University of Connecticut

Department of Economics Working Paper Series

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

by

Xiao Ma
UC San Diego

Alejandro Nakab
UC San Diego

Daniela Vidart
University of Connecticut

Working Paper 2021-10
May 2021

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

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

Xiao Ma — *UC San Diego*

Alejandro Nakab — *UC San Diego*

Daniela Vidart — *University of Connecticut*

November, 2020

Click here for the latest version

Abstract

This paper offers an explanation for why workers in richer countries have faster rates of wage growth over their lifetimes than workers in poorer countries by providing theory and evidence on the differences in firm-provided training across countries. We document that the share of workers who receive firm-provided training increases with development, and that this is a key determinant of worker human capital investments. We then build a general equilibrium search model with firm-training investments and frictional labor markets. Our model suggests firm-training accounts for a large share of the cross-country wage growth differences. We find that self-employment is the key factor explaining the lack of training in the poorest economies, whereas labor market frictions are key to explaining training differences as countries develop. Finally, our model predicts considerable inefficiencies in human capital investments and sizeable aggregate gains from training subsidies to firms, which may be particularly desirable in poor countries where economic environments disincentivize training.

*Email:xim032@ucsd.edu, anakab@ucsd.edu and daniela.vidart@uconn.edu. We are grateful to David Lagakos, Valerie Ramey and Natalia Ramondo for invaluable discussions and advice. We are also grateful to Titan Alon, Munseob Lee, Marc Muendler, Tommaso Porzio and seminar participants at UCSD for helpful comments. Finally, we 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 (Islam, Jedwab, Romer and Pereira, 2018; Lagakos, Moll, Porzio, Qian and Schoellman, 2018b). Different theories of life-cycle wage growth are consistent with this pattern, including differences in human capital accumulation, labor market frictions or long-term contracts. Since these possible drivers have massively different implications for policy and for explaining cross-country income differences, 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 worker human capital accumulation: firm-provided training. To that end, we carefully measure workers' post-schooling human capital accumulation investments, and explore how they differ across countries. Our results explain why post-schooling human capital accumulation is greater for workers in more developed economies.

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 and worker surveys covering more than 100 countries with vast information on workers' training investments. These surveys allow us to construct harmonized cross-country representative measures of on-the-job training provisions spanning a broad range of development with PPP-adjusted GDP per capita, ranging from $1,000 to $60,000. With different measures of on-the-job training, we document two novel facts.

The first fact we document is that the share of workers who receive firm-provided training rises strongly with the level of country per-capita GDP. We show that a key margin mediating this positive correlation in poor economies is the large share of self-employed workers who do not receive employer-provided training. Moreover, focusing only on workers employed by firms, we still find that the share of workers who are offered training rises with country-level GDP per capita. Richer countries exhibit a larger share of firms offering training, along with a larger share of trainees within the firms offering training 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.

For the second fact, we provide a detailed description of adults' human capital investments and document that job-related firm-provided training is a key determinant of on-the-job human capital accumulation. This evidence suggests that firms play a substantial role in adult 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, 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 post vacancies and wages and meet workers by random search following Mortensen and Pissarides (1994) and Pissarides (2000). We incorporate training investments based on the theoretical literature on general training investments 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.

Motivated by our empirical results and the training literature that considers job turnover as the fundamental barrier to training investments, we focus on three main channels that vastly differ across stages of development and directly impact training: differences in job destruction rates (Donovan, Lu and Schoellman, 2020), differences in contract quality and institutions, which shape labor market dynamics and contract length (e.g. Hall and Jones, 1999; Acemoglu, Johnson and Robinson, 2005), and differences in self-employment shares (Gollin, 2002, 2008). Thus, we calibrate our model to a representative country at different income levels and perform several exercises aimed at answering the following three questions: (1) how much of the wage-growth differences across countries can be accounted for with on-the-job training; (2) why do developed economies invest more in training; and (3) what is the optimal training policy at different stages of development?

We first show that our model matches all of the cross-country differences in wage-growth for countries above $10,000 of per-capita GDP. Nevertheless, it overpredicts the wage-growth for economies at the bottom of the world income distribution. When we include these countries, the model explains 55% of the wage-growth differences. Moreover, we find that the contribution from human capital in explaining workers wage growth is large for every economy and that it decreases with income. This happens because the high level of job destruction in the poorest economies prevents workers climbing up the job ladder. As income increases, fewer workers are separated from their jobs, which generates larger increases in wages through job-to-job transitions. Finally, we show that in our model 70% of the difference in wage-experience profiles is driven by on-the-job training. Moreover, focusing on productivity differences generated by this channel, we show that on-the-job training explains 10% to 15% of the income differences across countries in our quantitative model, which is a large share of the total differences stemming from life cycle human capital gains found in recent studies.

We then conduct a factor-decomposition of training to explore the evolving importance of the different channels at different stages of development and find three main results. First, most of the training gap between the poorest and richest economies is explained by differences in self-employment shares. This is driven by the high rates of self-employment prevalent in poor economies arising from the endogenous allocation of workers as a result of the wage-sector's high labor market frictions. Second, these labor market frictions remain key to explaining training investments as countries develop and self-employment shares fall. The mechanism driving this is the wage-sector's worker turnover. In particular, high job separation rates and low contract quality make worker turnover more likely, and thus depress the incentives to invest in training in low- and medium-income economies relative to richer economies. Third, 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 to explain the lack of training in poorer economies while frictions in job-to-job transitions are more important to explain the training differences between more developed economies.

In our framework, training investments are a negotiated outcome stemming from a joint decision of workers and firms. When deciding the optimal training level, workers and firms do not internalize the other party's and future firms' benefits from training, which generates inefficiencies in training investments. Motivated by the existence of these inefficiencies, we study the optimal training subsidy at different stages of development. We find that a training subsidy is a possible policy that could correct the distortions, and that these might be particularly desirable in poor countries where economic environments disincentivize training. We show that as self-employment and job separation increases or contract quality decreases, the optimal training subsidy rises to incentivize firms to invest in training.

Related Literature. This paper relates to several strands of the 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). Moreover, by embedding this firm-worker training investment dynamic into a search model, we relate to the literature that tries to disentangle the contributions of human capital and search dynamics on earnings (e.g. Bunzel et al., 1999; Rubinstein and Weiss, 2006; Barlevy, 2008; Yamaguchi, 2010; Burdett, Carrillo-Tudela and Coles, 2011; Bowlus and Liu, 2013; Bagger et al., 2014; Gregory, 2019). 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 facing tradeoffs between work and learning, and is simply an exogenous by-product of work.[1] 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.

By exploring the role of worker training in explaining differences in GDP per worker across countries, our paper relates to a large strand of the literature that measures the contribution 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 the ones focused on human capital.[2] Our paper focuses on one understudied source of cross-country human capital difference, namely on-the-job human capital accumulation. Thus, our work relates to the recent literature that highlights the potential importance of life-cycle human capital accumulation differences across countries (De la Croix, Doepke and Mokyr, 2018; Lagakos et al., 2018b,a; Islam et al., 2018). This literature, however, does not explain how these differences in on-the-job human capital accumulation patterns across countries emerge. Our paper attempts to fill this gap by delving into the processes and features giving rise to the low workers' skill acquisition prevalent in poor countries, and focuses its attention 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, Lu and Schoellman, 2020); (2) the vast literature on institutional quality differences across countries (reviewed in Acemoglu, Johnson and Robinson (2005)); and (3) the literature focusing on cross-country self-employment share differences (e.g. Gollin, 2002, 2008; Poschke, 2018). We contribute to this development literature by incorporating the interaction between these channels and firm-provided training. Moreover, through the interaction between employment distribution across firms and training, this paper is related to the misallocation literature, which studies the productivity losses stemming from the large contingent of small unproductive firms in developing countries (e.g. Hsieh and Klenow, 2009; Restuccia and Rogerson, 2013; Bento and

[1] Wasmer (2006) and Flinn, Gemici and Laufer (2017) incorporate micro-founded human capital investment decisions, but they focus on studying the distinction between firm-specific and general training.

[2] These focused on explaining cross-country productivity differences by quantitatively measuring the role of educational attainment (e.g. Hall and Jones, 1999; Erosa, Koreshkova and Restuccia, 2010; Jones, 2014), measuring school quality (e.g. Hanushek and Woessmann, 2012; Schoellman, 2012, 2016), considering the contribution of parental influence for human capital (De Philippis and Rossi, 2020) or differences in skill specialization in secondary and post-secondary curricula (Alon, 2017; Alon and Fershtman, 2019).

Restuccia, 2017; Poschke, 2018). Our paper focuses on documenting a new channel through which these generate productivity losses: lack of on-the-job training.

Finally, our focus on training and cross country analysis closely relates to two recent papers. The first is Doepke and Gaetani (2020) who study cross-country differences in on-the-job skill acquisition focusing on employment protection, which affects firms' and workers' incentives to invest in skills. The second is Engbom (2020) who studies differences in endogenous human capital formation in a search model, and focuses on how the costs of doing business affect human capital. Our work differs from theirs in the channels that we study, which include different labor market frictions 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 they focus on developed economies, we provide evidence and quantitative analysis for countries at all stages of development. The closest paper in analyzing human capital differences at all stages of development is Manuelli and Seshadri (2014). They focus on individual worker decisions, abstracting from firms, and suggest that lower TFP in developing economies raises the cost of accumulating human capital and this lowers households' incentives to invest in human capital after schooling. In this paper, we offer a very different explanation by focusing on firm-provided training, which we show, also empirically, is a key component of adults' human capital investments.

The paper is organized as follows. Section 2 introduces our empirical findings. Section 3 presents the model, and in Section 4, we calibrate a quantitative version of the model. Section 5 shows how the model explains the wage growth differences across countries and presents the income accounting results. Section 6 shows the factor-decomposition of training. Section 7 develops the optimal policy analysis, and in Section 8, we conclude.

2 Empirical Evidence on On-the-Job Training

In this section, we start by describing the data sources and defining key concepts. Then, we proceed to documenting 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 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 ranging from $1,000 to $60,000 of per-capita GDP.

The Enterprise Surveys (ES) from the World Bank are a collection of firm-level surveys of a representative sample of an economy's private manufacturing and service sectors covering approximately 140 low- and middle-income countries. The ES usually interviews owners and top managers of the establishments in the sample who can request assistance of accountants or human resources managers to answer certain questions. The ES has a set of country-specific questions according to each country's characteristics and a set of standardized questions that allow cross-country comparison. We rely on the two ES waves, between 2002 - 2005 and 2006 - 2017, for which they have standardized questions on worker training provisions.

For the EU enterprise data, we rely on the Continuing Vocational Training Survey (CVT). This survey provides information on enterprises' investment in continuing vocational training of their staff, providing information on participation, time spent, and the costs of such training. In our analysis, due to data availability, we rely on 3 of the 5 waves of CVTS conducted in 2005, 2010, and 2015, as these 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 sample survey that provides data on labor participation, unemployment and job characteristics, socio-economic characteristics, and education and training of adults (aged 15+). The survey is conducted on all members of the EU and the 3 European Free Trade Association countries. Although data collection dates back to 1983 for some countries, and the series are generally available from 1992 (according to EU membership), we use time series ranging from 2009 to 2018 for all countries for consistency.

The EU-AES' official objective is to collect information on participation in education and learning activities including job-related training, among other things. Thus, this survey is conducted specially to understand the patterns of adults' education. The AES is one of the main data sources for EU lifelong learning statistics and it covers adults aged 25 - 64. These data were collected during 2007, 2011, and 2017 in 26, 27, and 28 EU Member states, respectively.

Finally, we rely on two secondary United States' data sources for empirical robustness checks, calibration and model validation. First, we provide historical evidence on firm-provided training provision as robustness based on the National Household Education Surveys Program which consist on data on educational activities. Furthermore, to calibrate the model to this country, which we use as benchmark in the quantitative analysis, 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.

2.2 Defining On-the-Job Training

We first carefully define different types of on-the-job human capital acquisition to ensure consistency across different sources. We separate the sources of skill acquisition into four categories that allow for data comparability, while also having simple economic interpretations that can be mapped onto our model. We also present a summary of these learning sources in Table 1. The categories rank from the most structured and planned in advance (schooling) to the least structured (informal learning, which is not structured at all). For expositional purposes, and since 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).

Table 1: Human Capital Sources and Examples

		Firm Sponsored	Non-Firm Sponsored
How Structured	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)

Note: Our definition of schooling reflects "Formal education and training", according to the International Standard Classification of Education 2011 (ISCED 2011), while both Formal Training and Informal Training are categories within "Non-formal education and Training" from the International Standard Classification of Education 2011 (ISCED 2011). Formal and Informal training definitions follow definitions in the World Bank Enterprise Survey and EU continuing Vocational Training Survey. The definition of Informal Learning follows from the EU continuing Vocational Training Survey. 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).

The first category considered follows from the "Formal education and training" category of the International Standard Classification of Education 2011 (ISCED 2011)[3], which is defined as "education that is institutionalized, intentional and planned through public organizations and recognized private bodies which in their totality constitute the formal education system of a country..."[4] Following this definition, we refer to this category as

[3]ISCED 2011 provides "uniform and internationally agreed definitions to facilitate comparisons of education systems across countries." and was adopted by UNESCO.

[4]ISCED 2011 also adds that "...Formal education programs are recognized as such by the relevant national education authorities or equivalent authorities... Formal education consists mostly of initial education but it also includes vocational education, special needs education and some parts of adult education."

schooling for clarity, although specific institutionalized courses are also included.

The second and third categories, formal and informal training (collectively referred to as *training*), follow from the "Non-formal education and Training" category of ISCED (2011), which is any organized and structured learning activity outside the formal education system. *Formal Training* is defined as training that has a structured and defined curriculum, and includes classroom work, seminars, 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. Further, this training is typically more general, not geared to specific 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 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.

It is important to note that *Formal Training* may be organized and financed by the firm or it may be self-financed and self-directed by workers if workers attend these courses outside working hours and pay for them. On the other hand, *Informal training* is always firm-sponsored, as it happens during working hours and is directed by firms' managers, workers' supervisors, or colleagues.

Finally, for completeness, we define the last category, *Informal Learning*, as a type of a learning activity that is not structured and that is more related to workers' self-investments. Some categories described in the data encompassing this category are: 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 self-directed and employers are not usually involved.

2.3 On-the-Job Training Facts

The wide variety of data sources allow us to analyze training patterns for more than 100 countries, and to describe in detail the key sources giving rise to adult human capital accumulation. In this section, we document two key facts about firm-provided training. First, we document that on-the-job firm-provided training increases with GDP per capita. We find that this pattern is partially driven by the large share of self-employed workers who do not receive employer-provided training in poor economies, but also we show that the share of firms offering training, the share of participants per firm, the amount of hours per participant and the training cost per participant increases with income. Second, we show that job-related firm-provided training is the main source of adults' education, and that this type of training explains the bulk of the differences in adults' human capital investments across countries.

Fact 1 ***A positive cross-country correlation between firm-provided training and income exists***

We first focus on formal training, which is available from and consistent across enterprise surveys for all 100 countries in our data to study the correlation between on-the-job training and cross-country income. Initial vocational training, employee orientation, apprenticeships and informal training are explicitly excluded. After analyzing the formal training measure, we will focus on broader forms of training for a smaller sample of countries for which data is available. We construct country-year measures of the share of employees who receive formal training with the following formula:

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

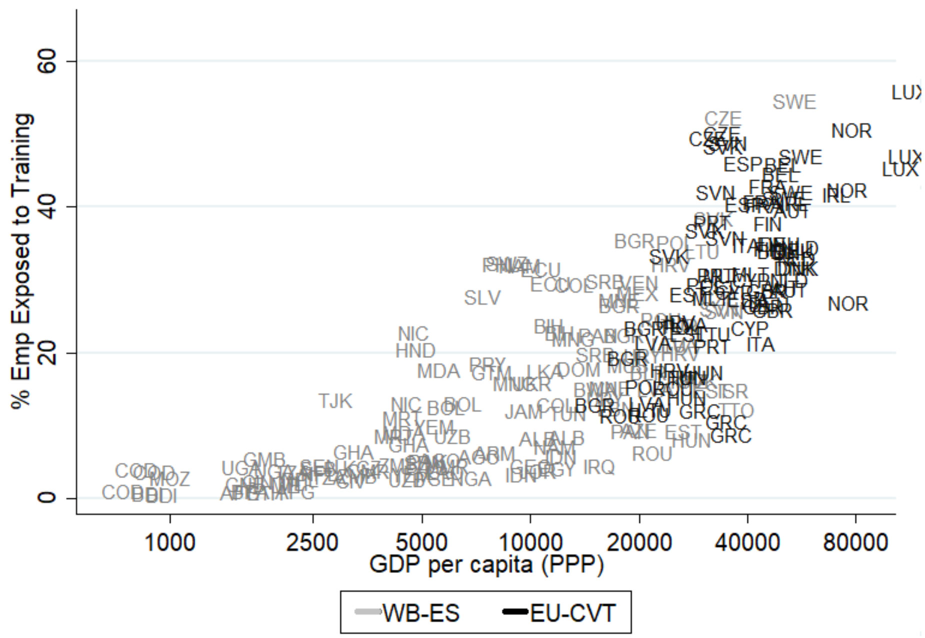
The WB-ES provides information about which firms provide training, along with the share of workers who were trained in those firms. We use these two measures to construct a country-year measure of the share of employees offered training. Since only firms are surveyed, we then adjust this measure by the share of self-employment for the main specification, assuming that self-employed workers do not receive training from employers.[5] In the EU-CVT, on the other hand, we have data on the share of workers who were trained in all the enterprises surveyed, which we also adjust by self-employment.[6]

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

[5]We also construct a second measure where we adjust our measure of trained workers by the share of workers who are not self-employed and who do not work in agriculture, assuming that workers in the agricultural sector do not receive training. This follows from recent findings suggesting that the returns to experience (hence human capital accumulation) are much lower in agriculture than in the manufacturing and service sectors (Lagakos et al., 2018b; Islam et al., 2018). The results do not change significantly when we use either of these two measures (see Appendix Figure C.2)

[6]We restrict the sample from the WB-ES to 2005 - 2015 for comparability with 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 USD 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 big firms. We show the same pattern with the unrestricted sample in Appendix Figure C.1.

Figure 1: Share of Employment Formally Trained and Development

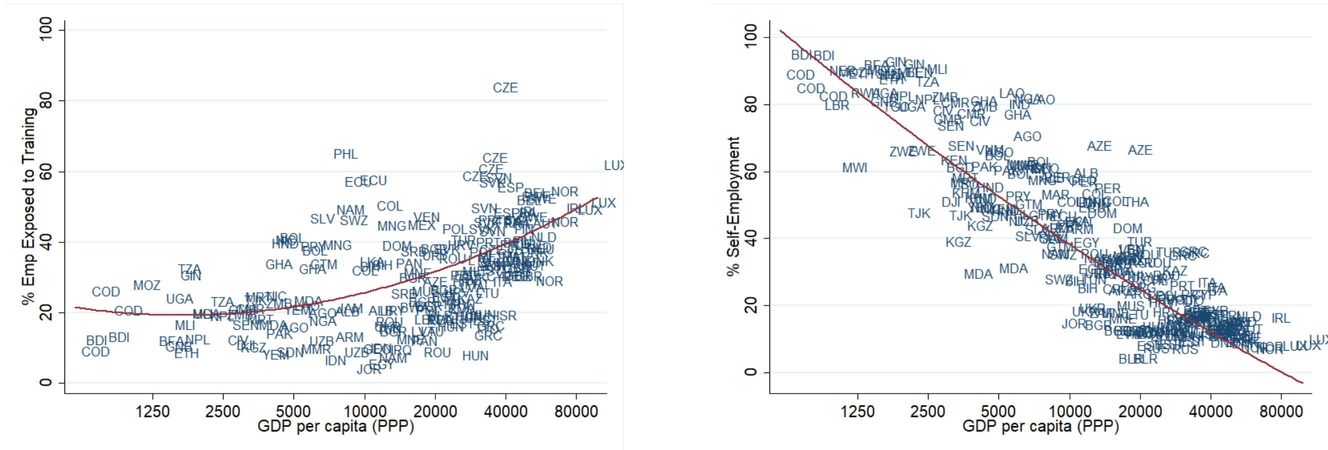


Note: The share of formally trained employment follows from adjusting the share of workers who were trained by firms by the share of self employment. Data on the share of employees trained inside firms comes from the World Bank Enterprise Survey (WB-ES) for all developing economies and from EU Continuing Vocational Training Survey (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 - 2017 for which we have firm weights. We restrict the sample from the WB-ES for the years between 2005 - 2015 to have the same years as the EU-CVT, and we restrict to countries with mean log employment in firms lower than 4. Data on GDP per capita and self-employment comes from the World Bank Indicators.

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 employees 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 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 mechanism driving low 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 settings. 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 much 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 in poor economies.

Figure 2: Unadjusted On-the-Job Training Shares and Self-Employment



(a) % Trained workers in Firms

(b) Self Employment

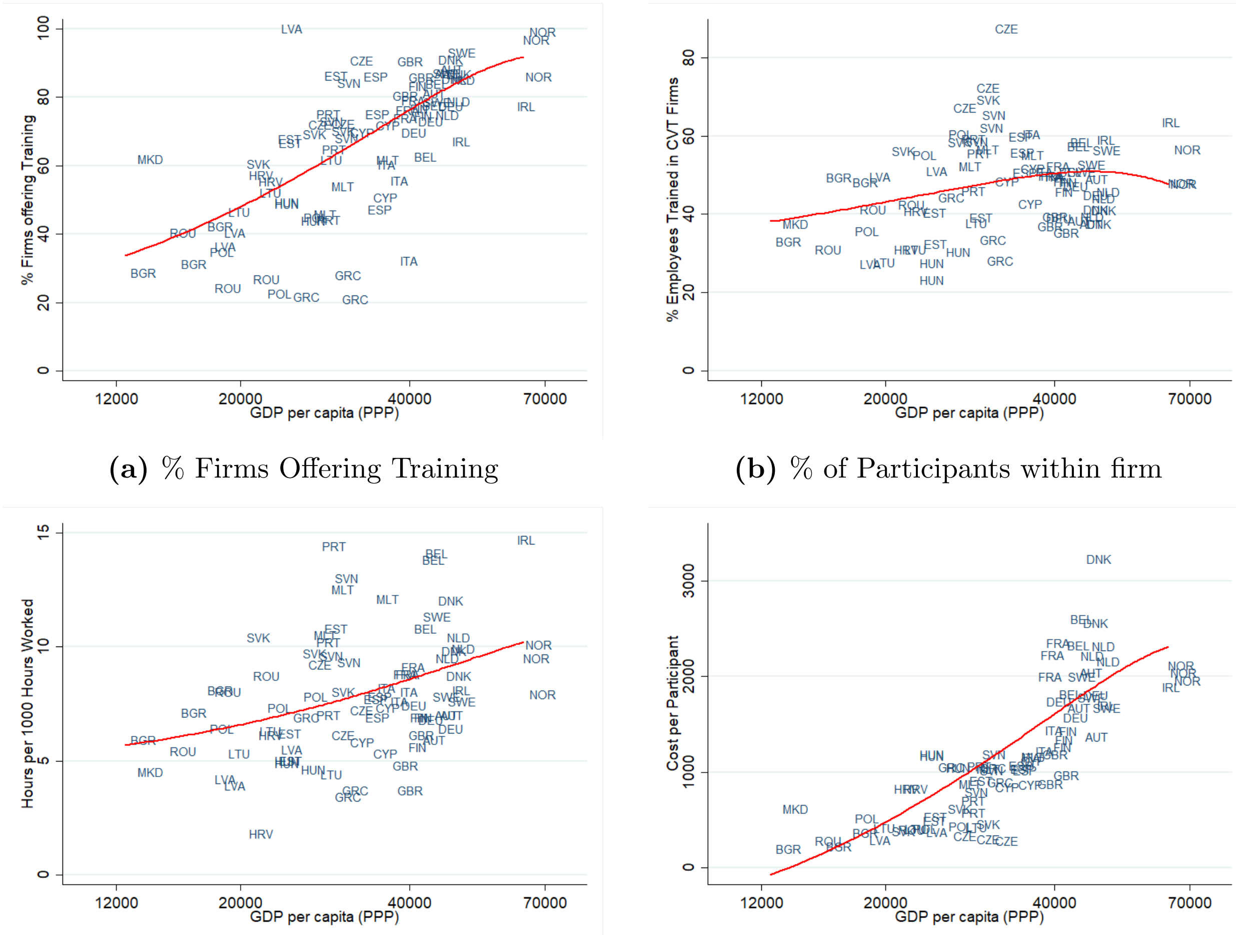
Note: This figure shows both margins from the formal training measure: in panel A we show workers who are trained by employer as a share of total workers in firms and panel B shows the share of workers who are self-employed. Data on the share of trained workers in the wage sector comes from the share of workers formally trained in the World Bank Enterprise Survey and the European Union Continuing Vocational Training Survey, which focuses on formal training and does not include initial vocational training or specific-worker targeted training. Self-employment data comes from the World Bank Indicators, which provide ILO estimates for each country-year.

