Skill Supply, Firm Size, and Economic Development

Charles Gottlieb University of Geneva
Markus Poschke McGill University
Michael Tueting LISER

November 25, 2026

Why do so few workers in low-income countries work in large firms? Largely, we argue, because skills are scarce. We harmonize microdata for 57 countries and show that where skills are scarce, the skill premium is higher and the large-firm employment share lower. Large firms are also more skill-intensive than small firms, especially in low-income countries. In a model where firms choose between a traditional and a skill-intensive modern technology with greater returns to scale, skill endowments account for 88% of cross-country variation in large-firm employment. Technology adoption raises the output gain from more skills by almost a third.

JEL codes: J24, L11, O11, O14, O47.

Keywords: firm size, skills, technology choice, skill premium, economic development.

Gottlieb: University of Geneva and Aix-Marseille School of Economics, gottlieb.charles@gmail.com; Poschke: Department of Economics, McGill University, markus.poschke@mcgill.ca; Tueting: Luxembourg Institute of Socio-Economic Research (LISER), michael.tueting@liser.lu.
This version: November 2026. First version: April 2024. Niklas Halfar and Selma Azladji provided outstanding research assistance. We thank Francesco Amodio, Pedro Bento, Johannes Boehm, Emanuele Brancati, Paco Buera, Julieta Caunedo, Mons Chan, Giacomo De Giorgi, Douglas Gollin, Jan Grobovsek, Ezra Oberfield, Michael Peters, Tommaso Porzio, Federico Rossi and Mathieu Taschereau-Dumouchel for their comments. We have also benefited from the input of seminar participants at McGill University, the University Carlos 3 Madrid, the University of St. Gallen, CEPREMAP (Paris) and CY Cergy Paris University, as well as participants at the Society for Economic Dynamics Annual Meeting (Copenhagen), the Canadian Macroeconomics Study Group (Hamilton), the North East Universities Development Consortium (Cornell), the Canadian Economics Association Annual Meeting (Toronto), the Firms, Labor Markets, and Development Workshop (Bari), the Conference of the Development Economics Network Switzerland (Zurich), the Verein für Socialpolitik Annual Conference (Cologne), the Brazilian Econometric Society Meeting (São Paulo), and the STEG Annual Conference (Nairobi). We gratefully acknowledge funding from STEG-CEPR LRG Grants 943 and 4685, a SSHRC Insight Grant and Swiss National Science Foundation Grant 10006576. This paper builds on and supersedes an earlier version, which was a background paper for the World Bank's 2024 World Development Report. All errors are our own.

The organization of production differs widely across countries. In high-income countries, four out of five employees work in firms with ten or more employees, while in low-income countries one out of two does. The shift of employment from small to large firms is one of the most salient regularities in how the organization of production changes with economic development. Why does so much of production in low-income countries take place in small firms? We argue that much of the answer lies in the skills of the workforce, and their prices. Operating at scale requires basic literacy and numeracy throughout the workforce. Where few workers have these skills, skills are expensive, and firms adopt technologies that do without them and stay small.

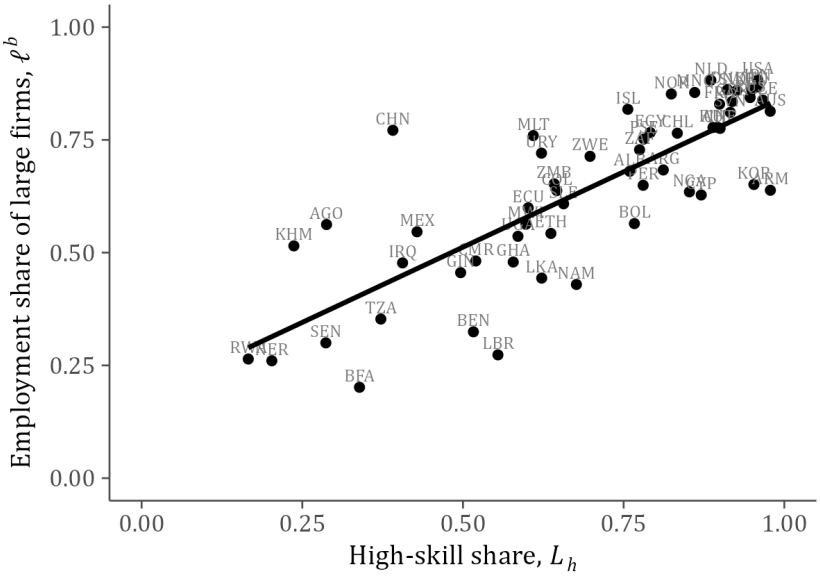


FIGURE I
Skills and employment in large firms.

