_s(z) = \delta_s\tilde{r}(z) + c_s\bar{w}$.

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

$$\sum_{a=1}^{T} \frac{q(\theta)}{\sum_a u_a + \eta(N_M - u_a)} \left[\eta(1-\delta)\int p_a(y,z)\overline{h}^s_a(y)m_{a-1}(y)dG(y) + u_a\overline{h}^u_a\right] \frac{\Psi(\phi,0,a)}{(1-d)^{-1}}$$

$$= \sum_{a=1}^{T} \frac{q(\theta)(\tilde{r}(z)-w(z))}{\sum_a u_a + \eta(N_M - u_a)} \frac{\partial\left[\eta(1-\delta)\int p_a(y,z)\overline{h}^s_a(y)m_{a-1}(y)dG(y) + u_a\overline{h}^u_a\right]\frac{\Psi(\phi,0,a)}{(1-d)^{-1}}}{\partial w(z)}$$

$$+ \sum_{a=1}^{T-1} \frac{q(\theta)}{\sum_a u_a + \eta(N_M - u_a)} \times$$

$$\frac{\partial\left[\eta(1-\delta)\int p(y,z)m_{a-1}(y)dG(y) + u_a\right]\sum_{t=0}^{T-a} D(z,t,a)\left[(\tilde{r}(z)-w(z))(1+\zeta s_{a+t}(z)^{\gamma_s}\Psi) - \mu_F c_s(z)s_{a+t}\right]}{\partial w(z)}$$

Note that this is a differential equation with regard to wage $w(z)$. To solve this, we can multiply each side by $w'(z)$. With this transformation, the right-hand side becomes the derivative with regard to productivity z, and thus we can numerically evaluate $w'(z)$. Combined with the lowest wage $b\bar{w}$, we can iterate the wage structure $w(z)$ until convergence.

F Identification of Model Parameters

We now illustrate how the moments we target help identify our model parameters. We calculate the elasticity of the model-predicted moments to each parameter and provide the results in Table F.1. First, we describe the parameters closely related to labor market outcomes. The moment most sensitive to the constant in the vacancy cost function c_v, is the ratio of vacancies to unemployment. Similarly, since c_m affects the economy's matching efficiency, the moments that identify this parameter are both the self-employment and the unemployment rates. The share of workers who switch jobs due to an idiosyncratic shock, $\underline{p}$, is identified through the wage growth from job-to-job switches and the share of workers who switch from high- to low-paying firms. The self-employment sector share in production, γ, has the largest impact on the self-employment share. A larger shape parameter of the Pareto productivity distribution κ implies fewer productive firms, which reduces the wage sector's relative return to the self-employment sector and the average wage growth after job-to-job transitions. The role of labor in production, μ, has a large effect on capital intensity in production and thus changes the aggregate output and workers' returns in the wage sector. Lastly, a larger productivity level A_M increases aggregate output, especially in the wage sector where a larger productivity level induces more intensive use of capital.

Second, we describe the parameters directly related to training. γ_s affects the degree of diminishing returns of training investments, thus impacting the training levels and wage profiles. c_s pins down the importance of direct training costs and is identified by the ratio of direct costs to wage costs of training. ζ governs the returns to training and has a large impact on training intensity, the wage increase after 40 years, and the self-employment share, because higher training returns make the wage sector more attractive.[36] Training intensity decreases with μ_F—the share of the training

[36]Moreover, γ_s, which defines the convexity on the training function, also has the largest impact

Table F.1: Elasticities of Targeted Moments to Parameters

	Labor Market Dynamics and Productivity							Training Dynamics					Frictions		
	c_v	c_m	$\underline{p}$	γ	κ	μ	A_M	γ_s	c_s	ζ	μ_F	d	γ_p	c_p	δ
Unemployment Rate	0.2	**-1.2**	-0.2	-0.4	**-0.5**	-0.3	0.0	0.0	0.0	0.1	0.0	0.0	-0.1	-0.4	**0.5**
Vacancies/Unemployed	**-0.6**	0.4	**0.7**	0.2	**1.1**	0.3	0.1	0.0	0.0	0.0	0.0	-0.1	0.3	**1.2**	-0.1
Self-Employment Share	0.1	**-0.8**	**0.6**	**5.2**	**1.7**	**2.9**	**-0.6**	**0.9**	0.1	**-1.4**	0.3	**0.6**	0.2	**1.0**	**0.8**
Pareto Parameter	0.1	-0.4	0.4	0.0	**0.8**	0.2	0.1	0.0	0.0	-0.1	0.0	0.0	0.1	**0.7**	0.3
% workers leaving Firm	-0.1	0.4	0.3	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.2	0.4
% workers J-to-U	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**0.7**
Av wage growth J-to-J	0.1	-0.4	**-0.9**	0.0	**-0.6**	-0.1	-0.1	0.0	0.0	0.0	0.0	0.0	-0.2	**-0.5**	0.2
% J-to-J high-to-low	0.0	0.2	**0.6**	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	-0.1
Training Intensity	0.1	-0.4	0.1	-0.4	-0.2	-0.2	0.0	**0.7**	-0.2	**0.6**	**-1.2**	-0.1	0.1	**0.5**	0.0
Trng ratio large-small	0.0	0.1	-0.1	0.0	-0.4	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.2	-0.1
Direct/wage cost (trng)	0.0	-0.1	0.1	0.0	0.3	0.0	0.0	0.0	**1.0**	0.0	0.0	0.0	0.0	0.1	0.0
% wage increase 20 yrs	0.0	-0.1	-0.1	0.0	-0.2	0.0	0.0	**-0.5**	0.0	**0.9**	-0.3	**-0.8**	0.0	0.0	-0.2
% wage increase 40 yrs	0.0	-0.1	-0.1	0.0	-0.1	0.0	0.0	**-0.6**	0.0	**1.0**	-0.4	**-1.0**	0.0	0.0	-0.2
Capital-to-output ratio	0.0	0.0	0.0	-0.2	0.3	**-0.6**	-0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0
Aggregate output	0.0	0.2	-0.2	**-1.4**	**-1.0**	**-1.8**	**1.2**	-0.4	0.0	**0.7**	-0.2	-0.4	-0.1	-0.2	-0.2