Wage-sector training increases with development in every margin. We can analyze the relationship between training and income within the wage sector using enterprise survey data from European countries. Although we rely on fewer countries, the survey time frame allows us to have large income variation. We find that this positive correlation is prevalent among 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), along with a larger share of trainees inside firms offering training, and a larger share of hours in training relative to total hours worked (intensive margin).[7] In addition, richer countries exhibit a larger cost of training, which potentially proxies training quality.

Continuing and initial vocational training increases with development. Our formal training measure is based on continuing vocational training (it does not include worker orientation or initial training), which seems the most relevant margin to explain life cycle increases in productivity. Nevertheless, it could be the case that continuing vocational training is larger in developed economies, but the initial vocational training (IVT), that takes

[7]In Appendix Figure C.4 we show that the share of hours in training relative to total hours worked increases with income, but total hours of training per participant remains constant. This is consistent with workers working fewer hours as income increases (Bick, Fuchs-Schuendeln and Lagakos, 2018). In Figure C.4 we plot the time spent in training for the EU-CVT and the EU-AES.

Figure 3: OTJ Training Margins within the Wage Sector



(a) % Firms Offering Training

(b) % of Participants within firm

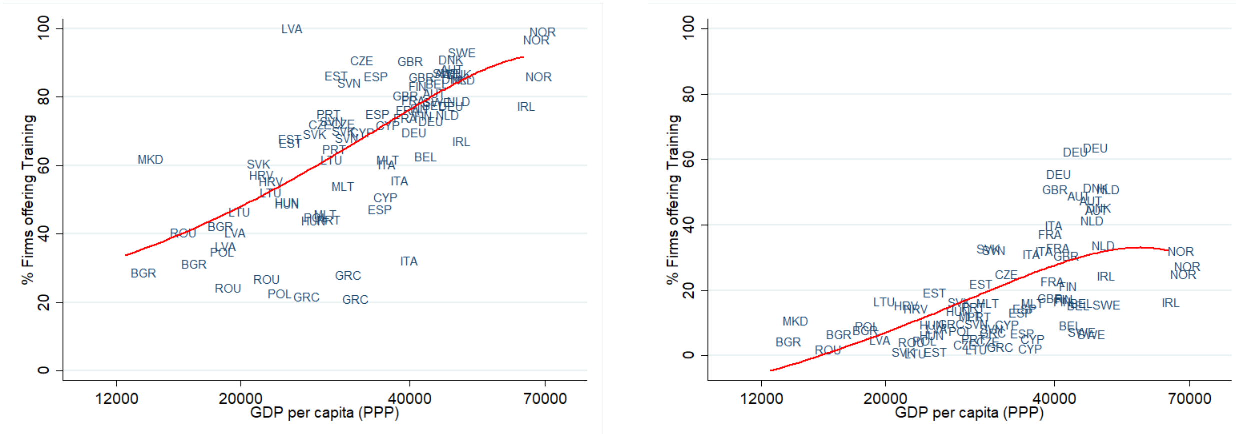
(c) Hours per 1000 Hrs Worked

(d) Cost per Participant

Note: This figure shows all margins of training coming from the EU Continuing Vocational Survey. Panel a shows the share of firms that offered training, which is defined by firms that offer 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 (participants and non-participants of the training). Panel d shows the total cost of training per participant that includes direct and indirect training costs (wages of trainers and wage lost by not working during training).

place when the worker starts the job, is larger 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 that there is a difference in the timing of human capital investment across countries. Interestingly, Germany has very high levels of IVT, which aligns with previous studies that analyzed German training programs.

Figure 4: Share of firms Offering CVT and IVT



(a) Firms offering CVT **(b)** Firms offering IVT

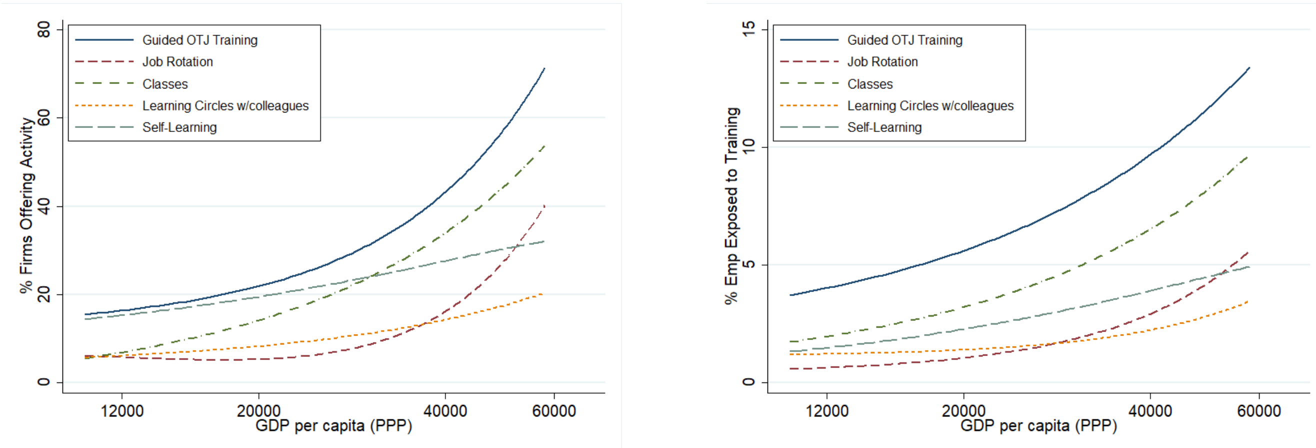
Note: These data come from the EU Continuing Vocational Training. Panel a shows the share of firms that offer continuing vocational training (CVT) and panel b shows the share of firms offering initial vocational training (IVT). CVT is defined as all training for workers except for the initial training to show workers job-specific skills for the new job or to teach workers general knowledge about the firm as they enter to a new job, which is included on the initial vocational training. Data on GDP per capita comes from the World Bank Indicators.

Informal on-the-job training increases with development. We focused on formal training to provide evidence for countries in all stages of development. Nevertheless, 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 the content 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 5 different training types: 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 Figures 5a and 5b 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.[8] We conclude that all informal training activities increase with development.[9]

[8]In Appendix Figure C.5 we show that as GDP per capita increases, the share of firms offering both formal and informal training increases, which suggest they might be complements.

[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 for other types of worker informal learning. However, this does not seem to be the case. Appendix Figure C.7 provides measures of all types of informal learning in the AES survey (learning from peers, colleagues; learning by using printed material, learning by using computers, learning through media; learning through guided tours in industrial sites/museums among others; learning by visiting learning centers as libraries. Further descriptions can be found in the Appendix data source description section), and we show these have positive correlations with development if at all.

Figure 5: Informal Training



(a) % Firms Offering Activities **(b)** % Workers Offered Activities

Note: This figure shows 5 types of "other forms of continuing vocational training" in the CVTS survey, which 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; planned training through self-directed learning/e-learning. This corresponds to questions B2 in the CVT3 (Year 2005) and CVT4 (Year 2010) surveys. We do not include CVT5 due to lack of data on exact shares of workers who received training.

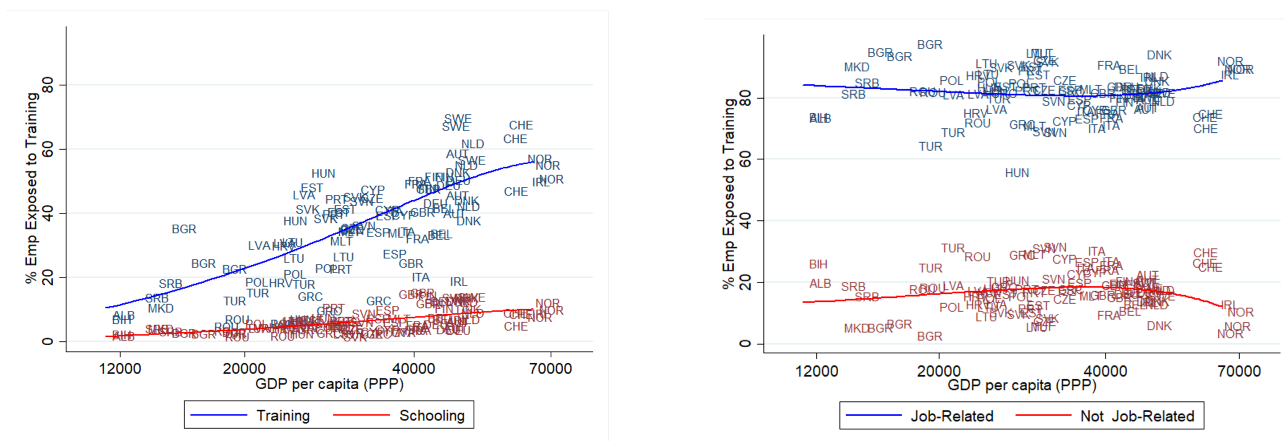
Fact 2 ***Firm-provided training is the main source of adults' education.***

Until this point, we have shown a strong correlation between on-the-job training and development using enterprise-level data. However, if on-the-job training is a small fraction of adults' human capital investments, this positive correlation will not be useful to explain cross-country human capital differences. Thus, we now turn our attention to labor force and worker surveys containing detailed information on workers' training activities and education, which allow us to quantify the role of on-the-job training relative to other human capital sources. In particular, we focus on data from the European Union Adult Education Survey and the Labor Force Survey, which contains information on the characteristics of all education and training investments in European countries. We document that the most important source of human capital investment for adults is job-related training financed by firms.[10]

[10]In Appendix Table C.2 and Table C.1, we further show that on-the-job training predominantly occurs during working hours, and its objective is to improve technical and job specific skills. Table C.1 shows that, on average, workers in the European Union were reportedly trained during paid working hours for 70% of the training length. In Table C.2, we first show the purpose of training reported by firms for the EU-CVT surveys in 2010 and 2015. These purposes range from general IT, management, and team working to problem solving and improve technical skills. Between 10% to 40% of firms reported having done some training on most of the categories, but the category with the highest level of reporting (around 70% of firms had training with this purpose) was technical and job-specific skills, which suggests a productivity-enhancing nature.

Most of adults' 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 last year) is training, while less than 10% is schooling. Additionally, panel (b) shows that around 80% of workers who reported participating in some type of training (blue measures in panel a) claimed that this was job-related, and interestingly this share is uncorrelated with cross-country income. Moreover, Appendix Table C.1 shows the same pattern when looking at the share of adults who reported being involved in training in the labor force survey (EU-LFS). On average 84% of adults in European countries reported that the education they received had been job-related and only 16% mentioned personal or social reasons as the purpose of their training or education. This evidence suggests that job-related training is a primary source of adult learning and human capital accumulation.

Figure 6: Adult Human Capital Accumulation Characteristics



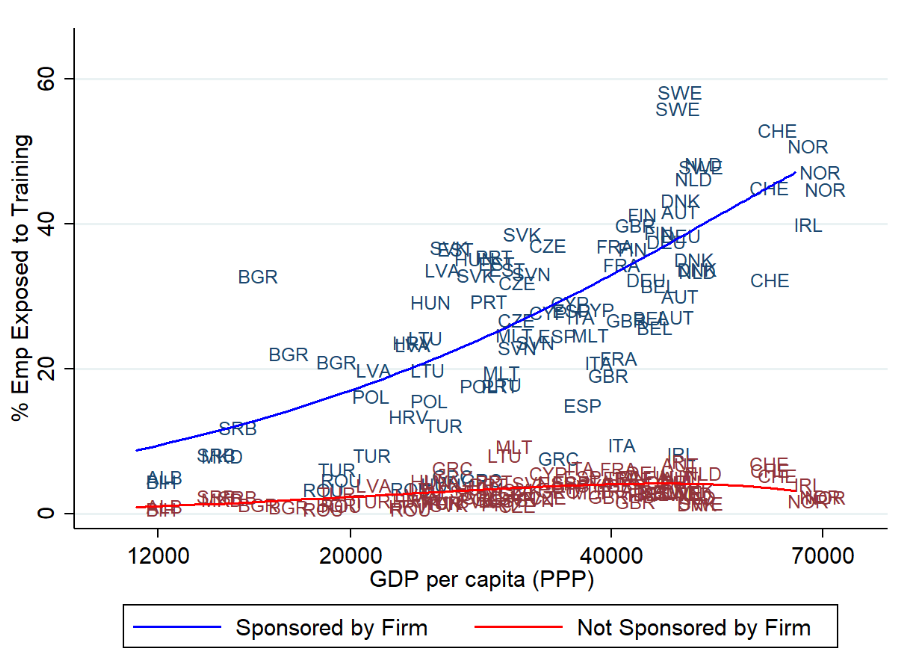
(a) Type of Education

(b) Job Relation of Training

Note: These data comes from the EU Adult Education Survey. Panel (a) shows the difference in share of adults who participated 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) while "schooling" refers to "Formal education and training" according to the International Standard Classification of Education 2011 (ISCED 2011). Panel (b) refers to the share of training that was job related from all the training reported in panel (a) (blue line). Data on GDP per capita comes from the World Bank Indicators.

Almost all of the job-related training is sponsored at least partially by firms. Figure 7 shows how the proportion of job-related training, financed at least partially by the firm or completely by the worker, varies across European countries. The graph shows that the vast majority of job-related training is sponsored by firms. In particular, less than 5% of workers for all countries receive some training that is directly related to their job and that is entirely self-financed. Moreover, the share of adults who fully self-financed their job-related education is constant as a function of per-capita GDP, which reflects the fact that the increase in job-related training with income is driven by firms offering more training, and not by workers themselves investing more in education. Doing some back-envelope accounting, our results show that 90% of all human capital investment is training, 80% of all training is job-related, and that almost 100% of this job-related training is financed by firms. This means around 72% of all human capital investments is at least partially provided and financed by firms. This striking result indicates that the most important source of adult human capital accumulation corresponds to firm investments in human capital, and not workers' self-investment outside the firm. Moreover, as robustness in Appendix Figure C.6, we provide United States' historical evidence from 1991 to 2005 and show job-related training provision accompanies economic growth in the time series as well.

Figure 7: Training Financing



Note: The data on training financing come from the EU Adult Education Survey. The graph shows the difference in share of adults who participated in sponsored and non-sponsored job-related training relative to GDP per capita. Data on GDP per capita come from the World Bank Indicators.

These patterns imply that employer-provided training is a key determinant of on-the-job human capital accumulation, and that firms play a substantial role in adult 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 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 understand cross-country human capital and income differences.

3 Model

To explain the positive correlation between training and development and its link to firm-level decision-making, we focus on two main themes: self-employment, and job turnover. The self-employment theme is motivated by our empirical evidence, and particularly by the fact that the high prevalence of this type of work is a key mechanism driving low training in poor economies. The job turnover theme is rooted in the literature, specifically on the fundamental problem of financing training investments first identified by Becker (1964). Training investments are less likely to occur if the probability of losing the worker is high. We focus on two factors affecting turnover, which are especially salient and have been widely studied in the cross-country context: differences in contract quality, and labor market frictions. As Acemoglu, Johnson and Robinson (2005) stress *"Economic institutions matter for economic growth because they shape the incentives of key economic actors in society, in particular, they influence investments in physical and human capital and technology, and the organization of production..."*

3.1 Environment

The model economy is populated by a continuum of workers whose life spans 2 periods. Every period, the same number of workers who die are born, and we normalize the size of each generation's population to be 1. 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:

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

where superscripts Y and O denote young and old ages, and $\rho > 0$ governs time preference. In the steady state, $\rho = r$, therefore workers are indifferent between consuming in each period.

3.1.1 Consumption Good Production and Self-employment

The consumption good is a composite of goods from two different sectors: the traditional sector good Y_T and the modern sector good Y_M,

$$Y = (\gamma Y_T^\sigma + (1-\gamma) Y_M^\sigma)^{\frac{1}{\sigma}}$$

Production in the traditional sector is characterized by a constant-returns-to-scale function:

$$Y_T = A_T N_T$$

where A_T is productivity and N_T is labor in that sector. This sector is related to self-employment and we assume training is not provided to workers. In our model, we normalize the price of the good produced by the modern sector to be 1. Therefore, the price of the good produced by the traditional sector is:

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

3.1.2 Modern Sector

This sector is characterized by frictional labor markets. There is a unit measure of firms, which are heterogeneous in productivity $z \sim H(z)$ and produce a homogeneous good. Once workers and firms are matched, worker i's production in firm j is given by:

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

where A_M is productivity in this sector, z_j is the firm-specific productivity and h_i is worker i's efficiency units of labor (human capital). This production function suggests human capital and technology are complementary.[11] By integrating this expression across all workers within firm j, we get total production of firm j:

$$y_j = \int_{i \epsilon j} A_M z_j h_i di,$$

and aggregating production over all firms, we get production in the modern sector:

$$Y_M = A_M \int_{j \epsilon J} z_j \int_{i \epsilon j} h_i \; di \; dj.$$

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 vacancies cost is defined by $c_v \frac{v^{1+\gamma_v}}{1+\gamma_v}$ and we require vacancy costs to be convex (i.e. $\gamma_v > 0$) such that firms with different productivity levels coexist. The total number of vacancies is then $V = \int v(z) dH(z)$.

[11]This assumption is consistent with studies finding complementarities between technology and human capital (Acemoglu and Zilibotti, 2001; Porzio, 2017)

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 job full time jointly with all newborn workers. Moreover, a portion η of 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 modern sector.[12]

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 Worker's Optimal Separation Policy. If the remaining old workers who search while on the job get an outside offer, they can make efforts to break the contract with probability p, by incurring the costs of breaking the contract $c_p^{\gamma_p} \frac{p^{1+\gamma_p}}{1+\gamma_p}$ per efficiency unit. The costs of breaking the contract represent costs of being caught and filing lawsuits, with a lower c_p representing weaker institutional environments. We assume $\gamma_p > 0$ such that the marginal cost of breaking the contract increases with probability p.[13]

We first solve workers' optimal choice of leaving probability taking the level of training investment as given for two reasons. First, when the new offer arrives at the end of the period, training already occurred. Second, firms and workers need to internalize workers' probability of leaving the firm to decide the optimal level of training. Thus, they must choose training according to the optimal breaking contract efforts conditional on each new offer. In a firm with productivity z, a worker chooses the optimal leaving probability $p \in [0, 1]$ when faced with an outside offer w' solving:

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

[12] Donovan, Lu and Schoellman (2020) suggest that the share of the higher turnover in developing economies that is not explained by observables (such as occupation, industry, education and firm size distribution) can be best characterized by the presence of match-specific productivity and screening. In Appendix Section N, we add match-specific productivity heterogeneity and show that better screening technologies or less productivity dispersion (which is consistent with the data) translates to lower separation rates which we capture by calibrating δ to the share of workers who are separated each quarter with no loss of generality.

[13] Consistent with previous literature (Acemoglu and Pischke, 1999), in our setting firms cannot break the contract as they pay for their share of training cost in the first period. Note that if firms have long-term reputation, the absence of contractual problems from firms breaking contracts is reasonable.

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

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

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

Training Determination. A young worker has initial human capital $h^Y = 1$ (normalization) and can be trained for s efficiency units of time to enjoy an increase in the next-period 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 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. In principle, training also reduces trainees' production time. As analytical properties will not be affected by training time costs, we omit them here, but we add them in the quantitative analysis as this is a key feature of the data.

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' perspective, respectively, training $s = \min\{s_W, s_F\}$. This assumption says the training level is determined by either party with lower affordability. It is a reasonable assumption: 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:*

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

where, in a firm with productivity z, current wage w, new offers of wage w', a wage distribution from 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:

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

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

Proposition 1 explains how optimal training is decided with different divisions of training costs. As the share each group pays 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) is due to frictional labor markets, because of which firms are able to extract partial rents from training (Acemoglu and Pischke, 1999). We differ from the general training literature (e.g. Acemoglu and Pischke, 1999; Engbom, 2020) in that we assume the cost shares workers and firms pay are common across firms and we add a time cost of training when we take the model to the data. In Appendix Section M, we show the model's results for three different specifications on how the training costs are financed: firms paying all training costs, joint internal efficiency where workers and firms distribute the costs efficiently (they pay the share proportional to their income gains), and the one we presented in our paper. The evidence suggests that if we include training time costs, the common share is the right assumption as it is the only one that can jointly match the training pattern with respect to firm productivity and the aggregate training levels.[14]

Moreover, in Appendix Section M, we show the training patters in the aggregate and the optimal training decisions for firms and workers for the calibration that matches all labor

[14] Training levels increase with firm productivity when the firm pays all the cost or when the share of the cost that the firm pays is constant. Nonetheless, the case of joint internal efficiency generates a decrease in training investments with firm productivity, which is counterfactual, as in the data, firms that are bigger and more productive invest more in training. Moreover, the model needs really implausible high levels of training productivity to match the levels of training observed in the data when firms pay the total cost of training. This suggests that the common share is the right assumption as it is the only one that can jointly match the training pattern with firm productivity and the aggregate levels of training.

market moments. We find that firm decisions define training investments as firms always want lower levels of training than workers. Thus, we now focus in understanding firm-level decisions.

Proposition 2 (Labor Market Frictions and Firms' Training) *In a firm with productivity level z, firms' optimal training $s_F(z)$:*
(1) increases with costs of breaking the contract c_p;
(2) decreases with on-the-job search probability η;
(3) decreases with exogenous separation rate δ.

Results in Proposition 2 indicate that a higher probability of job separation leads to lower training. This result provides the mechanisms for which better institutions and lower job destruction generates more training investments in our model.

Moreover, a wage increase has two opposing forces affecting training decisions. On the one hand, the incentives to invest in training decrease because firms capture a lower share of the surplus of the match but, on the other hand, the probability of keeping the worker increases, which generates higher training incentives. Interestingly, a wage increase in the aggregate economy do not impact the probability of keeping workers but 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. We show in Appendix Section O a simplified version of the model with different assumptions on wage formation. If the labor share does not change across firms, training investments will be proportional to productivity, but if it decreases in productivity (as happens in our model and data), training will be increasing and concave in productivity. In Appendix Section D, we show these correlations in the data.

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:

$$J(z, l^O_{-1}, w_{-1}, X_{-1}) = \max_{\{w,v,s\}} \underbrace{l^O_{-1}(1-\delta)\left(1-\eta\int p(w_{-1}, w')dF(w')\right)(A_M z - w_{-1})}_{\text{profits from remaining workers}}$$

$$+ \underbrace{\frac{v}{\theta}\frac{1}{1+\eta(1-\delta)+\delta}(A_M z - w - \mu c_s s)}_{\text{profits from hiring young workers}} + \underbrace{\frac{v}{\theta}\frac{\eta(1-\delta)\int p(w', w)dF_{-1}(w')\bar{l}(w) + \delta\bar{l}}{1+\eta(1-\delta)+\delta}(A_M z - w)}_{\text{profits from hiring old workers}}$$

$$- \underbrace{\frac{c_v v^{1+\gamma_v}}{1+\gamma_v}}_{\text{vacancy costs}} + \frac{J(z, l^O, w, X)}{1+r}$$

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

where we use the subscript -1 to denote pre-determined variables. l^O is the total supply of efficiency units by old workers before exogenous separations. $F_{-1}(w)$ is the wage distribution of job offers during the last period. The first term represents the net profits generated by all the workers who remain in the firm from the last period. The second term represents the profits (net of training costs) from hiring young workers. The third term represents the profits from poaching old workers who are willing to move to the current firm. The 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 $\bar{l} = 1 + \int \zeta s_{-1}(z)^{\gamma_s} dF_{-1}(w_{-1}(z))$. Note that $s(z)$ is determined according to Proposition 1, whereas $w(z)$ and $v(z)$ are determined according to FOCs. In particular, as shown by Burdett and Mortensen (1998), $w(z)$ is determined by a first-order differential equation, combined with the minimum wage $b\bar{w}$.

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 at some point, as costs of additional vacancies are increasing.

Equilibrium. In the steady-state equilibrium, labor and goods markets are clear for both traditional and modern sectors for each period. We abstract from workers' reshuffling between modern and traditional sectors, which needs tracking employment and training histories of each worker and is computationally intractable. Several assumptions can lead to immobility between sectors, including complete asset market, family networks, or large switching costs. Therefore, workers must also be indifferent in terms of expected utility between going to the traditional sector and the modern sector in the first period:

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

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. Finally, output per capita is defined as:

$$\frac{Y_i}{N_i} = A_T \left[\gamma(\Lambda_T)^\sigma + (1-\gamma) \left(\frac{A}{A_T} \Lambda_M \int z \left(1 + \frac{\zeta s(z)^{\gamma_s}}{2}\right) dF(z) \right)^\sigma \right]^{\frac{1}{\sigma}}$$

where Λ_i is the share of workers in sector i. Thus, output per capita depends on how workers are distributed between sectors and firms, sector-specific productivities and training investments in the economy.