Notes. The figure plots the share of wage employment in large firms against the share of high-skill wage employees for 57 countries. High-skill is defined as at least ten years of schooling. Large firms have at least ten employees. Each marker is one country, labeled with its ISO3 country code, and the solid line is a straight-line least-squares fit across countries. Sample: wage employees aged 25–55 with non-missing education and firm-size information. The surveys used for each country are listed in Table A.I.

Figure I shows how closely skills and firm size are related. Across 57 countries from Niger to Switzerland, the correlation between the share of wage workers with at least ten years of schooling, whom we call high-skill, and the share of employment in firms with ten or more employees is 0.81.[1] Conditional on log GDP per capita, the correlation is 0.56, so the association does not merely reflect that both education and

[1]We use the ten-year education threshold because it is informative in all countries, including low-income countries where few workers are college-educated, and ten employees because it is the size category that most surveys report. Our main analysis considers wage employees. All results are robust to including the self-employed as workers in small firms and to alternative definitions of skills and firm size – see Appendix C.

employer size rise with income. This finding raises four questions. Do large firms rely more on high-skill workers than small firms? Are skills more expensive where they are scarce? How much of the cross-country variation in the size structure of employment is accounted for by variation in skill endowments? And what does the link between firm size and skill endowments imply for aggregate output?

To answer these questions, we assemble a dataset that links workers' education and wages to the size of their employer for 57 countries at all levels of development. The data show that large firms are more skill-intensive than small firms across the development spectrum, and much more so in low-income countries. Skills are more expensive where they are scarce: the skill premium is higher in low-income countries. To answer the last two questions, we develop a heterogeneous-firm model in which firms choose between a traditional technology and a modern one that is more skill-intensive and has greater returns to scale. Calibrated to the United States, the model reproduces the decline in large-firm employment and the widening gap in skill intensity between large and small firms when we lower only the skill endowment, although neither outcome is targeted. Calibrated to each country, the model attributes 88% of the cross-country variation in the share of employment in large firms to differences in skill endowments. As the skill endowment rises, more firms adopt the modern technology, which adds almost a third to the resulting output gain.

Our first contribution is the data. We harmonize nationally representative labor force and household surveys for 57 countries, and measure for each country the skill endowment, the share of employment in large firms, the skill premium, and the skill intensity of large and small firms. Unlike business registers and firm surveys, which rarely record the characteristics of workers and miss informal firms, these surveys cover all workers, including the informal employment that accounts for most jobs in low-income countries. To our knowledge, no existing dataset combines workers' education, hourly wages, and employer size for as many countries across the development spectrum.[2]

We use the dataset to establish four facts. First, countries with lower skill endowments have less employment in large firms. The same positive link between skills and large-firm employment holds within broad sectors, within countries over time, and across workers within countries. Second, the skill premium is higher in lower-income countries. The ratio of the average hourly wage of high-skill to low-skill workers is 1.84 in low-income countries against 1.39 in high-income countries. Its estimated elasticity with respect to GDP per capita is -0.09 and statistically significant at the 1% level. This

[2] Engbom et al. [2025] also use labor force surveys to relate the skill composition of employment to firm size across countries. Our data also allow us to measure the relative price of skills, which is central to the relation between skill endowments and employment structure. We measure the skill premium from hourly wages for 56 countries, 20 of them low-income. This lets us document how the price of skills, the skill intensity of large and small firms and the size structure of employment vary together.

pattern is robust: all ten measures of the premium we compute decline with GDP per capita.[3] Third, large firms are more skill-intensive than small firms, across the development spectrum.[4] Fourth, the gap in skill intensity between large and small firms is larger in low-income countries. In high-income countries, the skill intensity of large firms only slightly exceeds that of small firms, while in low-income countries, it is twice as high.