Notes: This table reports the elasticity of the model-predicted moments to each parameter. We highlight in bold the elasticities greater than 0.5 in absolute values. The elasticities are measured by calculating the percent increase in each moment after a 1% change around the calibrated parameter value while keeping the rest of the parameters fixed.

cost firms pay. This indicates that optimal training levels are mostly determined by firm choices (as they are lower than workers' choices), which indicates the presence of inefficient training levels. Finally, the depreciation of human capital, d, mainly affects the shape of wage profiles, especially for older ages when the depreciation plays a larger role.

We now focus on the main parameters that mediate the role of our three channels to explain the differences in the training gap across countries and at different stages of development. The contract-breaking cost friction is composed of two parameters. The convexity in the cost, γ_p, has a large impact on the average wage growth from job-to-job transitions, as a higher γ_p makes it more costly to increase the leaving probability in response to higher wage offers. c_p has a large impact on wage growth in job-to-job transitions for the same reason, but it also impacts labor market outcomes such as market tightness, self-employment share, and the Pareto parameter more strongly. It also has a positive impact on training intensity, in line with our analytical model. Finally, the share of workers who are exogenously separated, δ, increases

on the wage growth and self-employment share through the impact on training intensity. The signs are more complicated to analyze, due to the training function choice, because an increase of γ_s increases the marginal returns ($\zeta\gamma_s s^{\gamma_s-1}$), but reduces the overall training returns (ζs^{γ_s}) for $s < 1$.

the unemployment rate and the job-to-unemployment rate, while reducing market tightness (due to having more unemployed people).

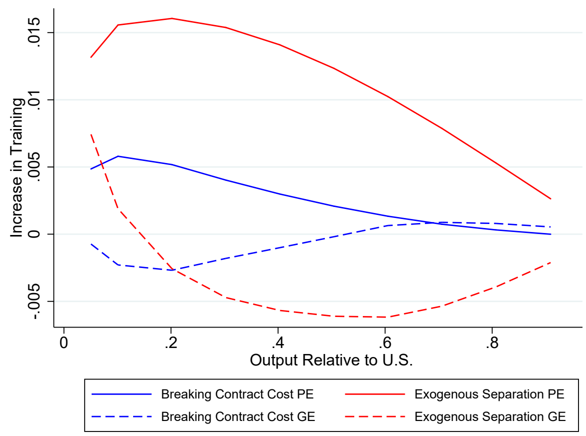
G Decomposition of Partial and General Equilibrium Effects

To further understand the model's dynamics, we go one step further and decompose the change in training stemming from each of the parameter changes above into partial and general equilibrium effects. The partial equilibrium effect captures the change in the firms' training level that occurs when parameter values change while the wage and employment distributions are kept fixed. The general equilibrium effect captures the change in training that occurs due to changes in the wage and employment distributions. We plot the decomposition from the change in the exogenous job destruction and the job-to-job transition friction in Appendix Figure G.1.

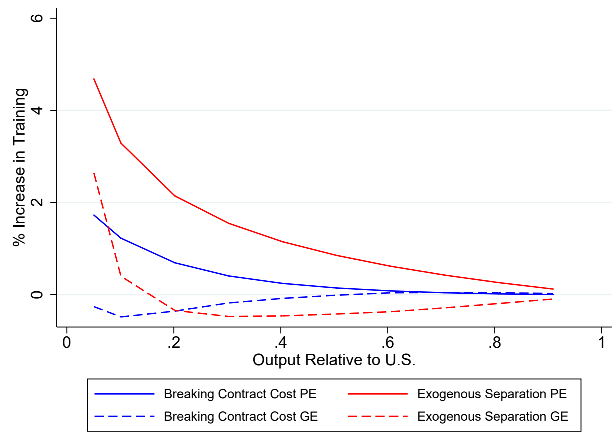
In partial equilibrium, the changes in training from altering each parameter are expected. As the exogenous separation rate δ decreases, the probability of keeping workers goes up, increasing training investments. Nevertheless, the general equilibrium effects may be negative. Lower separation rates decrease the number of unemployed workers among the pool of searchers, and thus, firms must post higher wages to attract workers. Moreover, a lower probability of job destruction means workers move to more productive firms faster, because they enter unemployment on fewer occasions. In this case, higher wages pull training investments down, while the shift of employment distribution toward more productive firms pushes training up. In the poorest economies, the effect from the increase in employment in bigger firms dominates while the wage effect predominates in middle- and high-income countries.