Self-Employment Shares and Training. Proposition 1 and 2 provide the mechanisms for which better institutions and lower job destruction generate more training investments in our model. We now focus on how training is affected due to changes in self-employment shares. Every change that lowers the returns of working in the wage sector relative to the traditional sector generates a higher share of workers allocated in the self-employment sector and a decrease in training in the aggregate, conditional on the wage-sector 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 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, since job search models are usually calibrated using high-frequency labor flows data. With this change in mind, an age-a worker's utility function can be written as:

$$\max_{\{c_\tau\}} \sum_{\tau \geq a}^{T} \left(\frac{1}{1+\rho}\right)^{\tau - a} c_\tau$$

We still normalize each generation's population to be 1, hence the total population in the economy is T. We assume human capital depreciates at rate d, and therefore evolves as $h' = (1-d)h + \zeta s^{\gamma_s}$. As in our analytical model, we abstract from workers' reshuffling between modern and traditional sectors, and thus, in the equilibrium, workers are indifferent between working in the wage sector and as self-employed in the first period.

Firms. Training costs are assumed to be proportional to the average wage $c_s \bar{w}$, while training per unit time also causes a δ_s decrease in efficiency units for production. With this assumption, firms' and aggregate training levels remain constant in response to proportional changes of A_T and A_M. As we aim to understand why developed countries invest more in training, we explicitly avoid capturing the training difference driven by the TFP level. With workers of different ages, firms can impose different training levels on different individuals. In particular, training intensities decrease with age.

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-estimated matching function $M(\tilde{U}, V) = c_M \tilde{U}^{\psi} V^{1-\psi}$ for the modern sector. This matching function produces unemployment and meaningful elasticities of matches with regard to searchers $\tilde{U}$ and vacancies V. 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.

Conditions for Simulations. To save space, we delegate the optimal conditions for the quantitative model to the Appendix Section E, as they provide the same intuition as in our analytical model. We show the optimal levels of training depending on firm productivity and age of workers, and the conditions for wages and vacancies for firms.

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 12 moments describing labor market dynamics and training investments, which allow us to identify model parameters. Then, we perform a second calibration for a representative economy 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, and A_T to match self-employment, job destruction rate, job-to-job transition, and income levels for each representative economy.

4.1.1 Calibrating the Model to the United States

Preassigned parameters. We first directly set some parameters following the literature. We calibrate a model of quarterly frequency and 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 $b = 0.5$ following Hornstein, Krusell and Violante (2011) who compute the mean-min ratio of wages to be around 2 from the U.S. Census 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 elasticity of substitution between the traditional and modern sectors in the aggregate production function as in Feng, Lagakos and Rauch (2018). We set the depreciation rate of human capital from training to be 0.7% quarterly according to Manuelli and Seshadri (2014). Finally, we set the relative intensity of on-the-job search intensity to be 0.4 following Faberman et al. (2017) who found that the average amount of offers per month for employed people is around 40% of that for unemployed people in the U.S. data.

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) found 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. Finally, we normalize the United States' aggregate productivity to be $A_M = A_T$ and unity.

Parameters to estimate. The remaining parameters to estimate are: the constant in the matching function, c_M; training costs per time as a share of the average wage rate, c_s; the

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	0.5	Hornstein, Krusell and Violante (2011)
ψ - Elasticity of matching function to searchers	0.7	Shimer (2005)
$\frac{1}{1-\sigma}$ - Elasticity of substitution	3	Feng, Lagakos and Rauch (2018)
d - Depreciation rate of human capital	0.007	Manuelli and Seshadri (2014)
γ_v - Convexity of vacancy costs	1	Dix-Carneiro et al. (2019) 0.8-2.3
δ_s - losses in production hours per unit time	0.7	EU-LFS 2004 Training Module
A_T, A_M - Productivity in M and T sectors	1	Normalization

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 traditional-sector share in the aggregate production function, γ; the convexity in the function of leaving costs, γ_p; the shape parameter of Pareto productivity distribution, κ; exogenous separation rates, δ; lower bound of leaving probability, $\underline{p}$; and the share of training costs paid by firms, μ_F.

Targeted moments and fit. To calibrate those remaining parameters, we target the following moments listed in Table 3: average unemployment rates in 1994 - 2007, as computed by Hornstein, Krusell and Violante (2011); the ratio of the amount of vacancies to the amount of unemployed people, from FRED for 2000 - 2007 (data available after 2000); the share of self-employment in total employment for 1994 - 2007, from the World Bank; the Pareto parameter of firm employment distribution, as estimated by Axtell (2001); workers' average wage growth after job-to-job transitions, 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 BLS, which has both employers' and employees' information. We add the percent wage growth of 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—provided by Donovan, Lu and Schoellman (2020) for 40 countries of different development levels. As we are interested in how institutional quality affects job transition at different income levels, we first regress probabilities for countries available in Donovan, Lu and Schoellman (2020) on all institutional measures from the World Bank Worldwide Governance Indicators and then predict the probabilities for all countries using the coefficients. Finally, we regress these probabilities on GDP per capita to have the average probabilities at different income levels.[15] We use the predicted values for the United States in our baseline calibration. Although these predicted values are a little higher than actual U.S. values, we choose to use the predicted values to be consistent with our calibration in the second step for representative economies at different income levels. Appendix Section F.1 discusses alternative ways of computing job transition probabilities and compares the results.

Table 3: Moments in the Model vs Data

Moments	Data	Model
Panel A: Targeted Moments		
1. Moments: labor market		
1.1 Unemployment rate (%)	6.5	6.4
1.2 Ratio of #Vacancies to #Unemployed	0.55	0.61
1.3 Traditional sector employment share (%)	7.0	6.6
1.4 Pareto parameter of firm size distribution	1.06	1.41
1.5 % workers remaining in same firm after one quarter	0.94	0.94
1.6 % workers remaining employed after one quarter	0.97	0.97
1.7 Workers' av wage growth after job-to-job transition	0.13	0.14
1.8 % job-to-job transition from high to low wage firms	0.22	0.24
2. Moments: training intensity and value		
2.1 Average training intensity (% time)	2.2	2.2
2.2 Ratio of training intensity in firms with 100-499 employees to that with 50-99 employees	1.2	1.1
2.3 Ratio of training costs to wage costs of training	0.24	0.24
2.4 Percent wage increase of 40 years' experience (%)	88	89

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

Finally, it is important to note that the available data do 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 our economies. Thus, we first measure the average hours of

[15]Alternatively, we construct a second measure, for which we directly regress the available probabilities in Donovan, Lu and Schoellman (2020) on GDP per capita and predict the probabilities at different income levels. The resulting fitted probabilities are very similar.

formal training per worker from the data.[16] We then impute overall training intensity for every economy, relying on two assumptions according to the US SEPT survey: the average worker spends two hours in informal training for each hour spent in formal training and there are 50% more workers participating in informal training than in formal training.[17]

Table 3 shows 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 although it slightly overestimates the share of job-to-job transitions from high-to-low wage firms (0.22 in the data vs 0.24 in the model) and the Pareto distribution parameter (1.06 in the data vs 1.41 in the model). Overall, the model does really well in matching the targeted moments.

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 investment (in terms of effective hours) in producing new human capital. Its calibrated value is in the ballpark of the estimates in the literature: for instance, Imai and Keane (2004) find this parameter to be 0.22, while Manuelli and Seshadri (2014) estimate this parameter to be 0.48.

Table 4: Calibrated Parameters

Parameter	Model
c_M - Constant in matching function	0.58
c_s - Ratio of training costs per time to average wage rate	0.20
c_v - Constant in vacancy function	1.28
γ_s - Convexity of training function	0.31
γ - Traditional sector share in production function	0.31
γ_p - Convexity of leaving costs	6.79
κ - Shape parameter of Pareto productivity distribution	3.92
ζ - Constant in training function	0.05
$\underline{p}$ - Exogenous job-to-job transition	0.14
μ_F - Share of Training costs paid by the firm	0.80
δ - Exogenous separation rate	0.03
c_p - Constant in leaving costs	5.13

Note: The table lists the parameters that were determined using simulated method of moments, as described in the text, and their values in the quantitative analysis.

[16] We multiply shares of workers exposed to formal training by hours spent in 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.

[17] In the United States, 60% of workers are receiving formal training and 90% are receiving informal training.

Moreover, training a young worker for the full quarter (480 working hours) increases her hourly wage by 5%, which lies in the range of empirical studies on U.S. training returns as reviewed by Leuven (2004) and Bassanini et al. (2005).[18] For more evidence on the model calibrated parameters and model dynamics, we illustrate in Appendix Section G how the moments help identify the model's parameters by calculating the elasticity for moments to each parameter.

4.1.2 Cross-Country Calibration

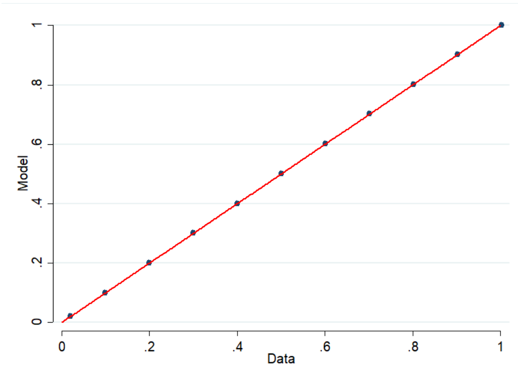
We focus our analysis on three main features that radically decrease with development: the share of workers who are self-employed, job-to-job transitions, and job destruction rates (Gollin, 2002, 2008; Donovan, Lu and Schoellman, 2020). In this section, we calibrate the model for a representative economy at 10 different income levels other than the United States' one. To do this, we keep the baseline calibrated parameters and re-calibrate δ, c_p, A_m, and A_T to match income levels, self-employment, the share of workers who stay in the same firm, and the share of workers who stay employed from quarter to quarter.

We first show how the model fits the targeted moments in Figure 8. On the x-axis we show the moments in the data, on the y-axis we show the moments in the model and we plot the 45-degree line. Overall, our model matches the targeted data moments well. We exactly match GDP per capita, self-employment shares, and the share of workers who remain employed from quarter to quarter. Moreover, the model slightly overestimates the share of workers who stay in the same firm from quarter to quarter (0.7 in the model and 0.66 in the data for the poorest economy).

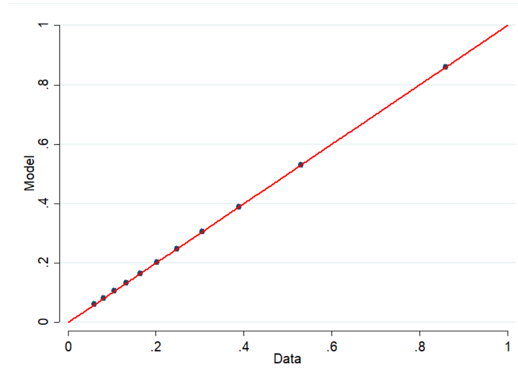
In Figure 9 we show the calibrated parameters given by the representative economy calibration. In Figure 9a we plot the level of productivity in the modern sector and in panel 9b we plot the level of productivity in the self-employment sector and, as expected, both productivity levels increase with GDP. The increase in the wage sector's productivity is faster than the self-employment sector's productivity increase, which partially shapes the decrease in self-employment with development. In Figure 9c and 9d we plot the parameters shaping the labor market dynamics. The job destruction rate δ decreases and the cost of breaking the contract c_p increases with income. These results align with larger turnover, more self-employment and worse institutional quality in developing economies.

[18] For example, using the NLSY data, Veum (1995) finds that increasing one hour of formal training improves hourly wage by 0.01%. Also using NLSY data, Frazis and Loewenstein (2005) find that 60 hours of formal training increases wage by $3 \sim 5\%$ — our calibration implies 2.7% wage growth for 60 hours of training in one quarter. The comparison with Veum (1995) and Frazis and Loewenstein (2005) is imperfect, because it is unclear whether training in their data happened within one quarter or in multiple quarters.

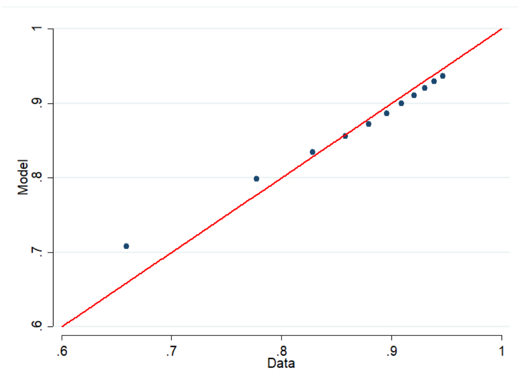
Figure 8: Cross Country Targeted Moments



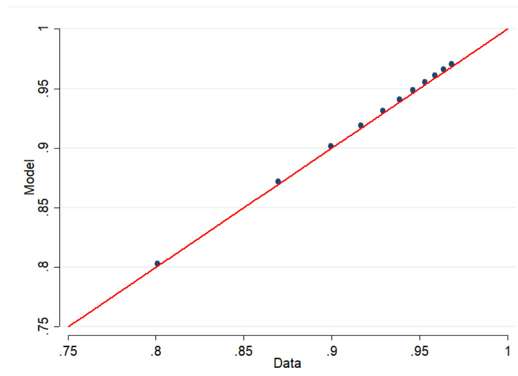
(a) GDP per capita



(b) Self employment share



(c) % workers in same firm



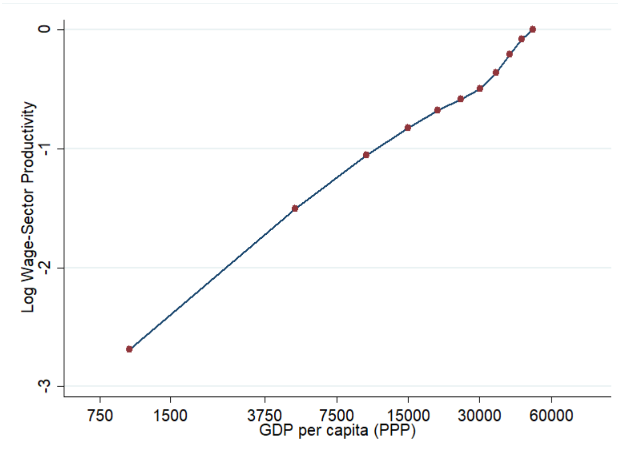
(d) % workers still employed

Note: 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 \$2,000, \$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 GDP per capita (relative to the U.S.). Panel B shows self-employment. Panel C shows the share of workers who remain in the same firm after one quarter. Panel D shows the share of workers who are employed in any firm for two consecutive quarters.

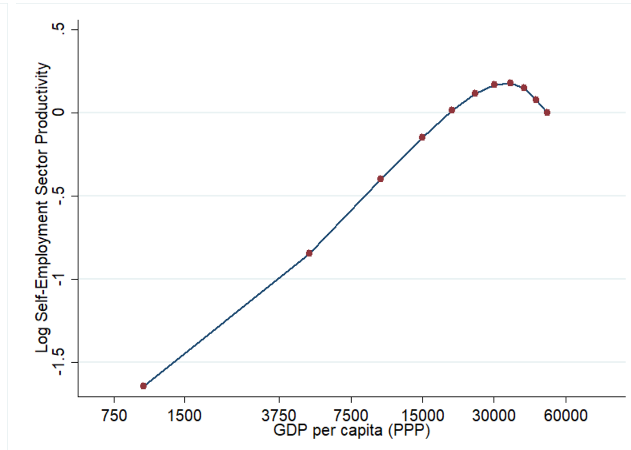
Non-targeted Moments and Model Validation. In Table 5, we compare several untargeted moments in the model to the data for the U.S. and across countries. In Panel 1, we first focus on the within-country moments. Our model implies a similar aggregate labor share and average unemployment duration as in the BLS data for the period 1994-2007. Our model has slightly higher slope of the labor share on firm market shares as estimated by Autor et al. (2020), which captures how concentration affects wage returns relative to firm revenues. Moreover, our model implies that bigger and more productive firms and firms with lower labor shares invest more in training, which aligns with our evidence in Appendix Section D.

Panel 2 in Table 5 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 the self-employment sector labor share is 1 while

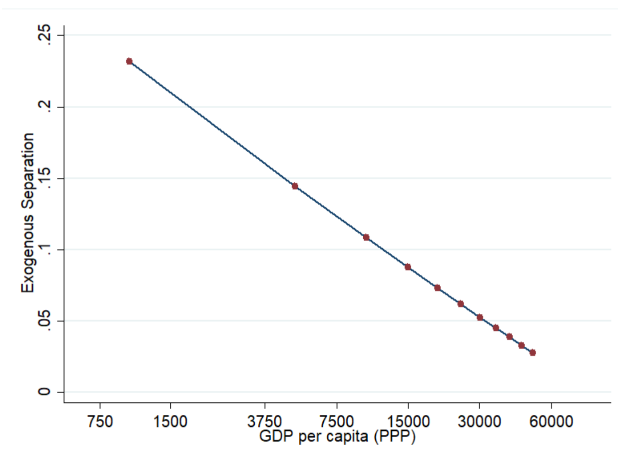
Figure 9: Cross Country Calibrated Parameters



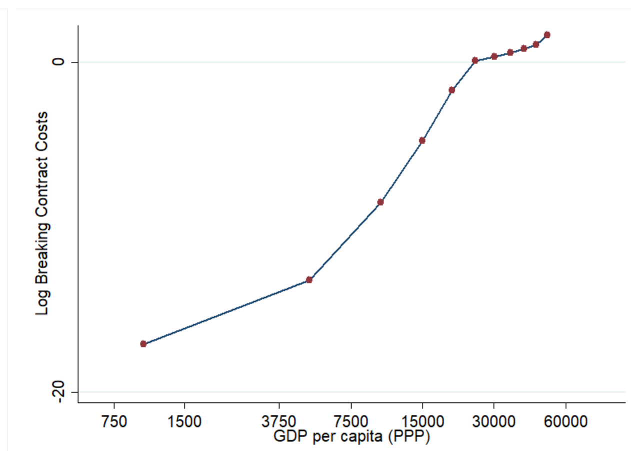
(a) Modern Sector Productivity



(b) Self-Employment Sector Productivity



(c) Exogenous Separation



(d) Breaking Contract Costs

Note: This figure shows the calibrated parameters for each economy in the model as a function of Log(GDP per capita). We consider 10 representative economies at income levels of \$2,000, \$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 in the model). Panel B shows the relative productivity between the self-employment sector and the wage sector (A_T/A_M). Panel C shows the quarterly exogenous separation rate implied by the model (δ). Panel D shows the log of the breaking contract costs, c_p.

the second measure (adjustment 2) assumes that labor share in the self-employment sector is identical to its counterpart in the wage sector. Furthermore, much of the literature shows how firms are smaller on average in developing economies and how this impacts productivity in the aggregate. We provide one informative moment of this distribution, which is the slope of the standard deviation of employment with respect to income, which is 0.1 in our model (slightly lower than 0.18, as found by Poschke (2018)).

Finally, we turn our attention to the main non-targeted moment we want to analyze. We plot the training intensity from the data and model as a function of GDP per capita in

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

Untargeted Moments	Data	Model
1. The US		
1.1 Employees' Labor Share (%)	55	63
1.2 Average unemployment duration (weeks)	17.5	15.3
1.3 Slope of labor share on firm market share	$[-2.37, -0.35]$	-0.69
2. Across countries		
2.1 Slope of labor shares on log GDPPC (adj 1)	-0.02	-0.03
2.2 Slope of labor shares on log GDPPC (adj 2)	0.02	0.05
2.3 Slope of std firm size on log GDPPC	0.18	0.10

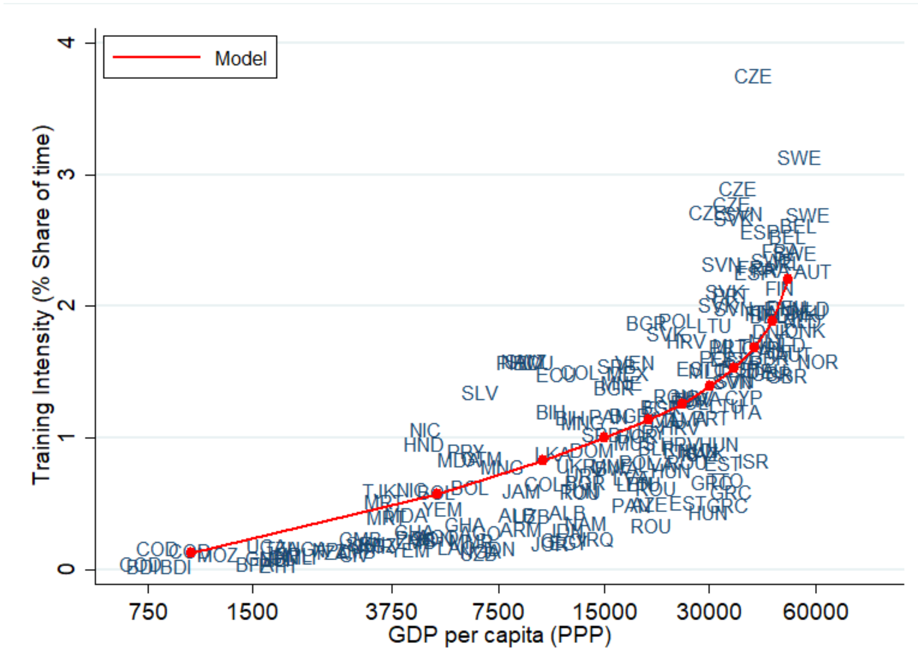
The table reports some non-targeted moments in the data and in the model. The aggregate labor share and average unemployment duration are calculated using BLS data for the period 1994-2007. The slope of the labor share on firm market shares comes from Autor et al. (2020). The measures of the labor share 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 wage increase in 20 years of experience on the logarithm of per-capita GDP comes from Lagakos et al. (2018a). The slope of the standard deviation of employment with respect to income comes from Poschke (2018).

Figure 10. The model does well in matching both the elasticity of training with respect to per-capita GDP and also the levels. Training intensity in the data may be noisier, especially for middle-income countries for which the WB-ES may overweight bigger firms and where self-employment is low enough to increase the importance of the wage sector's training intensity. Thus, it is possible that for this reason the model slightly underestimates the training intensity for middle-income countries.

These results suggest that our channels capture most of the difference in training across countries. In Appendix Table L.1 we show suggestive evidence on the correlations between training investments and job turnover measures from Donovan, Lu and Schoellman (2020), self-employment, firm size distribution, and institutional quality proxies. As we add each one of these explanatory variables, we show how the coefficient on GDP per capita decreases. Once we add the first principal component that includes all variables, we explain all the correlations between GDP per capita and training, which suggest that institutional quality, job separation, and self-employment capture most of the trend component of on-the-job training with respect to cross-country income.

These results do not imply that our channels explain all training differences across countries, but that we capture how training varies with income. From the model, it is clear that things 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

Figure 10: Training in Data and Model



Note: This graph shows the quadratic fit of the cross-country training intensity (measured in the share of time that an average worker spends in training) as a function of Los(GDP per capita). The green line represents the quadratic fit for the cross-country measure in the model and the blue line represents its counterpart in the data. The grey shadow represents the 95% confidence intervals.

L we rely on labor market institutions indexes constructed by Botero et al. (2004) to understand how the cost of firing workers, labor market institutions 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 do not account for part of the trend component between training and income. This is consistent with a result by Donovan, Lu and Schoellman (2020) that shows some labor market institutions are important determinants of cross-country variation in labor market flows but do not explain the trend relationship with respect to income.