Taken together, these facts point to the supply of skills as a constraint on large firms. The high skill premium in low-income countries (Fact 2) indicates that skills are scarce there because few workers acquire them, not because firms demand few of them.[5] A high skill premium raises the cost of the skill-intensive technology that large firms use (Fact 3), so that fewer firms adopt it and employment in large firms is low (Fact 1). Firms that do adopt it reduce their skill intensity by less than other firms when skills become expensive, which widens the gap between large and small firms (Fact 4). This chain of argument takes the skill premium as given, although it is an equilibrium outcome of skill supply and firms' employment choices. We therefore formalize it in a general equilibrium model, which we then use to quantify how much of the cross-country variation in large-firm employment skill supply can account for.

Our second contribution is a heterogeneous-firm model of skills and firm size in the tradition of Hopenhayn [1992], which we augment with the minimum ingredients needed to account for the four facts. Firms produce with low- and high-skill workers and choose between two technologies. The modern technology is more skill-intensive and has greater returns to scale than the traditional one, so firms that adopt it employ a larger share of high-skill workers and are larger, in line with Fact 3. The skill premium clears the labor market, so it is higher where the skill endowment is lower. We show that for any plausible value of the elasticity of substitution between high- and low-skill workers, a higher skill premium reduces the share of employment in modern firms and, under very weak additional conditions, the share of employment in large firms.[6]

Our third contribution is to quantify the mechanism. First, we calibrate the model

[3]Alternative measures vary the trimming of wages, the treatment of payment frequency, and also include the Mincer return to high skill conditional on age, sex, marital status, and payment frequency. The measure reported in the text yields the weakest relationship. The gradient is steeper when setting the high-skill threshold at thirteen years, and unchanged when the age window is widened to 16–64 – see Appendix C.1.

[4]This has previously been documented for a few rich countries [Brown and Medoff, 1989, Oi and Idson, 1999, Troske, 1999].

[5]Weak demand for skills, or for any input complementary to them, would depress the skill premium. For example, if large firms in low-income countries were held back by shallow financial markets, and capital and skills are complements, demand for skills would be low, and so would the skill premium. While the observed high skill premium does not rule out the presence of such factors reducing skill demand, it clearly indicates an important role of low skill supply. Hjort et al. [2026] make a similar argument for the price of managers.

[6]The conditions differ because technology and size are both choices in the model, so that "modern" and "large" are overlapping but distinct sets of firms.

to US data and vary the skill endowment only. As the endowment falls, employment in large firms declines as in Fact 1, and the gap in skill intensity between large and small firms widens as in Fact 4. These changes come quantitatively close to both the cross-country data and the U.S. time series, although neither is targeted. Second, we calibrate the model to each country in our data, allowing countries to differ also in the relative efficiency of high-skill labor in the two technologies and in the fixed cost of operating a firm. The calibrated model reproduces the targeted skill premium, skill intensities, and large-firm employment share of each country, and implies realistic values for the untargeted mean firm size. Decomposing the cross-country variance of large-firm employment into the contributions of each factor, we find that skill endowments account for between half and all of it, and for 88% by our preferred Shapley measure. Differences in the relative efficiency of high-skill labor, which can also capture schooling quality or distortions to relative skill demand like those discussed in footnote 5, account for close to half the dispersion in the skill premium, but do not help explain variation in the structure of employment.

Technology choice also amplifies the effect of skills on aggregate output. We compute the implied change in output from raising a low-income country's skill endowment to the US level, both with optimal technology choices and with the firm structure held fixed. This reveals that firms' optimal technology choices increase the gain from higher skill endowments by almost a third.

Our findings have implications for the measurement of distortions. In our calibrated model, the average product of labor of large firms exceeds that of small firms by about 8% in high-income countries and about 19% in low-income countries. This arises without misallocation or size-dependent distortions, since our model has no distortions and firms equate marginal products in a common labor market. One thus needs to be careful in interpreting observed differences in average products across firms: they do not necessarily indicate misallocation, but can also stem from firms' different choices of technology.