Moreover, as the cost to break contracts c_p increases, firms are able to keep workers for longer and it becomes harder to poach workers from other firms. This generates two main effects. On the one hand, because it is harder to poach workers, firms offer lower wages. This effect encourages firms to invest more in training as they capture a higher share of the surplus. On the other hand, fewer job-to-job switches shift the employment distribution towards less productive firms, which generates a decrease in training. Similar to the dynamics coming from δ, the effects from the change in the employment distribution predominate in the poorest economies, while the wage effect dominates in middle- and high-income countries.

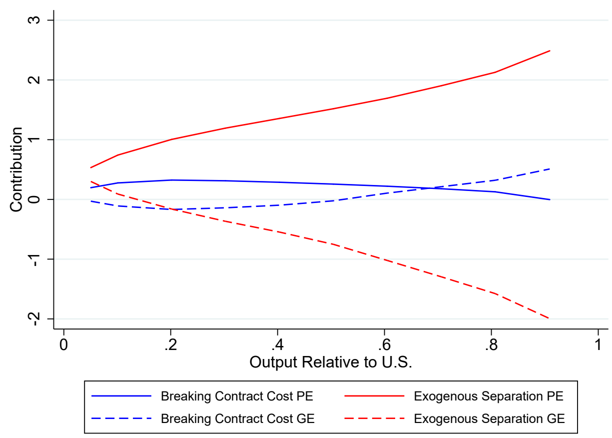
Figure G.1: Partial and General Equilibrium



(a) Training Level Increase



(b) % Increase in Training



(c) Contribution to the Training Gap relative to the US

H Other Channels in the Data

Our model predicts that any phenomena affecting separation rates, the probability of hiring, or the vacancy costs will affect contracts and thus training investments. It is therefore intuitive to think that higher unemployment benefits or firing costs could impact aggregate levels of training in the economy. We test this hypothesis in the data. We rely on the labor market institutional indexes constructed by Botero et al. (2004) to understand how the cost of firing workers and labor market institutions (such as the minimum wage and unemployment benefits) correlate with our measure of training. We regress our measure of training from the WB-ES and EU-CVT on GDP per capita and each of these indices separately, and include year and country fixed effects. We show the results in Table H.1.

Training increases as the legally mandated notice period to fire workers increases. This is intuitive, since when firing costs increase, turnover rates decrease, and agents stay longer in their jobs. In our sample, the amount of severance payment does not seem to be significant to explain on-the-job training. Furthermore, rows 3–5 indicate that as unemployment benefits increase, training investments decrease. This suggests that when the workers' outside options improve, it is harder for firms to retain workers and training investments decrease. This same pattern is observed when countries have meaningful minimum wages and better outside options. Nevertheless, although all these measures explain a portion of on-the-job training differences, their explanatory power pales in comparison to that of GDP per capita (see the last three rows). These results suggest that measures of unemployment benefits and labor market characteristics that are not included in our model (e.g, differences in minimum wages, laws to protect workers, or firing costs) are relevant to explain on-the-job training, but are not the key elements when it comes to explaining the positive correlation between training and income. This result is consistent with Donovan et al. (2020), who find that labor market institutions are an important determinant of cross-country variation in labor market flows (job separation, destruction, and job-to-job transitions), but do not explain the trend relationship between development and labor market flows.

Table H.1: Training and Labor Market frictions (Botero et al 2004)

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Log(GDP pc)	8.41***	8.50***	18.9***	17.8***	20.0***	7.66***
	(1.45)	(1.50)	(4.48)	(4.55)	(4.76)	(1.36)
Legally mandated notice period	0.67*					
	(0.35)					
Legally mandated sev payment		-0.059				
		(0.22)				
Months of contributions for U.B.			12.0			
			(12.2)			
% monthly salary deducted for U.B.				-14.1**		
				(6.94)		
Waiting period for U.B.					-32.7***	
Minimum Wage Index						-10.5**
						(4.04)
Constant	-42.7***	-39.6***	-150***	-117**	-121***	-22.4*
	(13.2)	(13.3)	(44.2)	(47.1)	(44.7)	(13.4)
Observations	183	183	132	132	132	184
R^2	0.421	0.412	0.389	0.395	0.430	0.440
Trend component with no controls						
Log(GDP pc)	8.42***	8.42***	18.8***	18.8***	18.8***	8.36***
	(1.48)	(1.48)	(4.52)	(4.52)	(4.52)	(1.43)
Observations	183	183	132	132	132	184
R^2	0.412	0.412	0.384	0.384	0.384	0.415

Note: Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.