As a robustness check, in Appendix Section H, we provide an alternative calibration in which we calibrate the model for 100 countries using country-specific levels of self-employment and training and let training productivity in the model vary across countries. We show that with this alternative calibration all our results hold.[19]

In the following sections we aim to answer three main questions: (1) how much of the wage-growth differences across countries can be accounted for with on-the-job training; (2)

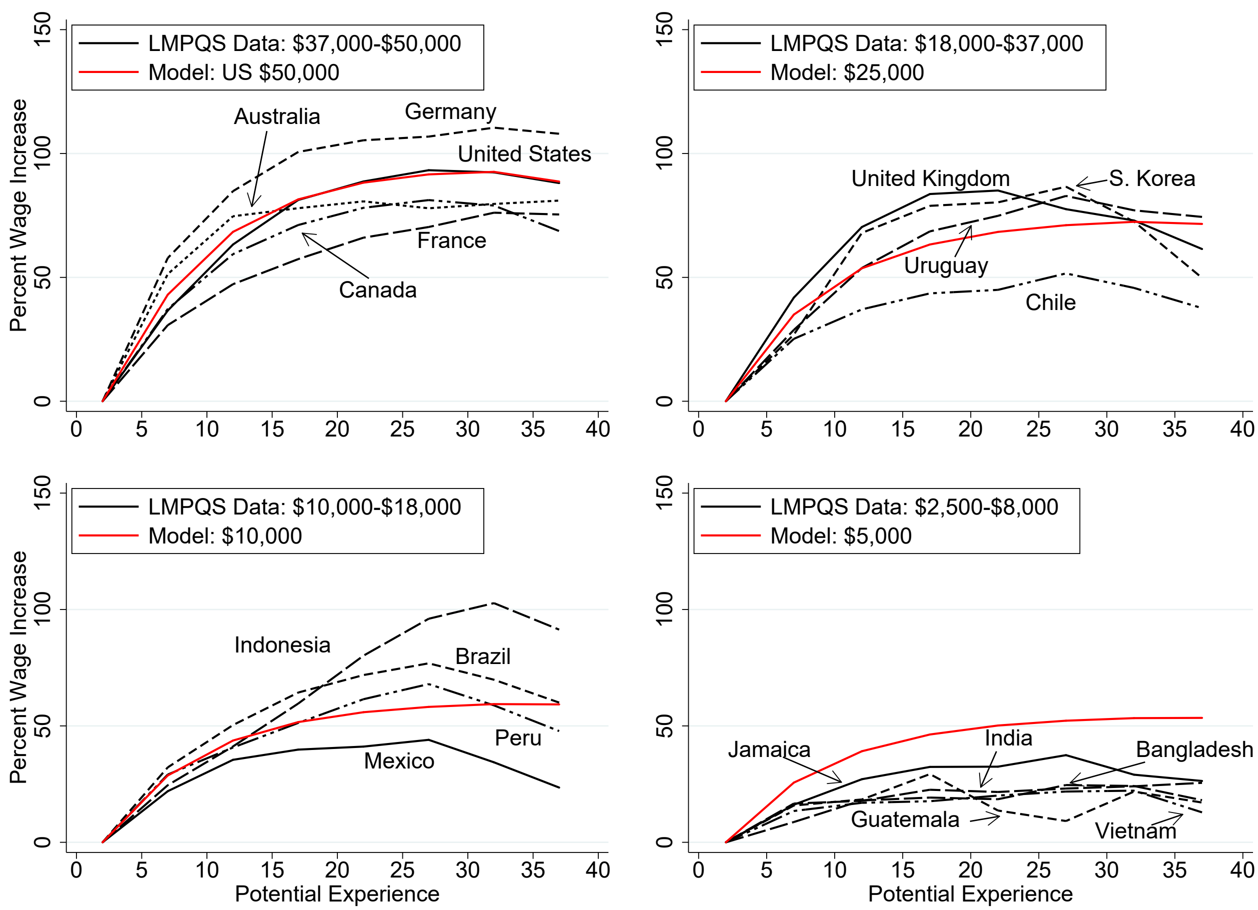
[19]By introducing country-specific training returns, we show the model matches all moments used in the baseline calibration and it also exactly matches training levels for all countries. As expected, we find that training productivity is mostly flat with income and thus there is not much unexplained by differences in training returns for the trend component of training levels.

why do developed economies invest more in training; and (3) what is the optimal training policy at different stages of development?

5 Cross-Country Wage Growth and Income Differences

In this section, we first analyze how much each of the channels in our model contributes to explaining the difference in workers' wage growth between developed and developing economies.

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

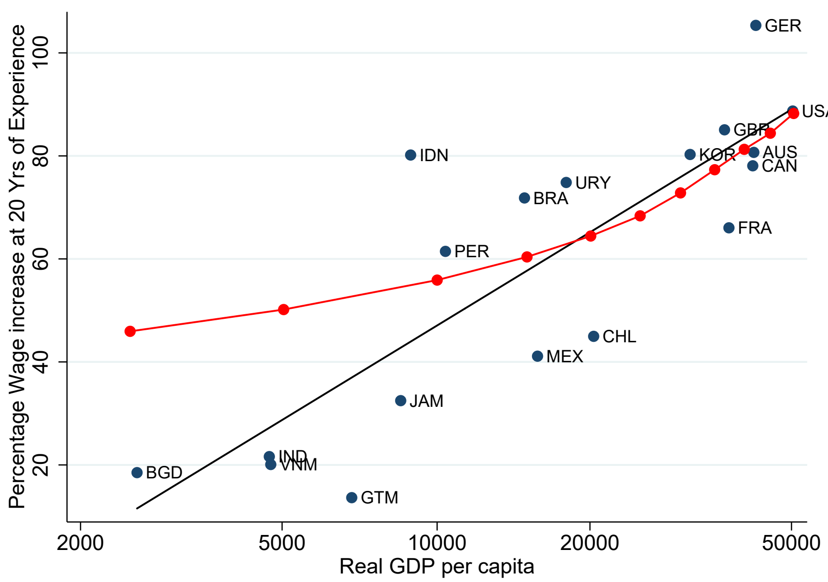


Note: This figure replicates Figure 2 from Lagakos, Moll, Porzio, Qian and Schoellman (2018b) (LMPQS) and adds the wage-experience profiles from our model in red. In the vertical axis we plot the percent increase in wages at each potential experience bin, and in the horizontal axis we plot potential experience in years. Panel a shows countries between $37,000 and $50,000 and the model outcome for the U.S calibrated economy ($50,000). Panel b shows countries between $18,000 and $37,000 and the model outcome for the calibrated economy at $25,000. Panel c shows countries between $10,000 and $18,000 and the model outcome for the calibrated economy at $10,000. Panel d shows countries between $2,500 and $8,000 and the model outcome for the calibrated economy at $5,000.

Figure 11 plots experience-wage profiles for 18 economies at all income levels from Lagakos et al. (2018a). 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 profiles well at all income levels except for the ones at the bottom of the world income distribution. The calibrated economy at \$5,000 has a steeper experience-wage profile than its counterparts in the data. This suggest that other factors that we do not include in our model may play an important role to explaining the low wage growth in these economies.

To analyze how much of the cross-country difference in returns to experience our model accounts for, we plot in Figure 12 the returns to 20 years of experience of LMPQS economies and the same measure from our model as a function of per-capita GDP. As expected, the model matches very well the wage growth for middle- and high-income countries and over-estimates the wage growth for workers in the poorest economies. Regressing the returns on log(per-capita GDP) we find a slope of 0.26 in LMPQS and a slope of 0.14 in our model which implies we capture 55% of the differences in returns to experience. Nonetheless, the model captures all of the difference for the economies above \$10,000.

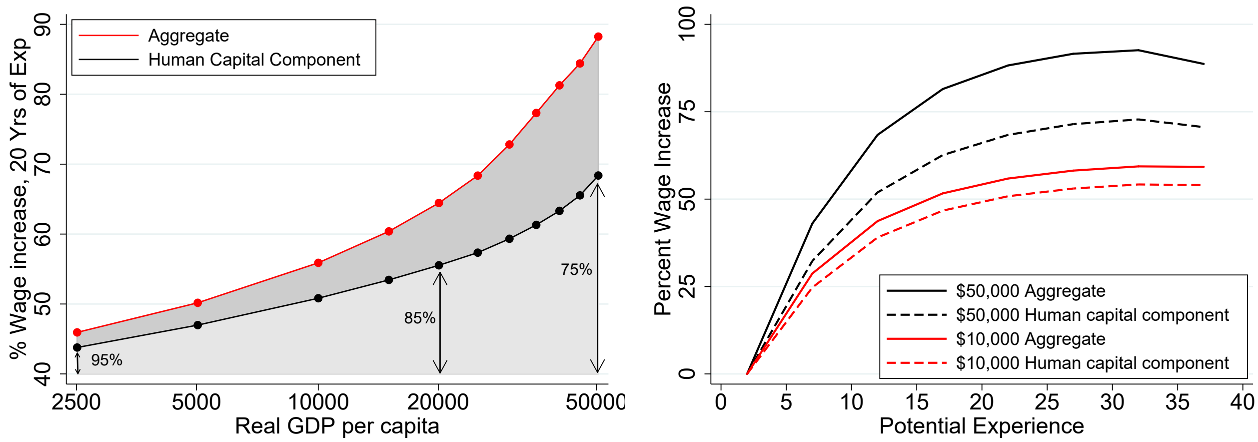
Figure 12: Cross-country Eperience-Wage Profiles: LMPQS vs Model



Note: This figure replicates Figure 3 from Lagakos, Moll, Porzio, Qian and Schoellman (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 14.

We continue by calculating how much of the wage growth is driven by firm-training and labor market dynamics. In Figure 13a, we decompose the effects on the human capital and the job-turnover components of wage-growth as a function of income. We find that the contribution of human capital is large for every economy and that it decreases with income. This happens because the high level of job destruction in the poorest economies prevents workers climbing up the job ladder. As income increases, fewer workers are separated from their jobs which generates larger increases in wages through job-to-job transitions. Moreover, Figure 13b shows the entire wage-experience profile decomposition for two economies at \$50,000 and \$10,000, which shows the same pattern. More importantly, we find that the human capital channel accounts for 75% of the difference in workers' wage growth between these economies.

Figure 13: Cross-country Eperience-Wage Profiles Composition



(a) Decomposition and Income **(b)** Profile Decomposition

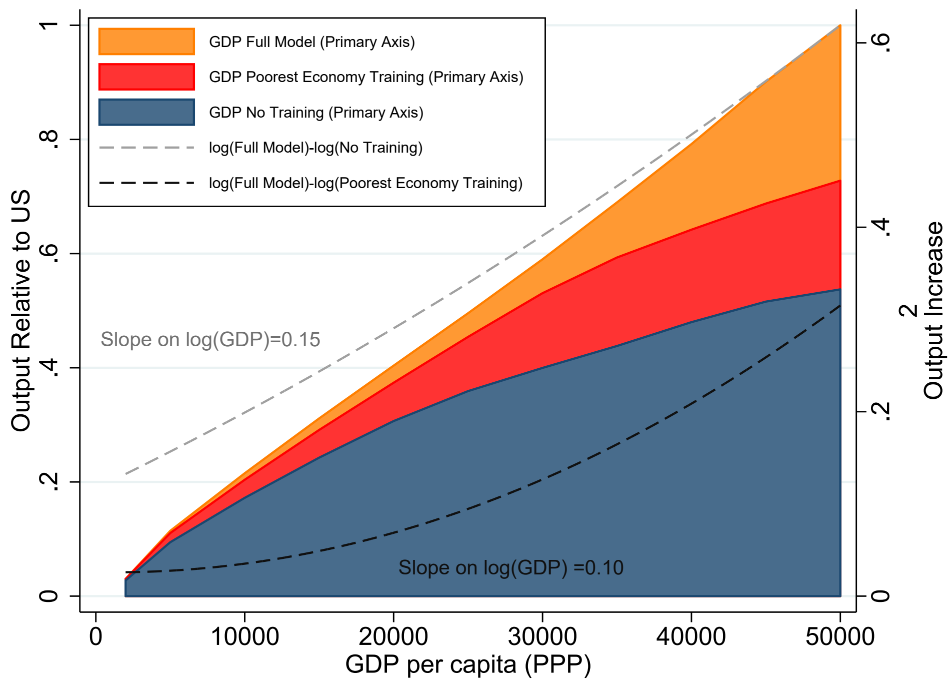
Note: This figure replicates Figure 2 from Lagakos, Moll, Porzio, Qian and Schoellman (2018b) (LMPQS) and adds the wage-experience profiles from our model in red. In the vertical axis we plot the percent increase in wages at each potential experience bin, and in the horizontal axis we plot potential experience in years. Panel a shows countries between \$37,000 and \$50,000 and the model outcome for the U.S calibrated economy (\$50,000). Panel b shows countries between \$18,000 and \$37,000 and the model outcome for the calibrated economy at \$25,000. Panel c shows countries between \$10,000 and \$18,000 and the model outcome for the calibrated economy at \$10,000. Panel d shows countries between \$2,500 and \$8,000 and the model outcome for the calibrated economy at \$5,000.

We now focus on productivity differences driven by this channel. 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 in Figure 14. In orange we plot the original model, in blue we plot the model with no training, and in red we plot a case where all economies have the same training investments as the poorest economy.[20] When there is no training output is the lowest. Output increases when we add the poorest economy level of training to the model with no training, and increases even more when we endogenice training, which reflects the fact that training boosts productivity in the aggregate. The heterogenous increase in output with respect to income suggest that adding training improves output more in developed economies than in developing economies.

Thus we ask: what is the share of the income differences across countries explained directly by training in our model? To answer this question, we plot the difference between

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

Figure 14: Income Increase due to Training



Note: This figure shows the percentage increase in output from training calculated as the log change in output from the model shutting down training (increasing c_s to an extremely large value) to the full model as a function of GDP per capita. Each observation comes from using the calibrated version of the model for each country. Data from GDP per capita comes from the World Bank Indicators. The slope of 0.16 represents the share of the increase in GDP per capita explained by training in the model.

the log(per-capita GDP) in the full model and its counterpart from the two models with no training and with the poorest economy training. We plot the fitted values in the secondary axis of Figure 14. This difference represents the percentage increase in output from the model with no training (or low training) relative to the full model. The slope of the percentage increase in output from the model with no training to the full model on log(per-capita GDP) provides the share of the income differences explained directly by training in our model. Our quantitative model suggests that on-the-job training explains 15% of the income differences across countries, which is sizeable.[21] Moreover, doing the same exercise but using the model with the poorest economy's training level for all economies, the model generates an income difference coming from training of 10%, which reflects a lower bound on how training explains income differences in the model.

In Appendix Section J we do some robustness checks for these results by using the calibration for all 100 countries. We construct two measures of per-capita GDP from the

[21] Lagakos et al. (2018a) shows that adding returns to experience helps explain around 20% of the income differences across countries, and thus, our result suggests that firm training is one of the main sources of workers' human capital accumulation post-schooling.

model, one with the full model that matches the data and a second from simulating the model shutting down training. We plot the increase in GDP driven by on-the-job training in the model as a function of per-capita GDP in Figure J.1 and we get similar results.

6 Training Decomposition

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

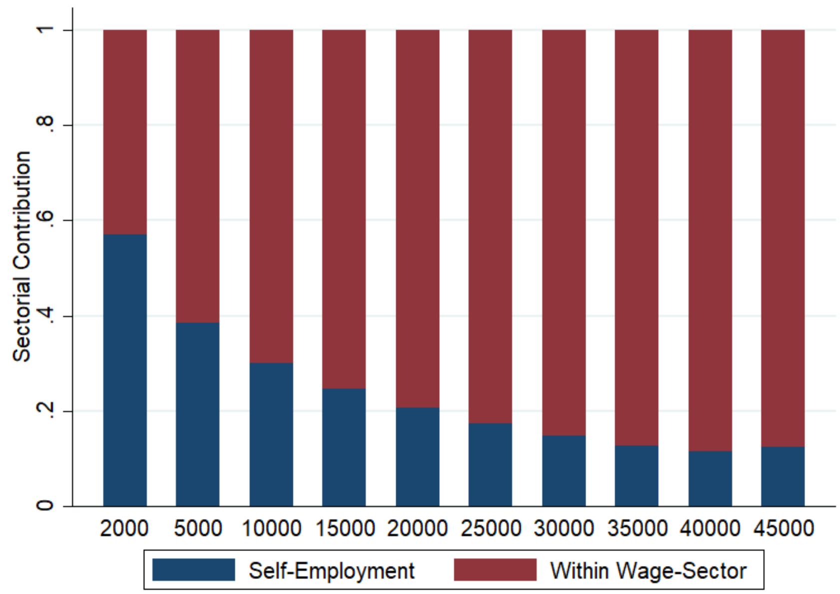
From the results on training intensity and self-employment shares, we first analyze how the sectorial allocation explains training differences in the aggregate. Denoting S as the training investment in the modern sector and T as the wage-sector employment share, the difference in training can be simply decomposed as:

$$log(S_{us}T_{us}/S_{base}T_{base}) = log(S_{us}T_{us}/S_{us}T_{base}) + log(S_{us}T_{base}/S_{base}T_{base})$$

The first term reflects the training increase due to the change in the self-employment share and the second term represents the increase in training in the modern sector, conditional on the sectoral allocation. We plot these two components in Figure 15. We find that the role of self-employment in explaining training differences decreases substantially with development. It goes from 60% in the poorest economies to 10% in the richest ones. This result is very intuitive; aggregate human capital will be low if there is a small proportion of workers in the modern sector, even if training levels are large in that sector. This result offers strong policy implications as this channel amplifies the productivity increases from reallocating workers away from self-employment studied in the structural transformation and migration literature.

There are 4 parameters that vary across country: δ, which shapes job destruction; c_p, which shapes job-to-job transitions; and A_T and A_M, which shapes income and self-employment shares. We proceed to do a factor-decomposition driven by these parameters. We explain the procedure through one example. Let's assume we want to capture how each parameter contributes to explaining the training gap between the United States and the poorest economy. We first start from the U.S. calibration and change c_p to its poorest economy counterpart and we measure the change in training. Then, we start from the U.S. calibration, but with the poorest economy's δ and change c_p to the poorest economy value again and compute the new change in training. We repeat this exercise for all possible combinations to calculate the average change in training coming from c_p. Then, we do the same

Figure 15: Training Decomposition by Sectoral Component



Note: This figure shows how (1) changing the self-employment share while keeping training in the wage sector fixed, and (2) changing the wage sector training level while keeping the self-employment share fixed contribute to explaining the difference in training between each economy and the U.S.

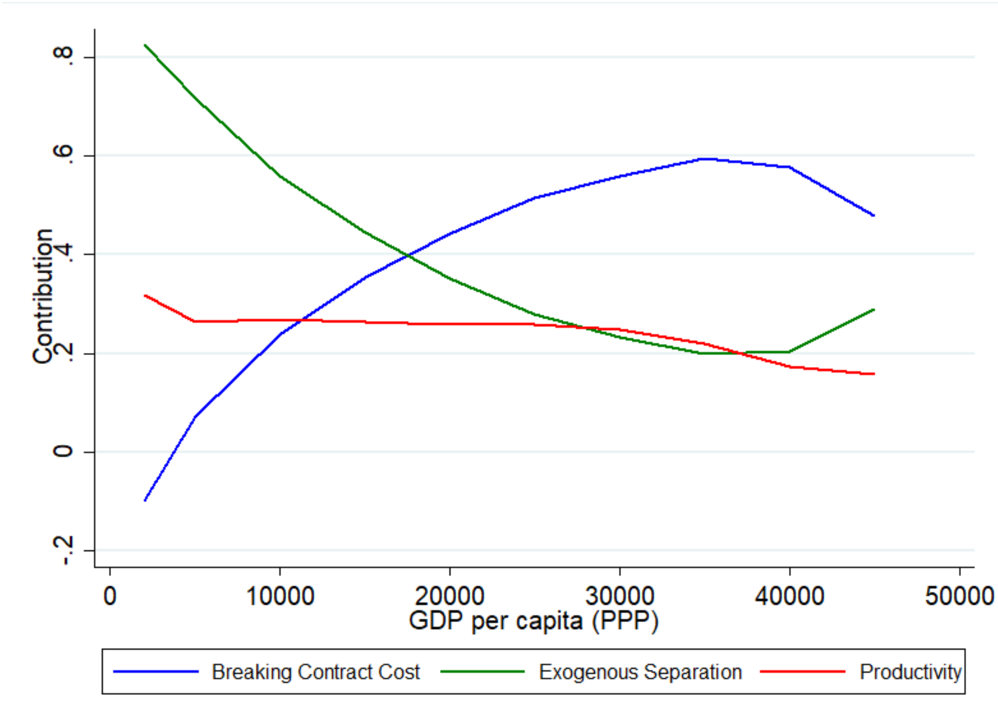
simulations to calculate the average change in training from the other parameters.[22] After these calculations we have the result on how each parameter contributes to explaining the training gap between the U.S. and the poorest economy. For our factor-decomposition, we repeat this procedure comparing the U.S. with the economy at each income level and plot the results in Figure 16.

The importance of job destruction, reflected in the exogenous separation, tends to decrease with income while the importance of the labor market friction coming from the cost of breaking contracts increases with income. On the other hand, the contribution from the change in productivity is roughly constant (around 20%). The first important result suggests that labor market frictions explain 80% of the differences in training across countries. The second take away is that, for poor economies, the most important channel to explain the lack of training is job destruction, but as income increases the difference in training comes largely from frictions in job-to-job transitions.

To dig deeper into the mechanisms driving these results, in the Appendix Figure I.1, we

[22]Note that there are 6 ways of doing these changes (from the baseline, we might first change c_p, then δ and finally A_T and A_M, or we might first include δ and then the other ones in another order, etc.). These different ways of calculating the change in training may provide different results depending on the complementarity between the mechanisms. Therefore, we simulate the model following all possible orders and calculate the average increase in training coming from each parameter in all simulations.

Figure 16: Share of Training Gap Covered by each Parameter Change



Note: This figure shows the contribution of each channel to explaining the training gap between the economies at each income level and the U.S. The green line presents the contribution from changing δ; the red line presents the contribution from changing A_t and A_M simultaneously; and the blue line presents the contribution from c_p.

decompose each channel into the wage sector and self-employment contributions:

$$log\left(\frac{S_i T_i}{S_{base} T_{base}}\right) = log\left(\frac{S_i T_{base}}{S_{base} T_{base}}\right) + log\left(\frac{T_i}{T_{base}}\right)$$

where S represents the wage-sector training level and T represents the share of workers in the wage sector, both by changing parameter i. The effect coming from changing sector-specific productivities only impacts the share of workers in each sector but does not generate changes in training levels within the wage sector (relative training cost and revenue are invariant across different income levels).[23] The effect coming from c_p and δ are driven mostly by the change in self-employment for poor economies and by changes in firm-level investments in training for richer economies.

Moreover, to understand the model's dynamics, we go one step further and decompose the increase in training within the wage sector into the partial equilibrium and the general equilibrium effects. The partial equilibrium effect represents firms' training level in the new parameter scenario while keeping the wage and employment distributions fixed. The

[23]The wage sector changes are not exactly zero because we solve the model numerically and the convergence may not be perfect in every case and some small noise can be reflected in the results.

general equilibrium effect represents the wage sector's extra increase in training level (letting the wage and employment distribution change). Thus, we plot the decomposition from the change in the exogenous job destruction and the job-to-job transition friction in Appendix Figure I.2.

In partial equilibrium we always observe the expected signs. As c_p increases and δ decreases, the probability of keeping workers goes up, increasing training investments. Nevertheless, the effects from general equilibrium may be negative. Lower separation to unemployment generate fewer unemployed workers in the pool of searchers, and thus, firms must post higher wages to attract workers. Moreover, lower probability of job destruction means workers move to more productive firms faster because they fall to unemployment in fewer occasions. In this case, higher wages pull training investments down while the employment distribution shift towards more productive firms pushes training up. The negative GE effect observed reflects the fact that the wage effect predominates.

Moreover, higher costs to break contracts through higher c_p means firms keep workers longer and it is more difficult to poach workers from other firms. These effects push wages downward and this encourages firms to invest more in training as they capture a higher share of the surplus. Although fewer job-to-job switches shifts the employment distribution towards less productive firms (which would generate a decrease in training) the wage effect predominates.

There are 3 main takeaways from the training decomposition. First, most of the training gap between the poorest and richest economies is explained by differences in self-employment shares. This is driven by the lack of training in the self-employment sector, and the high rates of self-employment prevalent in poor economies arising from the endogenous reallocation of workers from the wage sector to the self-employment sector as a result of high labor market frictions in the former sector. Second, these labor market frictions remain key to explaining training investments as countries develop, and self-employment shares fall. The mechanism driving this is worker turnover in the wage sector. In particular, high job separation rates and low contract quality make worker turnover more likely, and thus, depress the incentives to invest in training in medium-income economies relative to richer economies. Collectively, these two facts indicate that labor market frictions are key to explaining training differences across countries of different levels of development. Third, 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 to explain the lack of training in poorer economies while frictions in job-to-job transitions are more important to explain the training differences between more developed economies.

7 Cross-Country Optimal Policy

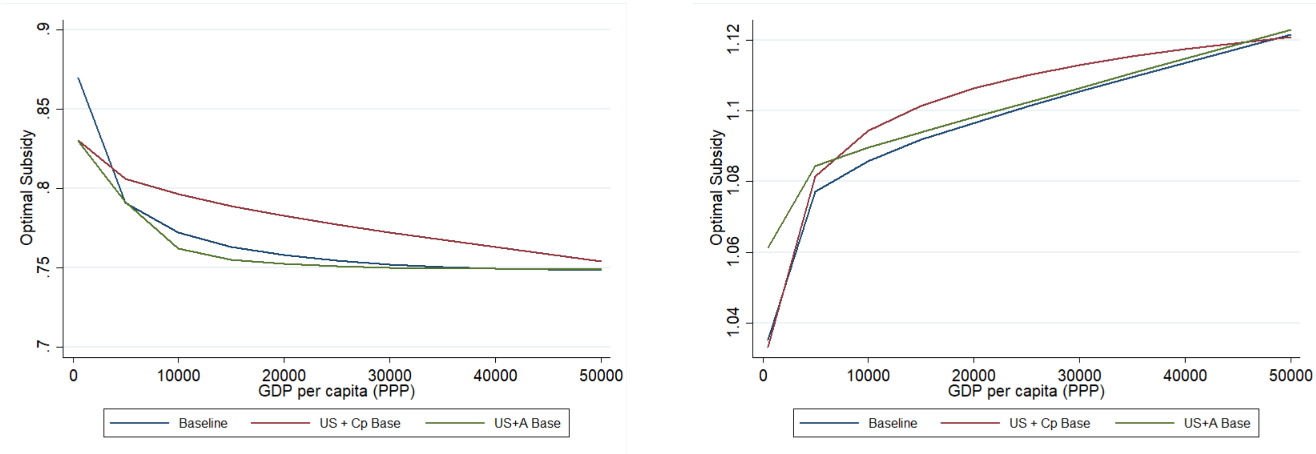
In this section, we discuss the inefficiencies present in our framework, show empirical evidence on the existence of policies aimed at increasing training around the world and conduct an optimal policy analysis.