This approach has limitations. We take skill endowments as given and do not model why workers in low-income countries acquire fewer skills, so the results speak to the consequences of skill scarcity, not its causes. Our cross-country evidence is consistent with the proposed mechanism but does not identify it causally. Studies of Germany and Brazil, discussed below, find that increases in skill supply raise employment in large firms, which supports a causal reading. Finally, the model abstracts from other determinants of firm size, such as capital, management, and distortions. The 88% is what skill endowments account for in a model without these factors, and we leave their interaction with skills to future work.

Related literature. Our work contributes to several active literatures. It is related to the literature on the size structure of firms across countries. Low-income economies have high self-employment rates [Gollin, 2008] and small firms [Poschke, 2018, Bento and Restuccia, 2017, 2021]. A large literature attributes these differences to frictions and distortions, among them entry costs [Poschke, 2010, Moscoso Boedo and Mukoyama, 2012], labor market regulation [Hopenhayn and Rogerson, 1993, Poschke, 2009, Ulyssea, 2010], financial frictions [Buera et al., 2011, Midrigan and Xu, 2014], contract enforcement [Boehm and Oberfield, 2020] and delegation frictions [Akcigit et al., 2021, Grobovšek, 2020, Guner et al., 2018] or to generic wedges in particular for large firms [Restuccia and Rogerson, 2008, Hsieh and Klenow, 2009, Bartelsman et al., 2013].[7] Others have argued that small firm size may be an optimal response to differences in the level of capital [Gollin, 2008] or technology [Poschke, 2018]. With an exception discussed below, this literature has not related small firm sizes to economies' skill levels.

This paper is also related to a literature that studies the role of education in explaining income differences across countries [Jones, 2014, Caselli and Ciccone, 2013, Hendricks and Schoellman, 2023, Bils et al., 2024]. Three recent papers study the effect of skill supply on firms using variation within single countries. Goensch [2025] finds that the lower supply of high-skill workers reduces firm sizes and output in East relative to West Germany, and Cox [2025] and Haanwinckel and Soares [2021] find that increases in the supply of skills in Brazil reduced the skill premium and raised the share of employment in large firms. These papers evaluate the effect within a single country, while we assess the role of skill supply for the relationship between the size structure of employment and economic development across the full development spectrum.[8]

An important empirical finding of our paper is that the skill premium is higher in low-income countries. The compilation of Mincerian returns by Psacharopoulos [1994], updated in Psacharopoulos and Patrinos [2004], which underlies the development accounting of Hall and Jones [1999] and much subsequent work, suggested higher returns to schooling where schooling is scarce, whereas Banerjee and Duflo [2005] found that returns vary little with development. Rossi [2022] shows the importance of harmonized measurement, and finds a negative gradient across twelve countries using IPUMS International data. We measure the skill premium in several ways and establish that it declines with income across 56 countries, including many low-income countries.

Our model brings the literature on appropriate technology to the level of the firm. In this literature, the productivity of a technology depends on the match between its factor

[7] See Davis et al. [2023] for a recent review.

[8] A related strand of work emphasizes that the skills of the workforce shape not only the productivity of individual workers, but also which goods an economy can produce and how it organizes production [Kremer, 1993, Hausmann and Hidalgo, 2011, Gottlieb et al., 2026]. These papers focus on the complementarity across tasks or production inputs, rather than the price of skills.

intensity and the economy's endowment, so that countries with few skilled workers rationally use technologies that need few of them [Basu and Weil, 1998, Acemoglu and Zilibotti, 2001, Caselli and Coleman, 2006]. We model this choice at the firm level and tie it to firm size, so that the skill endowment shapes both the technologies firms use and the size structure of employment. This lets us discipline technology choice with facts on skill intensity by firm size across countries, and quantify its contribution to the output gains from more skills.[9]

Finally, our work builds most directly on recent papers showing that skill scarcity shapes the organization of production in low-income countries. Hjort et al. [2026] find much higher relative salaries of managers and business professionals in poorer countries, using data from an international consulting firm. Engbom et al. [2025] show that large firms rely on hierarchies of white-collar workers and that low educational attainment limits their supply. In their model, skills create large firms mainly because more skilled individuals become entrepreneurs and professionals who run these firms. We take the next step by documenting patterns not only in the quantity but also in the price of skills, and study how this price shapes firms' hiring and technology adoption. We focus on skills of the workforce as a whole, not only those of white-collar workers, since ultimately it is the skills of production workers that shape the adoption of production technology.