Our model has three sources of inefficiencies in training investments. The optimal level for each party (workers and firms) depends on their own training marginal returns and on the share of the costs each pays. Therefore, there are a couple considerations worth noting. First, firms and workers do not internalize the benefits from training investments for the other party. If $s_F < s_W$ ($s_F > s_W$) the optimal level of training will be given by s_F (s_W), and thus, the MR_W (MR_F) will not be fully internalized. In Appendix Section M we show the training patters for firms, workers, and the aggregate, and show that the firm decision is the relevant margin as firms always want lower levels of training in the calibration that matches all labor market dynamics. Second, even if there is a joint decision and the optimal investment level is the same for both firms and workers, another source of inefficiency still exists. Workers and firms never internalize the benefits from training for future employers if separation occurs (Acemoglu, 1997). Therefore, in the model, higher output in future employers and the full marginal benefit for workers are not internalized, which means there is space for policies to improve welfare.

In Appendix Section K we do an extensive description of government subsidy policies around the world and show that training subsidies are indeed very common. We review government policies to incentivize employer-provided training from 36 countries from all continents and income levels and 21 U.S. states. The most commonly observed policies are direct reimbursements or employers' tax deductions for a share of the direct training costs. Nevertheless, in many cases, there are monetary incentives or subsidies for part of the wages of workers who are trained. There are a wide variety of training policies or subsidies ranging from 5% - 10% of training costs in Mozambique, or \$20 - \$100 in India, to 120% of training costs in Austria, or \$5,415 in New Zealand. For, U.S. states, on the other hand, the most common policy is a reimbursement for about 75% of the training cost. The large set of government programs described suggests governments all around the world consider training an important channel to improve productivity, and that training investments are probably inefficient.

Motivated by the existence of training investment inefficiencies and by the empirical evidence showing training subsidies are very common, we perform an optimal policy analysis. We simulate the model with the baseline calibration, adding a subsidy to the total training costs, which includes the direct costs of training and the workers' wages for the full training

Figure 17: Optimal Training Subsidy and Development



(a) Optimal Subsidy Rate

(b) Net Output Increase

Note: This figure shows the results from simulating the model adding a subsidy to the total training costs (direct costs of training and workers' wages for full training length), financed with lump-sum taxes. Panel (a) shows the optimal subsidy at each income level, and panel (b) shows the increase in net output driven by the policy. The blue line presents the results from the policy at each income level with the calibrated parameters of that economy. The red and green lines present the results from the optimal subsidy using the U.S. baseline calibration and only changing one parameter at a time to the value from each income level representative economy.

length. We assume the training program is financed with lump-sum taxes to agents and we analyze how output (net of training subsidy costs and vacancy costs) changes for different rates. In Figure 17 we plot the optimal subsidy rate and the net output increase for the optimal policy at different income levels. The blue line represents the optimal policy for each representative economy with all its calibrated parameters. The U.S. benchmark case shows that net output increases by 12% when it is maximized with a subsidy of 75%. Moreover, as countries become less developed, the optimal subsidy must be larger, although the increase in output is lower.

The other lines in Figure 17 represent the U.S. baseline calibration changing just one parameter at a time to the calibrated one for each income level. For instance, the red line at \$40,000 reflects the baseline calibration for the United States with the c_p from the \$40,000-calibrated economy. We do this exercise to show the fact that the decrease in the optimal subsidy rate with income comes partially from all our channels. Thus, we show that subsidies should be larger in developing economies when self-employment or job destruction is higher or when contract quality is lower.

8 Conclusion

Human capital accumulation plays a key role in economic growth and development. While recent research has highlighted the importance of on-the-job human capital accumulation in explaining cross-country income differences, how workers accumulate human capital on-the-job at different stages of development is still underexplored. In this paper, we study one key source of on-the-job human capital accumulation, namely firm-provided training. We exploit rich data sources to show that firm-provided training increases with development and that this happens in every margin of training. Moreover, we find that firm-provided training is the most important source of human capital investments in workers' careers. Then, we build a GE search model with firm heterogeneity and training investments that help us identify the mechanisms mediating these facts.

Our results have strong policy implications. A simple sectoral decomposition suggests that self-employment is key in explaining the lack of on-the-job training in the poorest economies. Thus, our channel amplifies the productivity increases from reallocating workers away from self-employment studied in the structural transformation and migration literature. Furthermore, by digging deeper into the channels generating the differences in training, we find that the high levels 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 imply that to increase productivity and improve net output, policies to improve match quality between firms and workers may be desirable in developing economies while policies to generate better contracts may be more beneficial in richer countries. Finally, our model predicts considerable inefficiencies in human capital investments and sizeable aggregate gains from training subsidies to firms. More importantly, training subsidies to firms may be particularly beneficial in poor countries where economic environments disincentivize training.

This paper finds that employer-provided training explains 15% of the income differences across countries. The importance of on-the-job training could be even larger if there are complementarities with other sources of human capital, such us schooling or co-worker spillovers. A fruitful area for future research is to study how different human capital accumulation sources interact with each other and what the implications of these interactions are to conduct more efficient public policy for countries at different stages of development.

References

Acemoglu, Daron. 1997. "Training and Innovation in an Imperfect Labour Market." Review of Economic Studies, 64(3).

Acemoglu, Daron, and Fabrizio Zilibotti. 2001. "Productivity Differences." Quarterly Journal of Economics, 116(2).

Acemoglu, Daron, and Jorn-Steffen Pischke. 1998. "Why Do Firms Train?" The Quarterly Journal of Economics, 113(1): 79–119.

Acemoglu, Daron, and Jrn-Steffen Pischke. 1999. "Beyond Becker: Training in Imperfect Labour Markets." The Economic Journal, 109.

Acemoglu, Daron, Simon Johnson, and James A Robinson. 2005. "Institutions as a fundamental cause of long-run growth." Handbook of economic growth, 1: 385–472.

Alon, Titan. 2017. "Earning more by doing less: Human capital specialization and the college wage premium." Unpublished Manuscript, Northwestern University.

Alon, Titan, and Daniel Fershtman. 2019. "Gradual Specialization and the Allocation of Talent."

Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen. 2020. "The fall of the labor share and the rise of superstar firms." The Quarterly Journal of Economics, 135(2): 645–709.

Axtell, Robert L. 2001. "Zipf Distribution of U.S. Firm Sizes." Science, 293.

Bagger, Jesper, François Fontaine, Fabien Postel-Vinay, and Jean-Marc Robin. 2014. "Tenure, experience, human capital, and wages: A tractable equilibrium search model of wage dynamics." American Economic Review, 104(6): 1551–96.

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

Bassanini, Andrea, Alison Booth, Giorgio Brunello, Maria De Paola, and Edwin Leuven. 2005. "Workplace Training in Europe." IZA Working Paper.

Becker, Gary. 1964. Human Capital. The University of Chicago Press.

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

Bick, Alexander, Nicola Fuchs-Schuendeln, and David Lagakos. 2018. "How do Hours Worked Vary With Income? Cross-Country Evidence and Implications." American Economic Review, 108(1).

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

Botero, Juan C, Simeon Djankov, Rafael La Porta, Florencio Lopez-de Silanes, and Andrei Shleifer. 2004. "The regulation of labor." The Quarterly Journal of Economics, 119(4): 1339–1382.

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

Bunzel, Henning, Bent J Christensen, Nicholas M Kiefer, Lars Korsholm, et al. 1999. Equilibrium search with human capital accumulation. Vol. 99, Centre for Labour Market and Social Research.

Burdett, Kenneth, and Dale T. Mortensen. 1998. "Wage Differentials, Employer Size, and Unemployment." International Economic Review, 39(2).

Burdett, Kenneth, Carlos Carrillo-Tudela, and Melvyn G Coles. 2011. "Human capital accumulation and labor market equilibrium." International Economic Review, 52(3): 657–677.

Caselli, Francesco. 2005. "Accounting for cross-country income differences." Handbook of economic growth, 1: 679–741.

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

De Philippis, Marta, and Federico Rossi. 2020. "Parents, schools and human capital differences across countries." Journal of the European Economic Association, Forthcoming.

Dix-Carneiro, Rafael, Pinelopi Goldberg, Costas Meghir, and Gabriel Ulyssea. 2019. "Trade and Informality in the Presence of Labor Market Frictions and Regulations ." Working Paper.

Doepke, Matthias, and Ruben Gaetani. 2020. "Why Didn't the College Premium Rise Everywhere? Employment Protection and On-the-Job Investment in Skills." National Bureau of Economic Research.

Donovan, Kevin, Jianyu Lu, and Todd Schoellman. 2020. "Labor Market Flows and Development." Working Paper, , (4).

Engbom, Niklas. 2020. "Labor Market Fluidity and Human Capital Accumulation."

Erosa, Andres, Tatyana Koreshkova, and Diego Restuccia. 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. Jason, Andreas I Mueller, Aysegul Sahin, and Giorgio Topa. 2017. "Job Search Behavior among the Employed and Non-Employed." NBER Working Paper.

Feng, Ying, David Lagakos, and James Rauch. 2018. "Unemployment and Development." NBER Working Paper.

Flinn, Christopher, Ahu Gemici, and Steven Laufer. 2017. "Search, matching and training." Review of Economic Dynamics, 25: 260–297.

Frazis, Harley, and Mark A. Loewenstein. 2005. "Reexamining the Returns to Training: Functional Form, Magnitude, and Interpretation." The Journal of Human Resources, 40(2).

Gollin, Douglas. 2002. "Getting income shares right." Journal of political Economy, 110(2): 458–474.

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

Gregory, Victoria. 2019. "Firms as learning environments: Implications for earnings dynamics and job search." Unpublished Manuscript.

Hall, Robert E, and Charles I Jones. 1999. "Why do some countries produce so much more output per worker than others?" The quarterly journal of economics, 114(1): 83–116.

Haltiwanger, John, Henry Hyatt, Lisa B. Kahn, and Erika McEntarfer. 2018. "Cyclical Job Ladders by Firm Size and Firm Wage." American Economic Journal: Macroeconomics, 10.

Hanushek, Eric A, and Ludger Woessmann. 2012. "Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation." Journal of economic growth, 17(4): 267–321.

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

Hsieh, Chang-Tai, and Peter J. Klenow. 2009. "Misallocation and manufacturing TFP in China and India." Quarterly Journal of Economics, 124(4).

Hsieh, Chang-Tai, and Peter J Klenow. 2010. "Development accounting." American Economic Journal: Macroeconomics, 2(1): 207–23.

Imai, Susumu, and Michael Keane. 2004. "Intertemporal Labor Supply and Human Capital Accumulation." International Economic Review, 45.

Islam, Asif, Remi Jedwab, Paul Romer, and Daniel Pereira. 2018. "Returns to Experience and the Misallocation of Labor."

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

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

Lagakos, David, Benjamin Moll, Tommaso Porzio, Nancy Qian, and Todd Schoellman. 2018a. "Life-cycle human capital accumulation across countries: lessons from US Immigrants." Journal of Human Capital, 12(2): 305–342.

Lagakos, David, Benjamin Moll, Tommaso Porzio, Nancy Qian, and Todd Schoellman. 2018b. "Life cycle wage growth across countries." Journal of Political Economy, 126(2): 797–849.

Leuven, Edwin. 2004. "A review of the wage returns to private sector training." Working Paper.

Manuelli, Rodolfo E, and Ananth Seshadri. 2014. "Human capital and the wealth of nations." American economic review, 104(9): 2736–62.

Moen, Espen, and Asa Rosén. 2004. "Does Poaching Distort Training?" The Review of Economic Studies, 71(4).

Mortensen, Dale, and Christopher Pissarides. 1994. "Job creation and job destruction in the theory of unemployment." Review of Economic Studies, 61(3).

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

Porzio, Tommaso. 2017. "Cross-Country Differences in the Optimal Allocation of Talent and Technology." Working Paper.

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

Restuccia, Diego, and Richard Rogerson. 2013. "Misallocation and productivity." Review of Economic Dynamics, 16(1): 1–10.

Rubinstein, Yona, and Yoram Weiss. 2006. "Post schooling wage growth: Investment, search and learning." Handbook of the Economics of Education, 1: 1–67.

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

Schoellman, Todd. 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).

Veum, Jonathan R. 1995. " Sources of Training and Their Impact on Wages." Industrial and Labor Relations Review, 48(4).

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

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

Appendix

A Data Sources

We rely on enterprise and workers' surveys for developed and developing economies. We rely on the World Bank Enterprise Survey for developing economies and we rely on European Union data for developed economies. We use the EU Labor Force Survey (EU-LFS), the Adult Education Survey (EU-AES) and the Continuing Vocational Training Survey (EU-CVT). Moreover, we use other data sources to test some implications from our quantitative model. With this purpose, we rely on the Chinese Industrial Census (an administrative dataset for all manufacturing firms in China) to test the correlation between firms' features and training investments. Finally, we rely on the World Bank Worldwide Governance Indicators to have measures of institutional quality, data from Botero et al. (2004) to proxy labor market indicators, and on data from Donovan, Lu and Schoellman (2020) to have measures of job destruction and job-to-job transitions to test cross-country correlations. To focus on the main robustness checks and supportive evidence we delegate to the Online Appendix all the deep details about the data sources used in this study, where we describe how each survey was conducted, the countries and years in each sample, the definitions of variables, and the references for the public and private sources of the data.

B Detailed Definitions on Educational Sources

Schooling: Formal education and training according to the International Standard Classification of Education 2011 (ISCED 2011) is defined as: "education that is institutionalized, intentional and planned through public organizations and recognized private bodies and in their totality constitute 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: 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 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 in total or 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 employed holding an apprenticeship or training contract should not be considered for CVT. Random learning and initial vocational training are explicitly excluded and measured separately. These courses are typically separated from the active workplace (e.g., the classroom or training institution), show a high degree of organization by a trainer, and its content is designed for a group of learners (e.g., a curriculum exists).

As defined by the CVT survey, "Other forms of CVT" that we refer to as *informal training*, have the purpose of 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 is often tailored according 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 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."

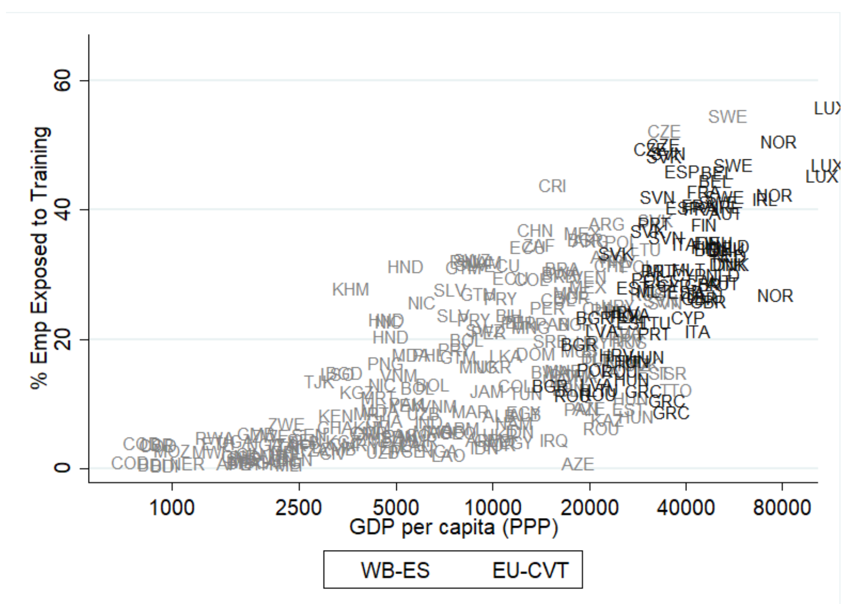
5. Participation in conferences, workshops, trade fairs and lectures: “Participation (instruction received) in conferences, workshops, trade fairs and lectures are considered as training actions only when they are planned in advance and if the primary intention of the person employed for participating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

Initial vocational training is defined as a formal education program (or a component of it) 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 initializing a job.

Informal learning: It 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.”

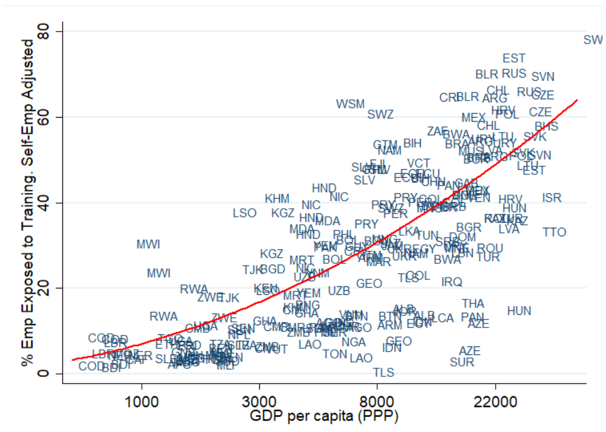
C Empirical Results: Additional Tables and Graphs

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

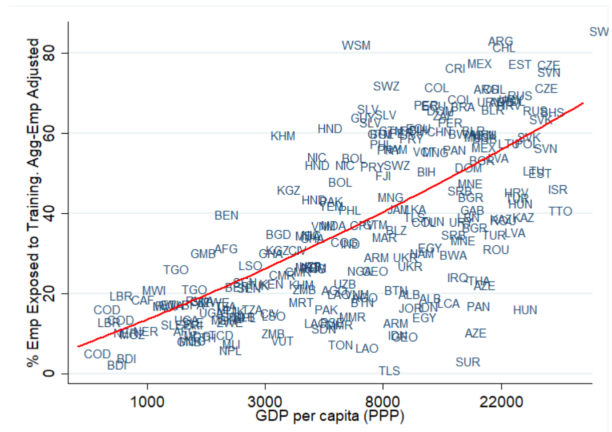


Note: The share of employment formally trained comes from adjusting the share of workers who were trained by firms by the share of self-employment. Data on share of employees trained inside the firms comes from the World Bank Enterprise Survey for all developing economies and from EU Continuing Vocational Training Survey for European countries. Both surveys ask if the firm 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 World Bank Indicators.

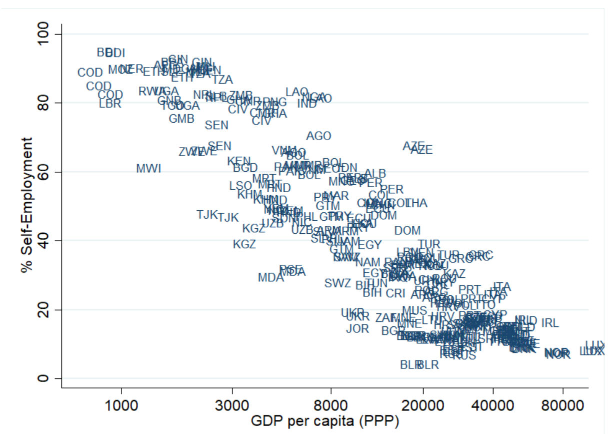
Figure C.2: Share of Workers Exposed to Firms that offer Training



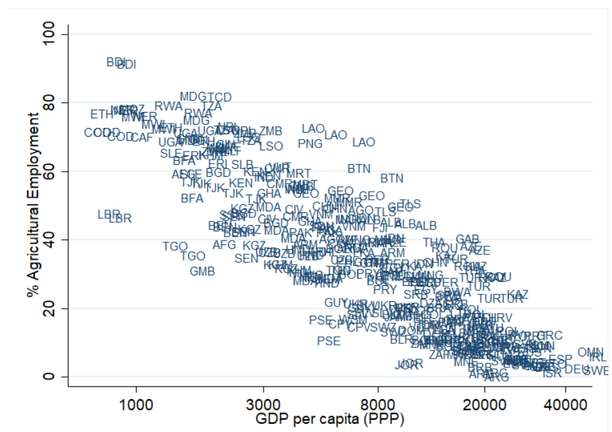
(a) Adjusted by Self Employment



(b) Adjusted by Agricultural Employment



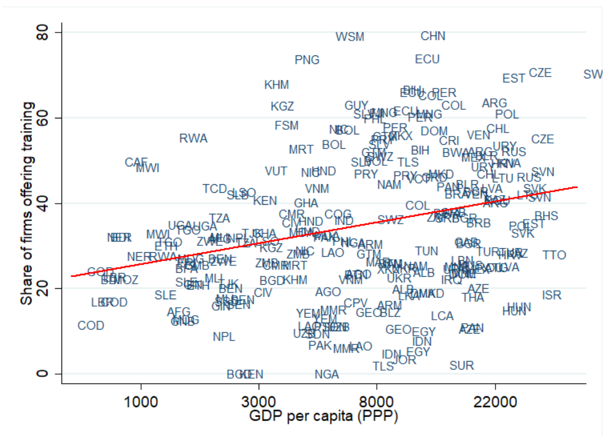
(c) Share of Self-Employed Workers



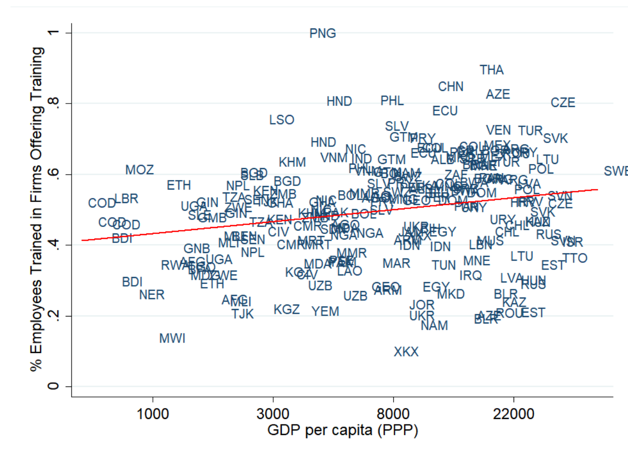
(d) Share of Agricultural Workers

Note: The share of employment trained in Panel a come from adjusting the share of workers who were trained in the wage-sector by the share of self-employment plotted in Panel c. The share of employment trained in Panel b comes from adjusting the share of workers in the wage-sector by the share of agricultural workers plotted in Panel d. Data on share of employees trained inside the firms comes from the World Bank Enterprise Survey for all developing economies. Data on self-employment and the share of agricultural workers come from the World Bank Indicators. Data on GDP per capita comes from the PWT.

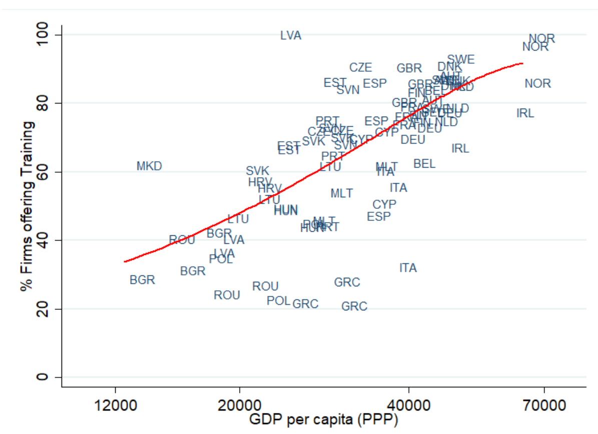
Figure C.3: Intensive and Extensive Margin



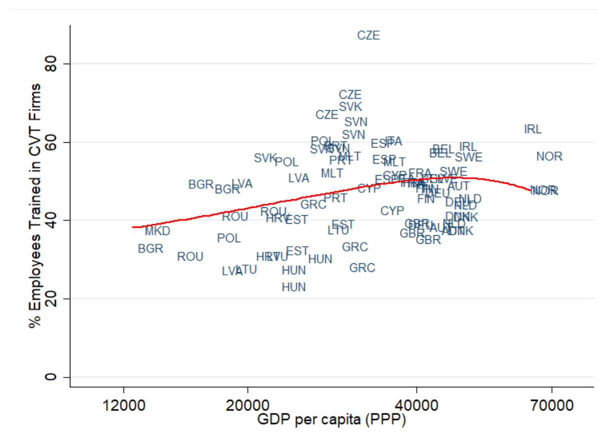
(a) %Firms Offering Training (WB-ES)



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



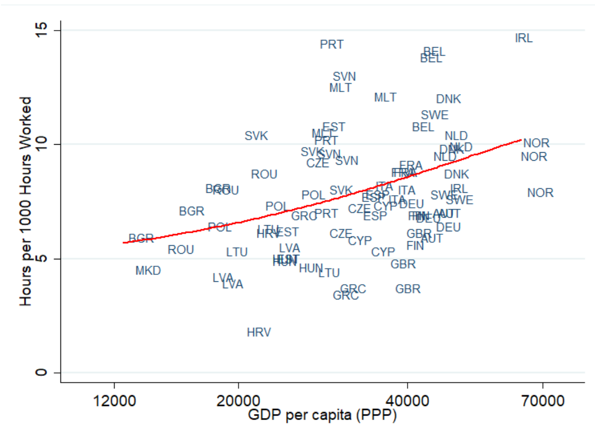
(c) %Firms Offering Training (EU-CVT)



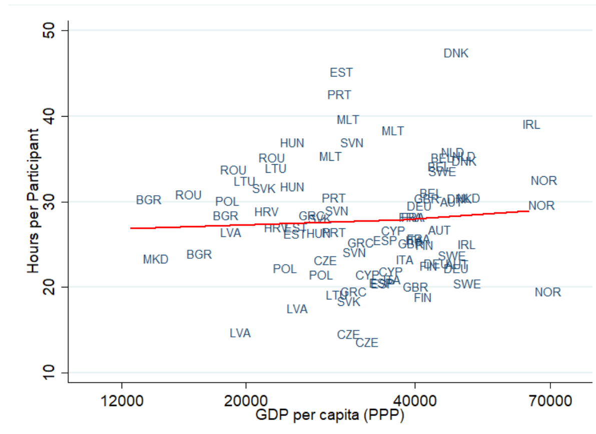
(d) %Trained Workers per Firm (EU-CVT)

Note: These figures show the share of firms offering training and the share of workers trained for the World Bank Enterprise Survey and the European Union Continuing Vocational Training Survey. Panel a shows the share of firms and Panel b shows the share of participants per firm in the manufacturing and service sector weighted by the WB-ES-provided weights. 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. Panel c and d show the counterparts from the EU-CVT provided by the publicly available results (trng_cvt_01s and trng_cvt_12s). Data on GDP per capita comes from the Penn World Table.

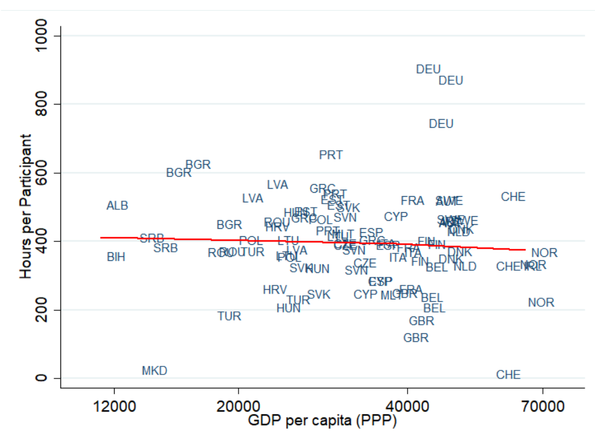
Figure C.4: Time Spent in Education



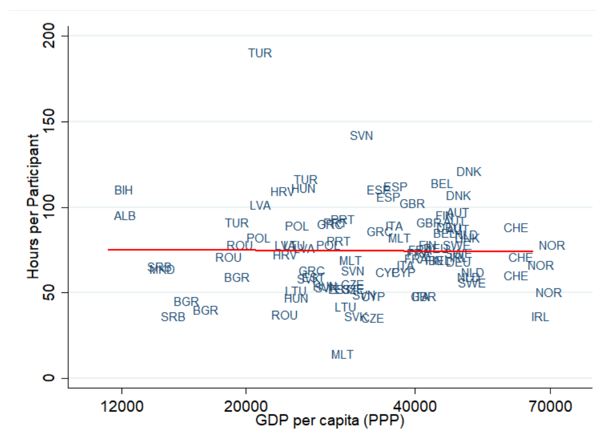
(a) EU-CVT: Hours per 1000 Hrs Worked



(b) EU-CVT: Hours per Participant



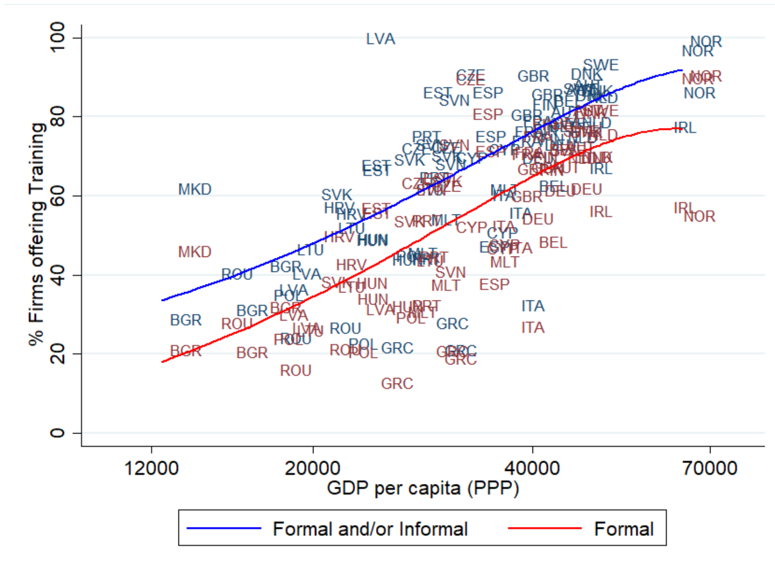
(c) EU-AES: Formal Education



(d) EU-AES: Training

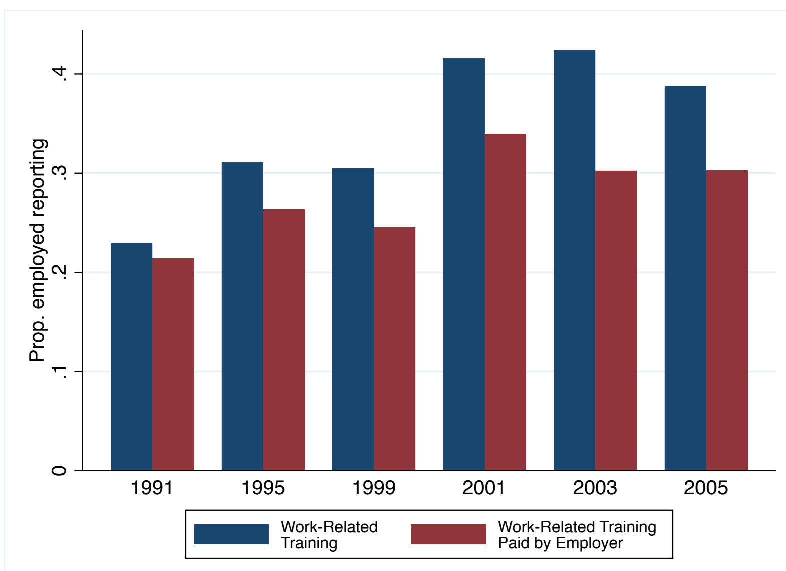
Note: These figures show the time spent in training for different data sources. Panel a represents the hours per 1000hrs worked for participants and Panel b represents the hours spent in training per participant coming from the publicly available EU-CVT data (trng_cvt_22s and trng_cvt_25s). Panel c represents the hours spent in formal education, which we call schooling, and Panel d represents the hours spent in training per participant coming from the EU-AES publicly available data (trng_aes_147). Data on per capita GDP comes from the Penn World Table.

Figure C.5: Share of Firms Offering Formal and Informal Training



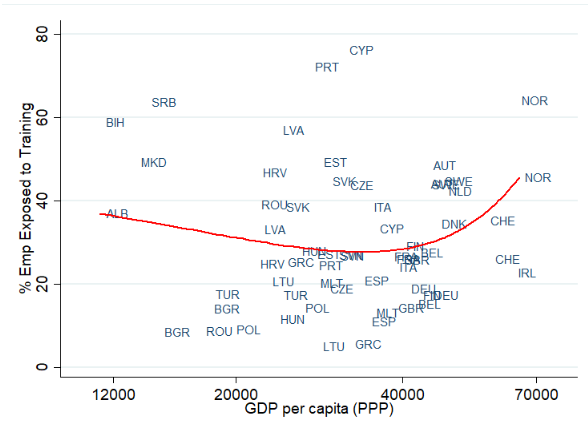
Note: This figure shows the share of firms offering formal training in red and the share of firms offering formal and informal training in blue as a function of per-capita GDP. Data comes from the EU- CVT survey and formal training is defined as CVT and informal training as "other forms of CVT" as defined in the data description section. The data comes from the EU-CVT publicly available data (trng_cvt_01s). Data on GDP per capita comes from the Penn World Table.

Figure C.6: Share of Workers Reporting Training by Year in the US

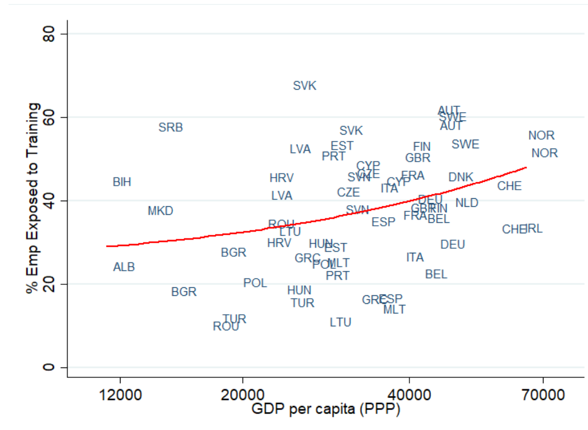


Note: This figure shows U.S. workers' participation rate in work-related training (training, workshops, seminars, courses, or classes for work related reasons in the past 12 months) and work-related training sponsored by employer (training paid at least partially by employer). We use all years with data 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).

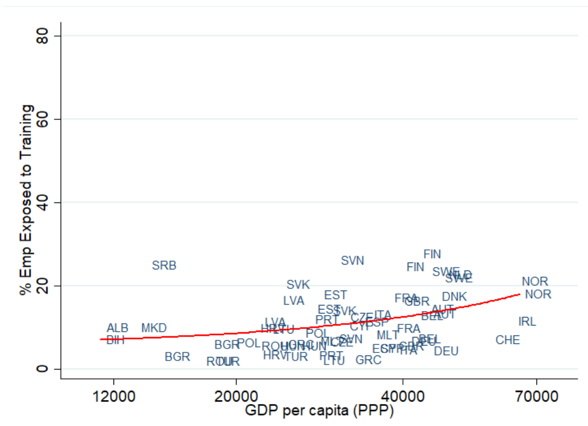
Figure C.7: Informal Learning AES



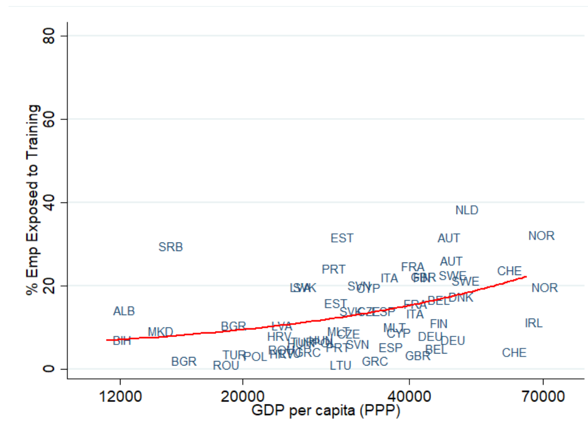
(a) Through Peers



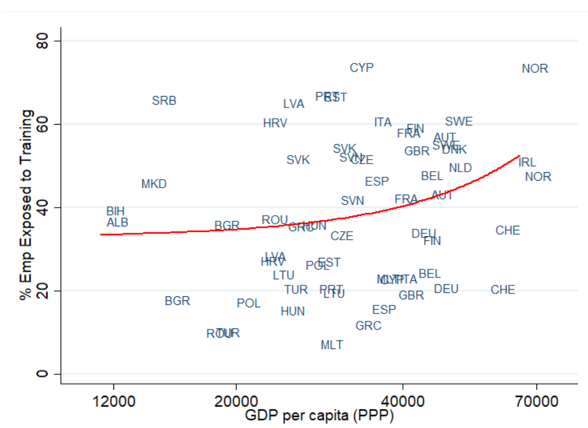
(b) Using Printed Material



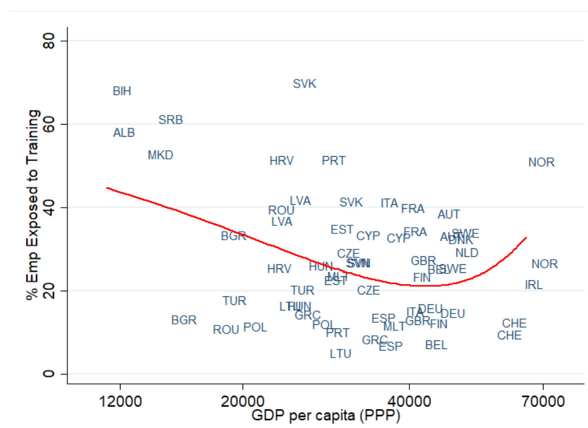
(c) Attending Learning Centers



(d) Tours on Relevant Sites



(e) Using Computer



(f) Using Media

Note: These figures show participation rate in informal learning that includes learning through peers (Panel a), using printed material (Panel b), attending to learning centers (Panel c), tours on learning sites (Panel d), using computers (Panel e) and using media (Panel f). Data are publicly available from the EU-AES (trng_aes_202). Data on per-capita GDP 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

This data comes from Eurostat, past series, LFS ad Hoc Module 2003 (trng_nfe6 for reason, trng_nfe7 for working hours and trng_nfe15 for hours). We show the outcomes from the most populated European countries ranked by population size. In the Online Appendix we provide the complete Table with the outcomes for all 31 countries in the sample.

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

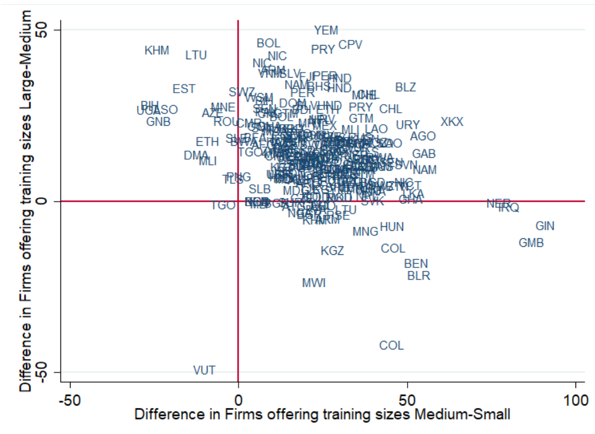
	Average By firm Size in 2010				Average By firm Size in 2015			
	All	10-49	50-249	250+	All	10-49	50-249	250+
General IT	27.3	23.7	34.5	54.7	12.8	12	15.2	15.8
Professional IT	16.9	14.5	21	37.5	10.2	9.8	11.6	11.2
Management	32	26.2	43.7	74.3	23.4	19.9	30.9	49.2
Team working	32.5	29	38.3	61.6	19.6	19.1	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

Note: This table shows the main skills targeted by CVT courses by type of skill and size class. This represents the share of enterprises providing CVT courses (publicly available data trng_cvt_29s). The table shows the share for all firms and for the firm size categories 10 - 49, 50 - 249, and 250+. A particular course may cover more than one category. In the EU-CVT survey, variables C5a to C5i stand for the different skills targeted by CVT courses. A skill is a main targeted skill if it is important in terms of training hours.

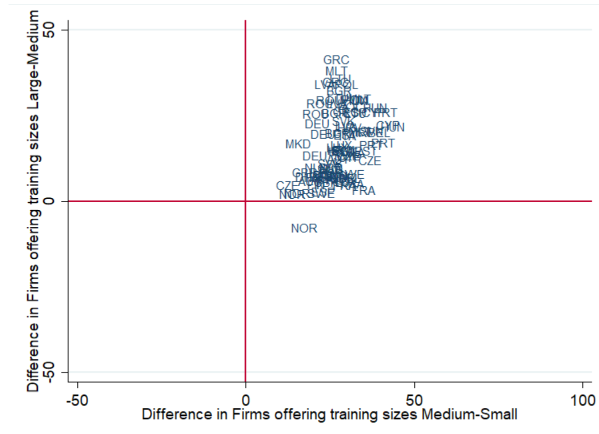
D Model Validation: Firm Features and Training

In this section we focus on the relation between firms' characteristics and training investments. Empirical evidence for some developed economies shows a positive correlation between size and training investments. Nonetheless, there is no evidence on the relationship for developing economies. We show that training increases with firms' size in virtually every country in our sample in Figure D.1. Moreover, we rely on Chinese firm administrative data to analyze how financial characteristics affect training at the firm level. We regress training expenditures on the labor share, firm's size and different TFP measures in Table D.1. We find that TFP positively correlates with training expenditures and that lower labor shares are associated with larger training investments validating the outcomes of our model. In the Online Appendix we provide further robustness checks on these relationship.

Figure D.1: Difference in Share of Firms Offering Training



(a) Enterprise Survey - World Bank



(b) Enterprise Survey - EU

Note: this Figure shows the difference between the share of firms offering training between the following size categories: small and medium, and medium and large. Using each survey we construct a measure of the share of firms offering formal training by size category in each country. Due to the size categories used for the stratification method in the sample, we keep the original sizes 5-20, 21-100 and 100+. For the EU CVT we use 10-49 50-249 and 250+.

Table D.1: China- Dependent Variable: Training Expenditures

	(1)	(2)	(3)	(4)	(5)	(6)
labor share	-2.962***	-.4937***	-2.291***	-0.402**	-3.543***	-0.568***
	(0.093)	(0.1683)	(0.094)	(0.168)	(0.092)	(0.168)
log(firm size)	1.137***	0.497***	0.977***	0.420***	1.137***	0.497***
	(.006)	(.0240)	(0.007)	(0.023)	(.007)	(0.023)
TFP (OP)	0.1695***	0.151***				
	(0.006)	(.012)				
TFP (LP)			0.288***	0.176***		
			(0.007)	(0.013)		
TFP (HK)					0.075***	0.140***
					(0.007)	(0.012)
Age, Year and Ind FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	No	Yes	No	Yes	No	Yes
Obs	685,673	685,673	685,673	685,673	681,094	681,094
R-squared	0.1061	0.7725	0.1073	0.7725	0.1053	0.7708

Note: Robust standard errors are in parenthesis. Significance levels: * 10%, ** 5%, *** 1%. We rely on measures of TFP following Olley and Pakes (1996), Levinsohn and Petrin (2003), and Hsieh and Klenow (2009) which are available in the dataset. Firms' size is measure through employment and labor share is measure as labor cost over total sales.

E Quantitative Model: Conditions for Simulations

Workers' Expected Utility. With linear utility and $\rho = r$, we assume workers do not save and spend all their income each period. Thus, for a worker of age a, their utility comes from (future) income flows produced by workers' 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 current human capital, and $J_{h,a}(z)$ as the expected value of income flows from human capital accumulation. With little abuse of notation, we use $J_{c,a}(u)$ and $J_{h,a}(u)$ for unemployed workers.

First, note that in the last period of workers' lifetime ($a = T$), workers have no incentive to accumulate human capital. Thus, we can obtain

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

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

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

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

$$J_{h,a}(z) = -\mu_W\left(c_s\bar{w} + \delta_s A_M z\right) + \frac{\zeta s_a(z)^{\gamma_s}}{1+r}\delta\left(\theta q(\theta)\int J_{c,a+1}(z)dF(w(z)) + (1-\theta q(\theta))J_{c,a+1}(u)\right)$$

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

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

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

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

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

$p_a(z, z')$ 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:[24]

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

$\bar{h}_a(z)$ is the average human capital of age a workers in firm z, which will be derived soon.

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

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

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 $u_{a+1} = \delta N_M + (1-\theta q(\theta))(1-\delta)u_a \ \forall\ 1 \le a \le T-1$.

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 proceeds as

$$m_{a+1}(z) = \underbrace{(1-\delta)\left[1 - \eta\theta q(\theta) \int p_{a+1}(z,z') dF(w(z'))\right] m_a(z)}_{\text{stayers}}$$

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

Training. Firms' optimal training is determined by:

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

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

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

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

[24]For computational tractability, we do not use different values of leaving probability for individual workers of age a in firm z. Because costs of leaving increase with human capital and income flows from current human capital are larger than benefits from future human capital accumulation in most cases of our simulation (except for early ages when workers have little human capital), this simplification is also reasonable.

our analytical model, the optimal training now depends on the present value of all future returns, adjusted for the depreciation rate of training and workers' separation rates (for firms). Notably, the 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. Specifically, define $\overline{h}_a(z)$ as the average human capital of age-a workers in firms with productivity z. The human capital of the youngest cohort is $\overline{h}_1(z) = 1$. We could obtain the dynamics of human capital as:

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

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 of unemployed people with $\overline{h}^u_1 = 1$.

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:

$$
\begin{aligned}
c_v v(z)^{\gamma_v} = & \underbrace{\sum_{a=1}^{T}\frac{q(\theta)(A_M z-w(z))}{\sum_a u_a+\eta(N_M-u_a)}\left[\eta(1-\delta)\int p_a(y,z)\overline{h}^s_a(y)(1-d)m_{a-1}(y)dG(y)+u_a\overline{h}^u_a\right]\frac{\Psi(\phi,0,a)}{(1-d)^{-1}}}_{\text{benefits from new hires' human capital}} \\
& + \sum_{a=1}^{T-1}\frac{q(\theta)\left[\eta(1-\delta)\int p(y,z)m_{a-1}(y)dG(y)+u_a\right]}{\sum_a u_a+\eta(N_M-u_a)} \\
& \times \underbrace{\sum_{t=0}^{T-a}D(z,t,a)\left[\zeta s_{a+t}(z)^{\gamma_s}(A_M z-w(z))\Psi(z,1,a+t)-\mu_F c_s(z)s_a(z)\right]}_{\text{benefits from training new hires}}.
\end{aligned}
$$

We define $D(z,t,a) = \prod_{k=1}^{t}\left(\frac{1-\eta(1-\delta)\theta q(\theta)\int p_{a+k}(z,z')dF(w(z'))-\delta}{1+r}\right)$ 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 $c_s(z) = \delta_s A_M z + c_s \bar{w}$.

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.

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

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

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

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

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 Alternative Measures for Calibration

In this study we use three moments that are key to performing our counterfactuals. In the baseline calibration, we add the job-to-job and job-to-unemployment probabilities and training intensity (the share of time workers spend in training on average). In this subsection, we explain in detail how we measure these moments and provide alternative calibrations using different measures and their implications for our results.

F.1 Calibrating Labor Market Dynamics

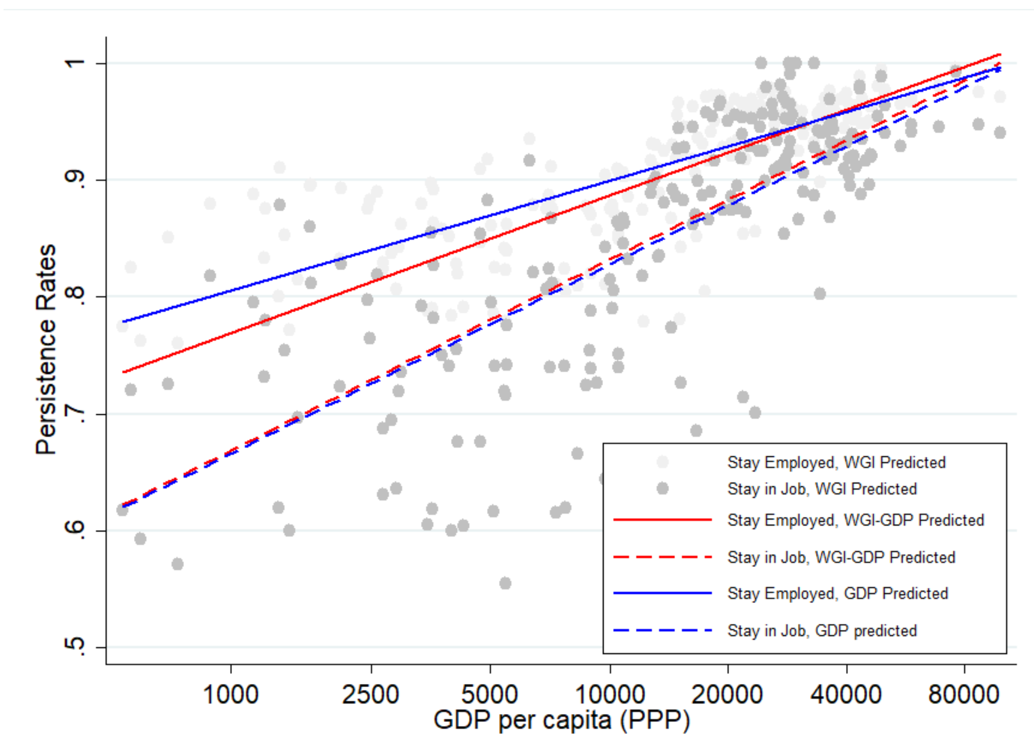
For job turnover dynamics, we use two measures in our calibration: the share of employed people remaining in the same firm and the share of employed people remaining employed after a quarter. We rely on data from Donovan, Lu and Schoellman (2020), which provide these two probabilities for many countries. Their study is the first, and only, in providing the relationship between these probabilities and development. Nevertheless, the countries in their sample do not match the countries and years in our sample in most cases and, therefore, we must build predicted measures using their data. Moreover, as the purpose of this study is to provide comparisons across countries, we must have consistent measures for all countries.