The remainder of the paper is structured as follows. Section 1 describes the data and the measurement of the size structure of employment, skills, and wages. Section 2 documents the four facts. Section 3 presents the model, Section 4 its calibration, and Section 5 the quantitative results. Section 6 explores implications for output, and relates our findings to the literature on misallocation. Section 7 concludes.

1. Data and Measurement

Assessing the role of skill scarcity for the size structure of firms across countries places four requirements on the data. The data must (i) link each worker's educational attainment to the size of the worker's employer, (ii) report individual wages and hours worked, (iii) cover countries at all levels of development, and (iv) be representative of total employment, including informal employment, which accounts for a large share of employment in low- and middle-income countries.

To our knowledge, no data source used in the literature meets all four requirements. Business registers and firm-level surveys report employer size but not the characteristics of individual workers, so they fail requirements (i) and (ii). Firm surveys that do

[9] Our assumption of heterogeneity in firms' returns to scale, and its relationship to firm size, is in line with recent evidence from Gao and Kehrig [2025] and Ozkan et al. [2026]. Kopytov et al. [2026] build a model where firms endogenously choose returns to scale.

collect workforce information, such as the World Bank Enterprise Surveys,[10] cover only registered firms with five or more employees. In low-income countries, where small and informal firms account for most employment, this sample is not representative, so they fail requirement (iv). Harmonized census microdata, such as IPUMS International, report educational attainment for many countries, but employer size and hourly wages for few. Only 13 countries meet requirements (i) and (ii), too few to span the development spectrum, so that this data source fails requirement (iii). The Demographic and Health Surveys (DHS) cover many low-income countries but are designed to study health rather than labor markets. They record employment status and occupation, but neither employer size nor earnings, and therefore fail requirements (i) and (ii).

Labor force and household surveys meet all four requirements. They are nationally representative, so they cover informal employment, and they record educational attainment, wages and hours worked, and the size of the employer for each respondent. We assemble labor force and household surveys that report these variables and harmonize the measurement of firm size, education, wages, and hours across them.[11] The result is a dataset that measures skill endowments, the firm size structure of employment, and the skill premium consistently for 57 countries at all levels of development.

Survey Selection. Our main analysis uses a cross-section of countries around 2015. We use one survey per country, choosing the year closest to 2015, the most recent year with broad coverage. The sample spans the full range of income per capita, from Niger in 2012 ($1,198 per capita) to Switzerland in 2015 ($67,268 per capita).[12] Of these 57 countries, all but Angola also report wages. Some countries are observed in several years and we use this panel dimension as well. Appendix Table A.I lists the countries, survey years, data sources and whether a country-survey is part of the panel.

Sample selection. Firm size and wages are only reported for wage employees, who form our baseline sample. Within this group, we keep working-age (25–55) individuals with non-missing information on their level of education and the size of the firm they work in. For wage analysis, we further restrict the sample to workers reporting positive hours, positive wages, and a valid reference period for the reported wage. Overall, our baseline sample consists of 1,162,815 individual observations.[13] Because the self-employed account for a large share of employment in low-income countries, we report

[10] See https://www.enterprisesurveys.org/en/methodology

[11] The harmonization is part of a larger effort to build the Harmonized World Labor Force Survey (HWLFS), a microdataset that harmonizes nationally representative labor force and household surveys for more than 120 countries. The surveys used here are those in which respondents report the size of their employer.

[12] GDP per capita is from Feenstra et al. [2015].

[13] Observations with a missing or invalid reference period cannot be converted to an hourly wage. We also drop the rare cases in which survey weights are missing.

a robustness exercise in which we assign them to small-firm employment; most are own-account workers without employees, who effectively operate single-person firms.

Measuring firm size. All surveys ask wage workers a version of the question:

How many employees work at your place of work (establishment/work site)?