Our main measure uses the variation on institutional quality across countries, which shapes labor market dynamics, particularly job turnover. If contracts are better and easily enforced, job turnover will be lower. Thus, we first regress Donovan, Lu and Schoellman (2020) probabilities in all institutional measures from the World Bank Worldwide Governance

Indicators and predict the probabilities for all countries in our sample using these variables (imposing an upper bound of 1 given that we are predicting probabilities). Figure F.1 shows these two predicted probabilities in grey. This measure has two issues that are the high noise in the predicted value and that these two probabilities are really close together for some developed countries. This second issue is more relevant, as it implies there are almost no job-to-job transitions, which is counterfactual. Thus, we construct a smooth probability measure by predicting our previously built measure with GDP per capita. We plot these new predicted values with respect to per-capita GDP in red in Figure F.1.

We construct a second measure to use as robustness for our main specification. We directly predict the probabilities from Donovan, Lu and Schoellman (2020) with per-capita GDP and plot the outcome in blue in Figure F.1. We show that the probability of a worker staying in the same firm is the virtually the same for both measures. For the probability of a worker staying employed after a quarter, our main measure is a little higher than the one just predicted with GDP for poorer economies, which imply that there is even higher job-turnover in developing economies. Finally, it is worth noting that our results do not change substantially by adding either one of these two measures.

Figure F.1: Job Transition Probabilities



Note: This figure shows the three measures for the share of employed people remaining in the same firm and the share of employed people remaining employed after a quarter. The first measures in the grey scatter plot represents the result of first regressing Donovan, Lu and Schoellman (2020) probabilities in all institutional measures from the World Bank Worldwide Governance Indicators and then predicting the probabilities for all countries in our sample. The predicted red lines are the linear prediction of the first measure with respect to per capita GDP and the blue lines are the linear prediction of Donovan, Lu and Schoellman (2020) probabilities with respect to per capita GDP directly. Data on per-capita GDP comes from the Penn World Table.

F.2 Calibrating Training Intensity

As mentioned in the calibration section of the paper, we do not have a direct measure of total firm-training for all countries. For instance, for most of our economies we have measures of formal training but not of informal sources of training. Thus, we build different measures to show how each one impacts our training estimation. We show the quadratic prediction with respect to log per-capita GDP for these measures in Figure F.2.

1. **Main specification.** We use two assumptions coming from the SEPT survey; the informally trained average worker spends two hours in training for each hour spent by a worker who is formally trained and that there are 50% more workers participating in informal training than in formal training. Moreover, we use the hours spent in formal training per participant predicted with per capita GDP coming from the EU-CVT data presented in the empirical portion of the paper. Thus, with our measure of formal training for each country and the measure of hours spent in formal training per participant, we construct a measure of training intensity for every economy.

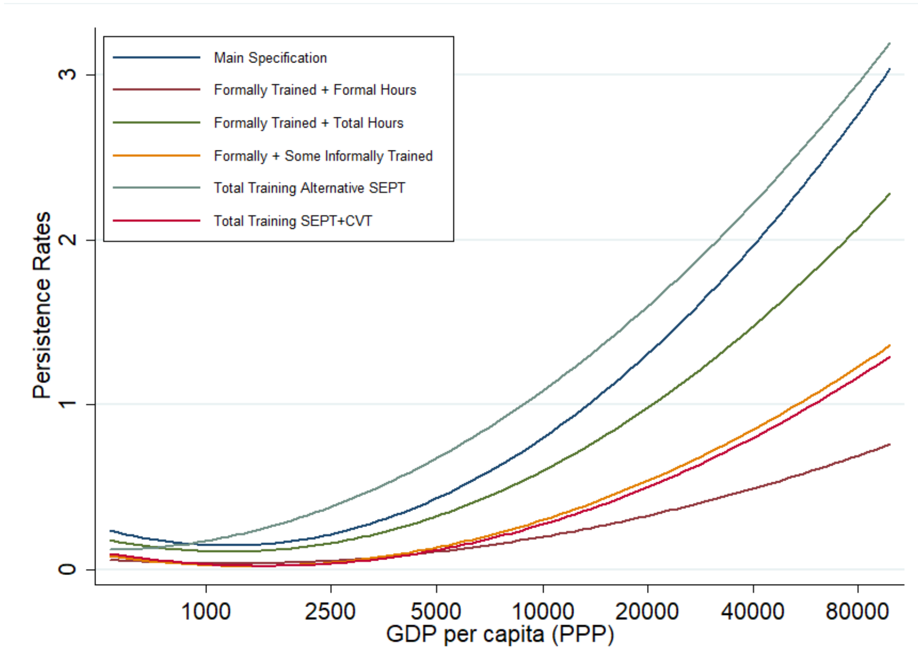
2. **Formally Trained + Formal Hours.** We use the share of formally trained workers from the WB-ES and EU-CVT and the amount of hours coming from the EU-CVT data shown in the empirical portion of the paper. This measure does not consider any informal training for which we need extra assumptions to estimate. This measure is our lower bound.

3. **Formally Trained + Total Hours.** We use the share of formally trained workers from the WB-ES and EU-CVT, assuming these workers are the only ones trained (which is a lower bound) but that they are trained for the amount of hours in formal and informal training. We follow the assumption of our main specification that there are two hours of informal training per one hour of formal training. This is still lower than the most realistic measure as there are more workers who are informally trained than just the formally trained share of workers. We rely on the amount of hours calculated using the EU-CVT as the main specification.

4. **Formally + Some Informally Trained.** We have data on the share of workers who participated in some specific activities of informal training from the private EU-CVT data source. Nonetheless, there may be other activities that are not included, such as informal interactions with supervisors and initial vocational training. Thus, this measure is a lower bound for the share of informally trained workers. We use the same measures of hours spent in training as before for each type of worker (formally and informally trained).

5. **Total Training Alternative SEPT.** This measure is similar to the main measure used but instead of assuming that the informally trained workers are 50% more of the formally trained ones we use the elasticity of the "some informally trained" share of workers elasticity with respect to per-capita GDP from the EU-CVT (but normalized to match that in the U.S. 90% of the workers were informally trained). Then we also use the share of workers formally trained and hours spent in training as the main specification.

Figure F.2: Training Measures Alternatives



Note: This figure shows the six measures of training intensity constructed from the data. The main specification used is the one depicted with a blue line. Then we add the fitted values for the specification with only formal training as the lower bound and the specification using assumptions from the CVT and SEPT surveys as the upper bound. In red orange and green, we add specifications including formal training and different assumptions to represent informal training. All these measures are explained in appendix subsection F.2. Data on per-capita GDP comes from the Penn World Table

From all these measures in Figure F.2 the lower bound is the measure "Formally Trained + Formal Hours," which uses direct data from our sources with no extra assumption. Although this measure does not consider informal training sources, it is useful to consider what the lower bound is for our results. Moreover, from all these measures, the upper bound is the alternative measure to the main specification that uses the elasticity of informally trained workers from the EU-CVT and hours in training as in the main specification. Nevertheless, none of these measures include initial vocational training.

G Identification of Model Parameters

We now illustrate how the moments help identify parameters. We calculate the elasticity for moments to each parameter and provide the results in Table G.1. First, we describe the parameters closely related to labor market outcomes. For the constant in the vacancy cost function c_v, the most sensitive moment is the ratio of vacancies to unemployment. Similarly, c_m affects the economy's matching efficiency, and the moments that identify this parameter are both the vacancy-to-unemployment ratio and the unemployment rate. As expected, 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 traditional sector share in production, γ, has the main impact in the self-employment share. Lastly, a larger shape parameter of the Pareto productivity distribution,

κ implies fewer productive firms, which reduce the wage sector's relative return to the self employment sector and the average wage growth after job-to-job transitions.

Table G.1: Elasticities of Targeted Moments to Parameters

	Labor Market Dynamics					Training Dynamics				Frictions		
	c_v	c_m	$\underline{p}$	γ	κ	γ_s	c_s	ζ	μ_F	γ_p	c_p	δ
Unemployment Rate	0.2	**-1.4**	-0.1	-0.3	-0.2	0.1	0.0	-0.1	0.0	0.0	-0.1	**0.8**
Vacancies/Unemployed	**-0.7**	**1.0**	0.1	0.1	0.2	**-0.6**	0.0	**0.5**	-0.2	0.1	0.3	**-0.8**
Self-Employment Share	0.1	**-0.5**	0.4	**4.1**	**2.1**	1.1	0.1	**-1.5**	0.4	0.2	**0.5**	**0.6**
Pareto Parameter	0.3	0.0	0.4	0.0	**0.6**	0.0	0.0	0.0	0.0	0.2	**0.6**	0.3
% workers leaving Firm	-0.1	**0.5**	0.2	0.0	-0.1	-0.1	0.0	0.1	0.0	-0.1	-0.2	0.4
% workers J-to-U	0.0	-0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**0.8**
Av wage growth J-to-J	0.2	-0.2	**-1.2**	0.0	**-1.5**	0.0	0.0	0.1	0.0	**-0.5**	**-0.9**	0.2
% J-to-J high-to-low	-0.1	0.2	**0.5**	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.3	-0.2
Training Intensity	0.0	**-0.6**	-0.2	-0.3	0.0	**0.6**	-0.2	**0.6**	**-1.2**	0.1	0.2	-0.4
Trng ratio large-small	-0.1	-0.1	0.0	0.0	-0.2	0.1	0.0	0.0	0.0	0.0	-0.1	0.0
Direct/wage cost (trng)	0.0	-0.1	0.1	0.0	0.4	-0.1	**1.0**	0.0	0.0	0.1	0.1	0.1
% wage increase 40 yrs	0.0	-0.1	-0.2	0.0	-0.2	**-1.5**	-0.1	**2.0**	**-0.7**	0.0	0.0	**-0.5**

The table reports the elasticity for moments to each parameter where 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 keeping the rest parameters fixed.

Second, we focus on the parameters directly related to training. The parameter c_s pins down the importance of direct training costs and is identified by the ratio of direct costs to wage costs of training. The parameter ζ determines how training translates to efficiency units 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.[25] Finally, training intensity decreases with μ_F — the share of the training 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. We will discuss optimal policies to reduce training inefficiency in Section 8.

We now focus on the main parameters that mediate our channels. The breaking contract cost friction is composed of two parameters. The convexity in the cost, γ_p, has the biggest 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. Moreover, c_p has the greatest 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 the unemployment rate and the job-to-unemployment rate, while reducing market tightness (due to more unemployed people). Moreover, it has a relatively strong negative effect on training intensity, in line with our analytical model as well.

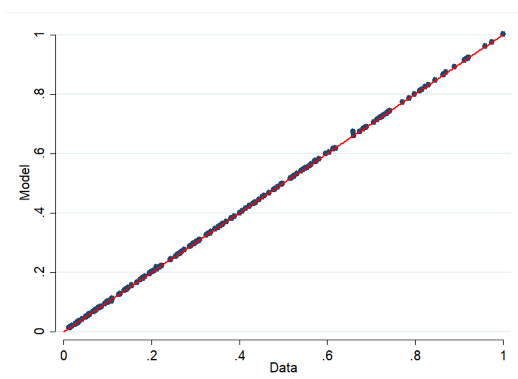
[25] Moreover, γ_s that defines the convexity on the training function also has the biggest impact on the wage growth and self employment share through the impact on training intensity. The signs are more complicated to analyze due to the training function choice because an increase of γ_s increases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

H Alternative Cross-Country Calibration

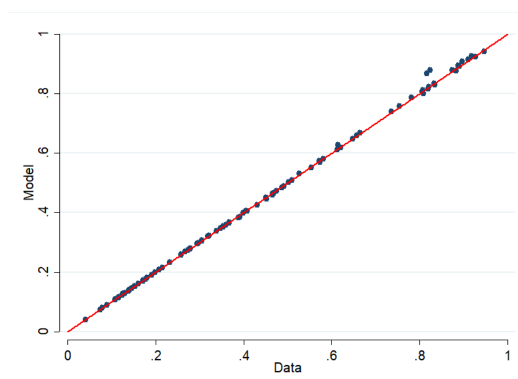
In this section, we calibrate the model to all 100 countries for which we have data on training. The difference with the main specification is that we use the observed variation in self-employment and also match training intensity directly. Therefore, in this calibration we target 5 different moments: (1) Real GDP per capita, (2) traditional sector employment share, (3) exogenous separation, (4) difference in endogenous separation from job-to-job transitions and (5) training intensity. We keep the calibrated parameters from the U.S. baseline calibration except for δ, c_p, A_m, A_T and differently from the baseline calibration we also let ζ vary across countries.

We first show how the model fits the targeted moments in Figure H.1. On the x-axis we show the moments in the data, on the y-axis we show the moments in the model and we plot the 45-degree line for the targeted moments except training. Overall, our model matches the targeted moments well. In Figure H.2a we plot training intensity from the data and the model as a function of GDP per capita, and in Figure H.2b we plot the training intensity in the model (y-axis) and data (x-axis). When we let ζ change across countries, the model exactly matches training intensity for every country.

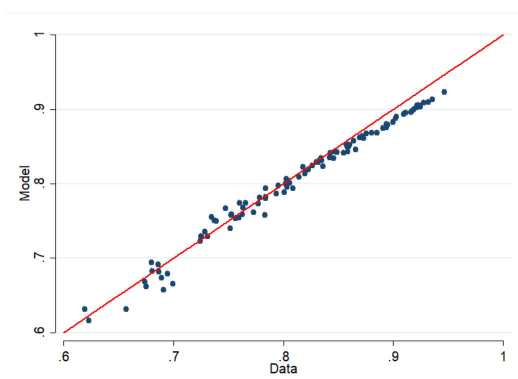
Figure H.1: Cross Country Targeted Moments



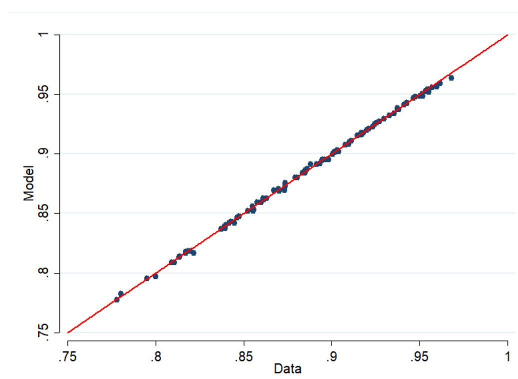
(a) GDP per capita



(b) Self employment share



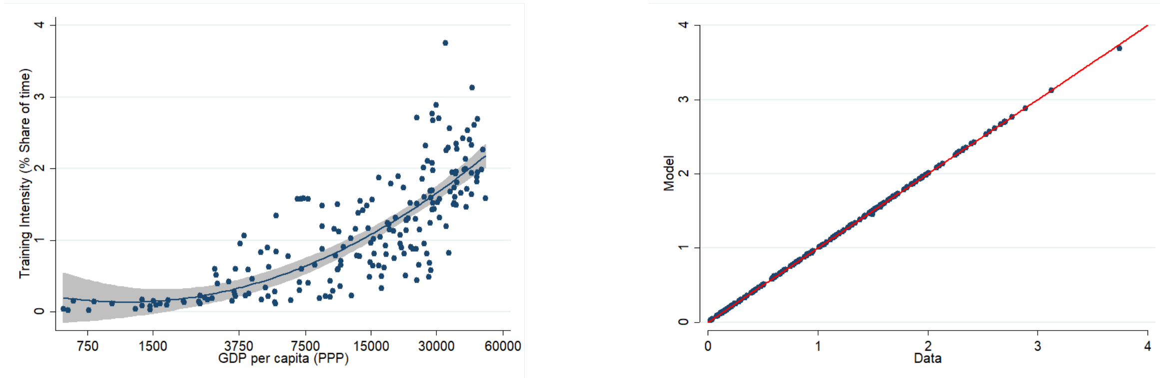
(c) % workers in same firm



(d) % workers still employed

Note: This figure shows the targeted moments in the model (vertical axis) and in the data (horizontal axis). Panel A shows GDP per capita. Panel B shows self employment. Panel C shows the share of workers who remain in the same firm after one quarter. Panel D shows the share of workers who are employed in any firm for two consecutive quarters.

Figure H.2: Training in Data and Model



(a) Training Intensity **(b)** Training Data vs Model

Note: This graph shows the quadratic fit of the cross-country training intensity (measured in the share of time that an average worker spends in training) as a function of Los(GDP per capita). The green line represents the quadratic fit for the cross-country measure in the model and the blue line represents its counterpart in the data. The grey shadow represents the 95% confidence intervals.

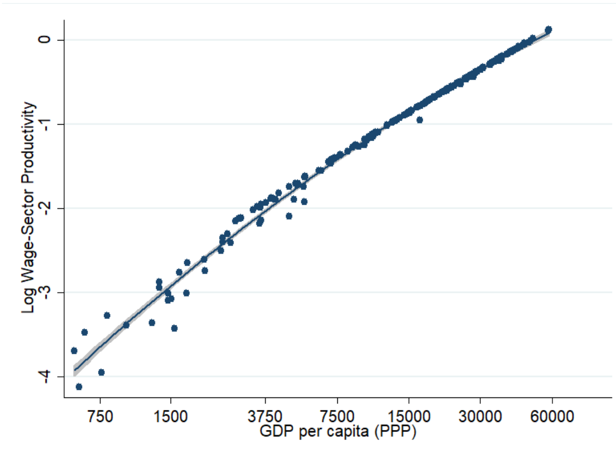
We provide some more evidence to show how the model matches the data. In Table H.1, we show the slopes of training intensity with respect to different moments in the data and model. In all cases the model replicates the data well. Finally, we show the calibrated parameters given by the cross-country calibration in Figure H.3. We get the same elasticities and patterns from the main specification. Interestingly, we add the dynamics on the training productivity and show that this parameter is mostly flat with respect to GDP per capita. This result reinforces our conclusion that most training differences come from factors captured in our parameters.

Table H.1: Training Accounting: Cross-Country Correlations

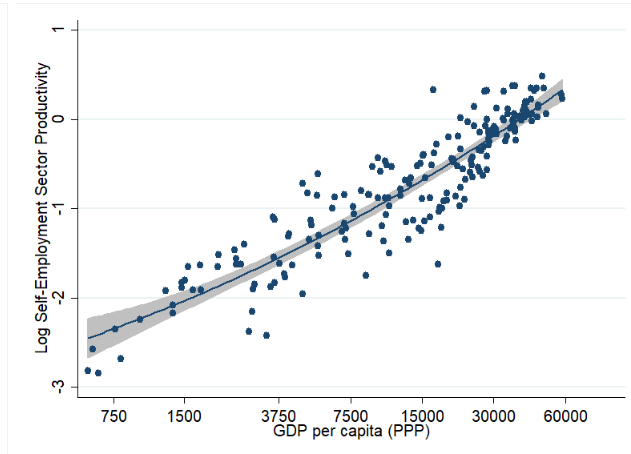
Slopes	Data	Model With ζ	Model Benchmark
Training in modern sector wrt to GDP per capita	0.33	0.33	0.26
Training in Aggregate wrt GDP per capita	0.84	0.84	0.77
Training in Aggregate wrt P(worker staying employed)	17	18	16
Training in Aggregate wrt P(worker staying in firm)	9	10	10

Note: This table shows elasticities in the model and data for model validation. All elasticities in the data column are calculated with data moments and elasticities in the model column are calculated with their counterparts in the model. The first row regresses log training intensity in the modern sector on per-capita GDP. The second row regresses log training intensity in the aggregate economy on per-capita GDP. The third row regresses log training intensity in the aggregate on the probability of workers staying employed from quarter to quarter. The fourth row regresses log training intensity in the aggregate on the share of workers who stay in the same firm from quarter to quarter. Data on GDP per capita comes from the Penn World Table and data on the share of workers in the third and fourth row comes from the predicted values in our sample from Donovan, Lu and Schoellman (2020) data.

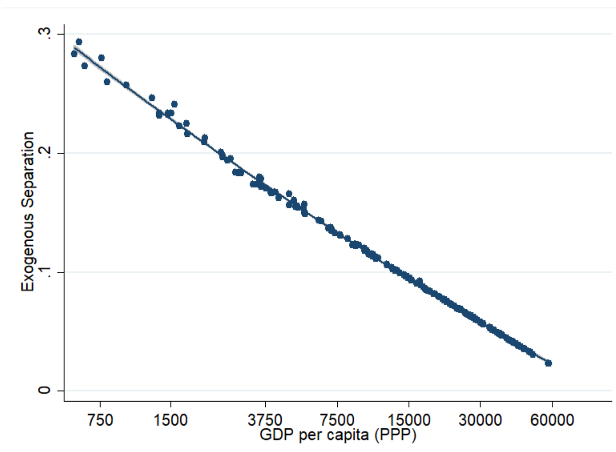
Figure H.3: Cross Country Calibrated Parameters



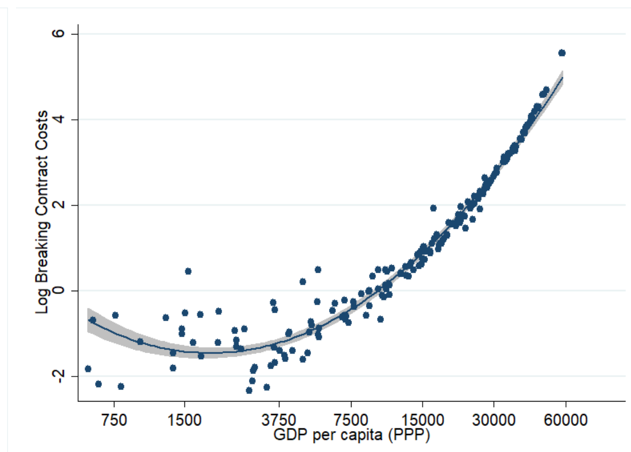
(a) Modern Sector Productivity



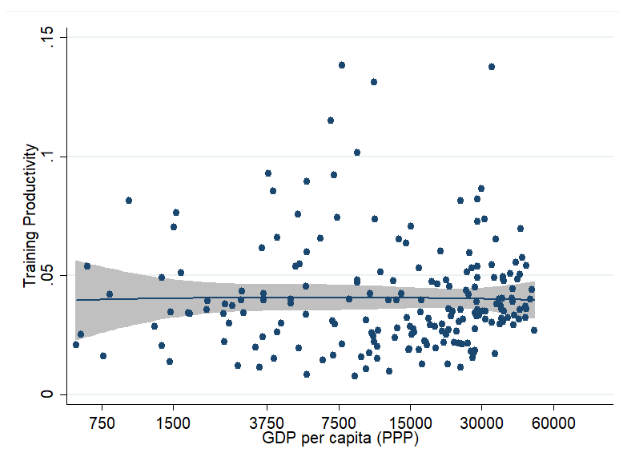
(b) Self-Employment Sector Productivity



(c) Exogenous Separation



(d) Breaking Contract Costs

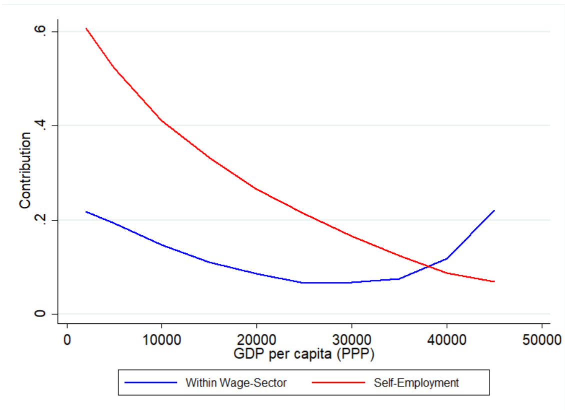


(e) Productivity of training

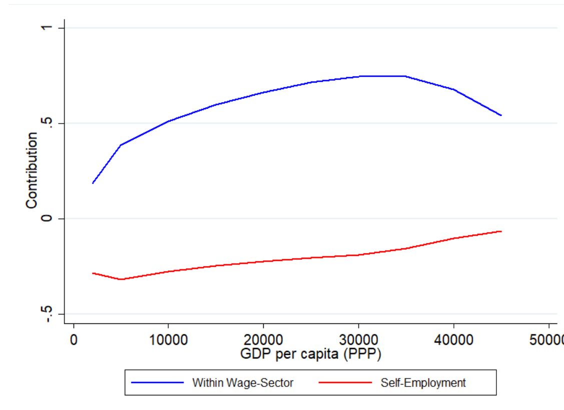
Note: This figure shows the calibrated parameters for each economy in the model as a function of Log(GDP per capita).Panel A shows the wage sector productivity (A_M in the model). Panel B shows the relative productivity between the self-employment sector and the wage sector (A_T/A_M). Panel C shows the quarterly exogenous separation rate implied by the model (δ). Panel D shows the log of the breaking contract costs ($\log(c_p * A_M)$).

I Training Decomposition

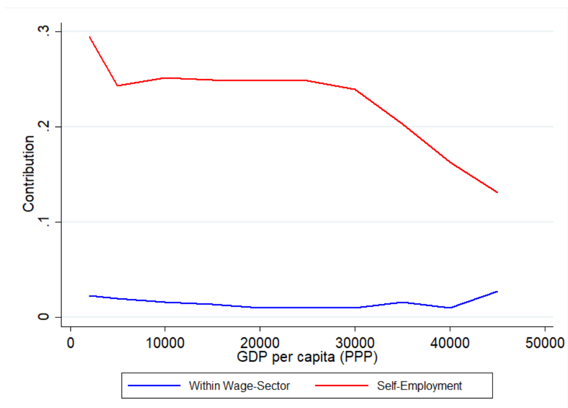
Figure I.1: Cross Country Calibrated Parameters



(a) Exogenous Separation

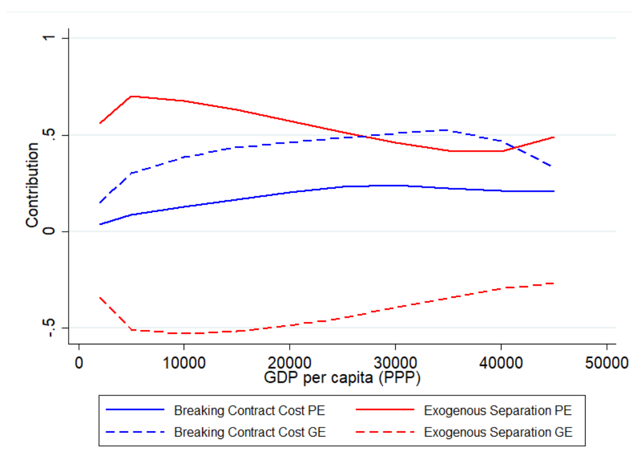


(b) Breaking Contract Costs



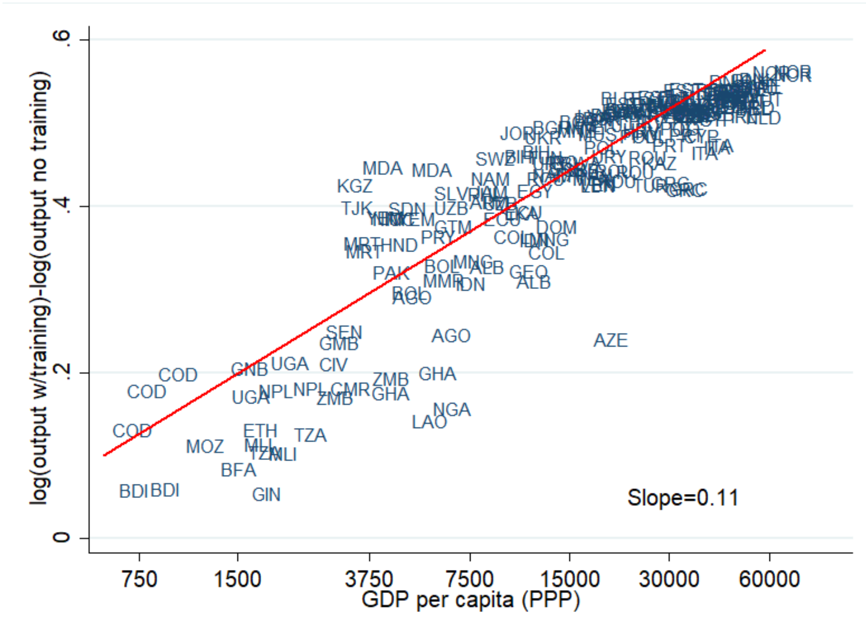
(c) Productivity

Figure I.2: Partial and General Equilibrium



J Income Accounting Robustness

Figure J.1: Income Increase due to Training



Note: This figure shows the percentage increase in output from training calculated as the log change in output from the model shutting down training (increasing c_s to an extreme large value) to the full model as a function of GDP per capita. Each observation comes from using the calibrated version of the model for each country. Data from GDP per capita comes from the World Bank Indicators. The slope of 0.16 represents the share of the increase in GDP per capita explained by training in the model.

Table J.1: Cross-country Income Differences and Training Contribution

Representative Economy In Full model	Per capita GDP Ratio With Training	No Training	Inc Accounting Contribution
\$50,000 to \$45,000	1.1	1.0	6%
\$50,000 to \$25,000	2.0	1.5	25%
\$50,000 to \$5,000	8.7	5.7	34%

Note: The table lists measures to explain income differences across countries. The first column shows the ratio of per-capita GDP between different percentiles of the full model's income distribution. The second column shows the ratio between different economies shutting down the training channel. The third column represents the share of the difference explained by training, which is calculated using the ratio of column 3 over 2.

We continue by focusing on other measures. We calculate the ratio of per-capita GDP between different economies in the model with and without training and show the results in Table J.1. We use the calibrated parameters from the full model for different economies provided in Column 1. We show the ratio of per-capita GDP in the model with training in Column 2 and the ratio of per capita GDP in the model shutting down training in Column 3. We observe that the income differences expand for all ratios when training is added to the model. In Column 4, we take the log difference between the results in Column 2 and 3, which represents how the income differences expand due to training in our model. For more developed economies, training accounts for a smaller share of the income differences as these economies are similar in terms of self-employment, turnover rates, and institutional quality. As the difference in GDP per capita increases, the share of income explained by training increases.

K Training Investment Inefficiency and Subsidies

It is clear that training investments are usually not efficient due to worker turnover and incomplete contracts. In this section, we do an extensive description of government subsidy policies and show that training subsidies are indeed very common. Figure K.1 presents the countries for which we found data on training incentives. We review government policies from countries in all continents (we provide examples for 36 countries from all income levels) and we also show examples for 21 U.S. States for which there is data available on government policies to incentivize employer-provided training. We present the survey for the cross-country policy examples in Table K.1 and the examples for the U.S.' States in Table K.2.

Figure K.1: Examples for Government Training Incentives

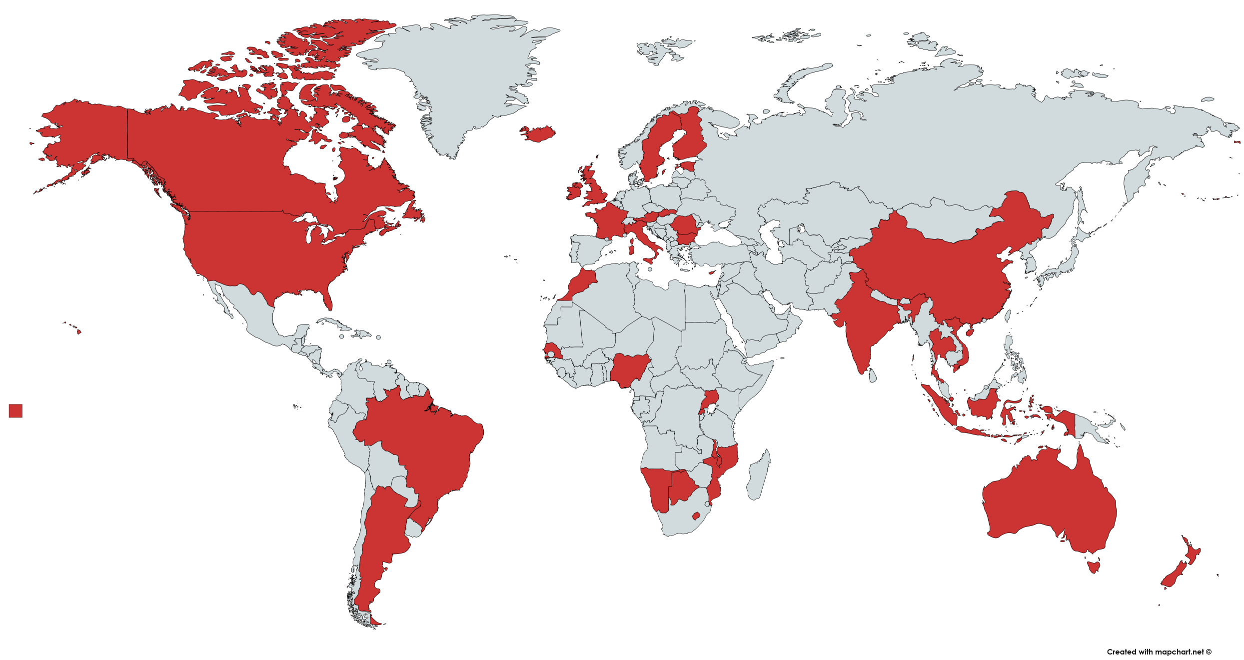


Table K.1: Training Subsidies Across Countries

Country	Year	Subsidy or Incentive to employer
Argentina	2000 - present	30% of training costs are tax deductible
Australia	2019 - present	50% or $2,200 of training costs granted (or to employee)
Austria	2002 - 2015	120% of training costs tax deductible
Botswana	1985 - present	200% training costs tax deductible
Brazil	2012 - present	No set maximum of tax deductible training expenses
Bulgaria	2007 - present	80-90% of training costs granted
Canada	2014 - present	50-83% of training costs granted
China	2015 - present	8% of total payroll may be deducted from taxable income
Cyprus	1979 - present	60-80% of training costs granted
Czech Republic	1992 - present	100% of training costs tax deductible
Estonia	2012 - present	100% of training costs tax deductible
France	2005 - present	Wages of trainees are paid
Finland	2014 - present	50% of employee's average wage is tax deductible
Iceland	1998 - present	100% of training courses reimbursed
India	2016 - present	Funding of $20-100 for IVT (apprenticeship)
Indonesia	2019 - present	200% of learning costs tax deductible for corporate taxpayers
Ireland	1999 - present	Maximum not specified, but depends on project and industry
Italy	2017 - present	50% of training costs of tax deductible
Lesotho	1980 - present	50% of wage bill reimbursed
Malawi	1999 - present	20-50% of training costs reimbursed
Mauritius	2003 - present	75% of training costs reimbursed
Morocco	2014 - 2020	20% of training costs reimbursed (large projects)
Mozambique	2002 - present	5-10% of taxable income may be deducted
Namibia	1995 - 2020	75% of training costs reimbursed
New Zealand	1983 - present	Funding of $5,415 (or to trainee)
Nigeria	1971 - present	Reimbursement of 50% of payroll tax paid
Romania	2000 - present	No set maximum of tax deductible training expenses
Rwanda	2014 - present	70% of training costs granted
Senegal	2014 - present	80-90% of training costs granted
Singapore	2016 - present	90% of training course fees reimbursed
Slovakia	2003 - present	100% of training tuition tax deductible
Sweden	1996 - present	Training costs are tax deductible (no specified maximum)
Thailand	2002 - present	200% of training costs tax deductible
Uganda	1997 - present	100% of training costs tax deductible
United Kingdom	2017 - present	95% of IVT (apprenticeship) costs paid
Vietnam	2019 - present	100% of training costs subsidized for female-owned enterprises

Table K.2: Training Subsidies within United States

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

L Our Channels in the Data

In Table L.1 we show suggestive evidence on the correlations between training investments and job turnover measures from Donovan, Lu and Schoellman (2020), self-employment, firm size distribution, and institutional quality proxies in the data. It shows the results of regressing the share of employment exposed to training on GDP per capita, the share of employment in small firms (to account for the composition effect not captured by job separation), the probability of staying in the same job, and the first principal component of all 5 institutional measures from the World Bank Worldwide Governance Indicators.[26] The semi-elasticity of GDP per capita with respect to our training measure is 8.69. As we add each one of the explanatory variables, we show how the coefficient on GDP per capita decreases. Once we add the first principal component that includes all the other variables, we explain all the correlation between GDP per capita and training, which suggests that institutional quality, job separation, and self-employment captures most of the pattern described. In the Online Appendix we provide further robustness checks on these relationships, regressing our training measure in all individual institutional variables and our conclusions hold.

Labor Market Institutions. From the model, it is clear that things affecting separation rates, the probability of hiring, or the vacancy costs will affect contracts and training investments. It is intuitive to think that higher unemployment benefits or higher firing costs will generate lower levels of training. These could be potential mechanisms to explain why more developed economies invest more in training, and thus, 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 ES and CVT on GDP per capita and each index separately, year and country fixed effects and show the results in Table L.2. Training increases as the legally mandated notice period to fire workers increases, meaning that as the 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 training on-the-job. Moreover, rows 4, 5, and 6 have different measures on the strength of unemployment benefits in different countries. As unemployment benefits increase, training investment decreases. This shows that when the workers' outside option is better, they are harder to retain workers and training investments decrease. This same pattern is observed when countries have meaningful minimum wages and outside options are higher. Nevertheless, although all these measures

[26] We use data from the PWT and the World Bank Indicators for GDP per capita, self-employment, and capital stock. For institutional quality we rely on the World Bank Worldwide Governance Indicators, which provides indexes on country-specific institutional characteristics. The characteristics provided by the WGI are: "Voice and Accountability," "Political Stability," "Government Effectiveness," "Regulatory Quality," "Rule of Law," and "Control of Corruption." Moreover, we use data on separation rates estimated using the results provided by Donovan, Lu and Schoellman (2020). Due to a mismatch between their sample and our sample, we are not able to relate these two measures directly. Nevertheless, we can conduct a 2 step estimation process. We first regress the probability of staying in the same job on all the institutional variables, which gives us an R-squared of almost 80%. Then, we predict the probability of staying in the same job for all countries using the institutional indicators which gives us predicted separation rates for most of the country-years in our sample.

Table L.1: Share of Workers Exposed to Training

(a) WB-ES and EU-CVT

	(1)	(2)	(3)	(4)	(5)	(6)
GDP per Capita	8.69***	4.69**	5.77***	7.43***	5.98***	1.32
	(0.61)	(1.89)	(1.14)	(0.70)	(0.70)	(1.32)
Log per Capita K		3.58**				
		(1.69)				
self-employed			-0.14***			
			(0.047)			
Prob Same Job				20.2***		
				(7.58)		
1st comp Institutions					1.87***	
					(0.44)	
1st comp All						5.61***
						(1.04)
Constant	-57.8***	-57.0***	-25.3*	-63.5***	-33.4***	11.6
	(6.75)	(6.68)	(12.9)	(7.38)	(7.52)	(12.8)
Year FE	YES	YES	YES	YES	YES	YES
Observations	211	211	211	211	211	211
R^2	0.626	0.635	0.640	0.637	0.651	0.663

(b) WB-ES

	(1)	(2)	(3)	(4)	(5)	(6)
GDP per Capita	8.09***	3.54***	7.63***	7.17***	6.63***	2.87***
	(0.64)	(1.02)	(0.62)	(0.66)	(0.70)	(0.83)
self-employed		-0.22***				
		(0.041)				
% Emp in Small Firms			-25.1***			
			(3.91)			
Prob Same Job				18.2**		
				(7.40)		
1st comp Institutions					1.33***	
					(0.50)	
1st comp All						-5.03***
						(0.75)
Constant	-52.5***	-1.98	-43.3***	-58.4***	-37.5***	-7.02
	(5.33)	(10.6)	(5.34)	(6.41)	(6.44)	(7.33)
Year FE	YES	YES	YES	YES	YES	YES
Observations	194	194	194	194	194	194
R^2	0.517	0.570	0.567	0.532	0.534	0.600

increase the explanatory power over training on-the-job, they do not account for part of the explanatory power of GDP per capita. These results reflect the fact that, although important, these measures as unemployment benefits and labor market characteristics, which are not included in our model (i.e., differences in minimum wages, laws to protect workers,or firing costs) are not the key elements to explain the positive correlation between training and income. This result is consistent with Donovan, Lu and Schoellman (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 that they do not explain the trend relationship between development and labor market flows.

Table L.2: 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
Log(GDP pc) restricted sample	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

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

M Cost Shares

In this section, we provide a cost share analysis. In our model, worker and firm choices of training depend on the marginal revenue and the cost shares as shown. Note that we could make different assumptions on what the cost shares are, and thus, we could have different training patterns.

Proposition 3 (Joint Internal Efficiency) *In a firm with productivity level ϕ and wage $w(\phi)$ if*

$$\mu(\phi)^* = \frac{MR_W(\phi)}{MR_W(\phi) + MR_F(\phi)}$$

then

$$s^*(\phi) = \left(\frac{\zeta e^{\alpha} \gamma_s (MR_W(\phi) + MR_F(\phi))}{(1+r)c_s} \right)^{\frac{1}{1-\gamma_s}}$$

which maximizes the joint surplus from training

$$\max_s \frac{\zeta e^{\alpha} s^{\gamma_s}}{1+r}(MR_W(\phi) + MR_F(\phi)) - c_s s$$

Proposition 3 suggests there is a unique division of training costs that maximizes the joint surplus of firms and workers from training. However, there is still under-investment in training because of the incomplete contract (Acemoglu, 1997) — workers and firms cannot internalize the benefits of training for future employers if separation occurs.

We first show that the marginal return from training increases with firm productivity for both workers and firms, and that this increase is faster for firms. On the one hand, as firms become more productive, the probability of losing the worker is lower, which means firms will enjoy higher revenue from workers for longer. That dynamic, jointly with the increase in training returns with firm productivity, generates the increase in firms' training marginal returns depicted in Figure M.1.b. Moreover, workers have larger expected revenue from training as they will capture the increase in human capital if separation occurs, thus having larger marginal revenue than firms as shown in Figure M.1.a for every firm.

To think about human capital investments, we must also consider the investment costs, which are also increasing due to the opportunity cost. Note that although there is a constant direct cost –C_s– workers lose 70% of production time while being trained. Figure M.2 shows the optimal training levels workers and firms would choose when the firm pays for all training costs (Figure M.2.a in blue), when the worker pays all the costs (Figure M.2.a in orange) and when each one pays for the share they capture from the investment (joint internal efficiency case in Figure M.2.b). Workers have larger probabilities of leaving to better firms when their firms are small and unproductive, and thus, the difference between firms and workers' marginal revenues from training are the largest at the bottom of the firm productivity distribution and workers will want to invest in training more than firms. When firms become more productive they are willing to invest more in training as the increase in revenue is larger than the increase in costs, but the opposite is true for workers. In the joint internal efficiency case, as the ratio MR_F/MR_W increases, firms will start paying a higher share of the training cost which decreases the training investment as firms want lower levels of training than workers.

Figure M.1: Marginal Returns from Training

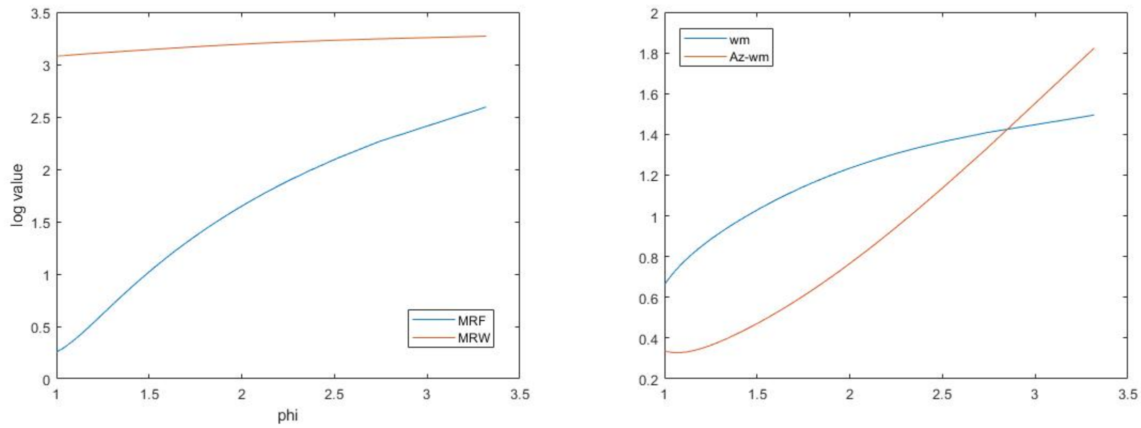


Figure M.2: Workers and Firms Optimal Training Levels

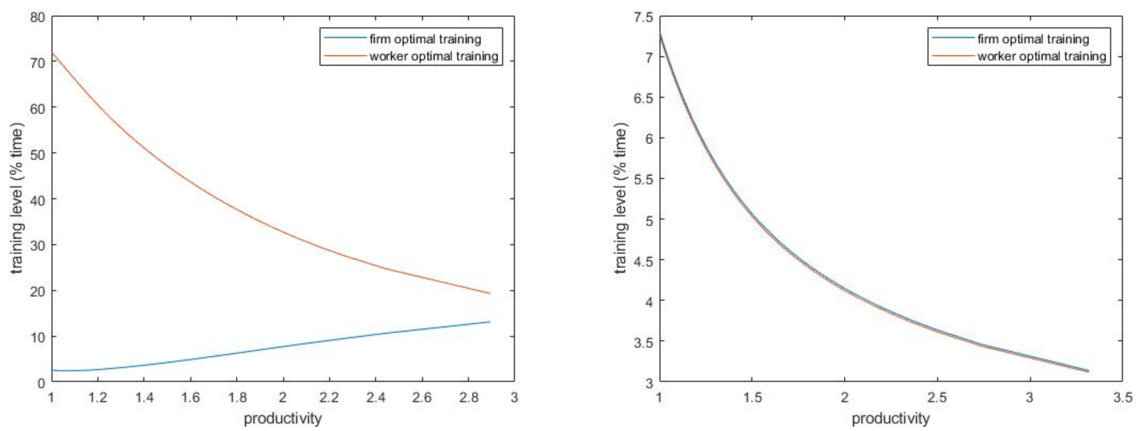


Figure M.3: Training and Cost Shares

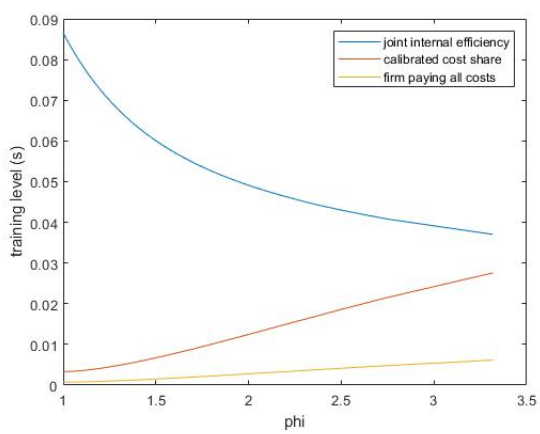


Figure M.3 shows the optimal level of training for three relevant cases (joint internal efficiency, calibrated shares and firm paying all the costs). When firms pay all the cost or when the share of the cost firms pay is constant, training levels increase with productivity. Nonetheless, the case of joint internal efficiency is different because training investments decrease with productivity, which is counterfactual. This evidence suggests that more productive firms do not seem to finance substantially larger shares of training costs than smaller firms, which, if true, would generate a decrease in training with productivity.

N Endogenous Separation Rates: Screening Model

Donovan, Lu and Schoellman (2020) not only show that job separation decreases with development but also that the patterns observed can be characterized by the presence of screening where firms are uncertain of workers' quality. We develop an extension of the model adding screening in the Online Appendix. We assume there is a firm-worker specific productivity and that employers observe workers productivity with some probability. This extension consists of the baseline model with only one simplification, which is that workers can break contracts at no cost (which means that p will be 1 in case the new offer has a wage that is higher than the current wage or 0 otherwise). This simplification is without loss of generality to analyze job destruction and screening. This model would include a mechanisms to add endogenous job destruction which would imply higher job destruction in developing economies when screening technologies are worse, or when firm-worker match productivity dispersion is higher. Nevertheless, the same dynamics and main patterns hold in this extension, although it also adds an extra inefficiency coming from the productivity dispersion.

O Training and the Labor Share

We now provide some extra details on the effects from firm specific labor shares on training. A strand of the literature has focused on studying the relationship between firms' productivity, size and the labor share. In this section, we focus on the relationship between the labor share and training. In the baseline model we show training decrease with the labor share, all else equal. Nonetheless, the labor share depends on productivity, and thus, it is difficult to disentangle the effect from these two. To understand how the labor share behaves and what its implications are, we provide an extension of the model that features different assumptions on wage formation in the Online Appendix. We simplify the baseline model by abstracting from exogenous separation, assuming that the cost of braking the contract is zero and by assuming firms pay all the cost of training. This simplifications will not affect the results and relationship between the bargaining power of workers and firm investments. We find that the labor share directly affects training investments. All else equal, if there is a larger labor share, firms will tend to decrease human capital investments as they can extract a lower share of the surplus, and therefore the marginal return of the investment decreases.