Survey responses are typically reported in discrete size bins (e.g., 1–4, 5–9, 10–49, 50 or more). We harmonize responses into two categories: *small firms* (fewer than ten employees) and *large firms* (ten or more employees). The ten-employee cutoff is a bin boundary in every survey, and this binary split is the finest distinction the data support consistently across countries. The question refers to the respondent's establishment rather than the firm. For single-establishment firms, which include almost all small firms, the distinction is immaterial. For multi-establishment firms, establishment size understates firm size. Because such firms are concentrated in high-income countries, our measure, if anything, understates the difference in large-firm employment between rich and low-income countries. We denote the share of wage employment in large firms by ℓ^b ("big") and that in small firms by $\ell^s = 1 - \ell^b$ ("small").

Measuring skills. We classify workers by years of schooling.[14] Workers with at least ten years of schooling are *high-skill*, and those with nine years or fewer are *low-skill*. We denote their shares in a country's wage employment by L_h and $L_l = 1 - L_h$.

In most countries, the nine-year threshold coincides with completing lower secondary education, corresponding to level 2 of the International Standard Classification of Education (ISCED), which marks the transition from generalist to subject-oriented instruction. We adopt this threshold rather than a college-education cutoff commonly used in the literature for two reasons. First, the college-educated share of the population is very small in low-income countries, leaving too few observations for a meaningful comparison across the development spectrum. Second, the skills most relevant for productive participation in a large organization are plausibly advanced literacy and numeracy, for which lower secondary completion is a natural threshold. We show below that the facts we document are robust to an alternative thirteen-year threshold.

Wages. Our earnings measure is total labor income from wage employment, including the value of in-kind payments where surveys distinguish them from cash. Survey respondents report an earnings amount together with its reference period, typically a day, a week, or a month. We convert earnings to an hourly wage by expressing them

[14] Some surveys only provide information on completed degrees. For these, we impute years of schooling based on the standard duration of the corresponding level in the country's education system.

on a weekly basis and dividing by hours worked in the reference week.[15] To limit the influence of outliers, we trim hourly wages at the 1st and 99th percentiles within each country. We denote the average hourly wage of high- and low-skill workers by w_h and w_l, and define the skill premium as their ratio, $\omega \equiv w_h/w_l$. As a second measure of the skill premium, we estimate for each country the conditional return to high skill, the coefficient on a high-skill indicator in a Mincer regression of log hourly wages that controls for age, age squared, sex, marital status, and the wage reporting frequency.[16]

Validation. We validate two of our measures against published cross-country datasets. The correlation of countries' average years of schooling in our data with that from the Barro–Lee series [Barro and Lee, 2013], the standard source on educational attainment, is 0.89. Our levels are slightly higher because they cover wage employees rather than the population. No published source reports the share of employment in large firms, so we compare the cross-country gradient of our measure with two firm-side sources, the average establishment size that Bento and Restuccia [2021] collect from economic censuses and mean firm size that Poschke [2018] computes from the Global Entrepreneurship Monitor. Our measure correlates at 0.77 with the former and 0.60 with the latter. Our data thus reproduce the cross-country patterns of firm size found in sources covering firms. Further details on the dataset comparisons can be found in Appendix B.

2. Cross-country evidence

In this section, we use our dataset to document four facts about the skill endowment of the workforce, the size structure of employment, the skill premium, and the skill intensity of large and small firms across countries. For each fact, we present the evidence, state the fact, and summarize robustness checks, which we report in detail in the appendix.[17]

[15] For surveys in which actual hours in the reference week are not reported, we use usual weekly hours. The conversion of daily earnings rests on an assumed number of five days worked, and monthly earnings are divided by 4.33. As a robustness check, we exclude workers paid by the day (Table C.I).

[16] Payment frequency matters because in lower-income countries, low-skill workers are more often paid daily or weekly. Higher-frequency pay carries a premium, presumably reflecting the instability of such jobs. This implies that, although high-skill workers earn more at every frequency, the overall ratio of average high- to low-skill wages is biased downward, and in some countries can fall below one (by Simpson's paradox), if high-skill workers are paid less frequently. The Mincer regression therefore includes payment frequency as a control.

[17] They include alternative definitions of high skill and of large firm size, alternative measures of the skill premium, including the self-employed, widening the age window, the role of sectors, changes over time, and a worker- rather than a country-level perspective.

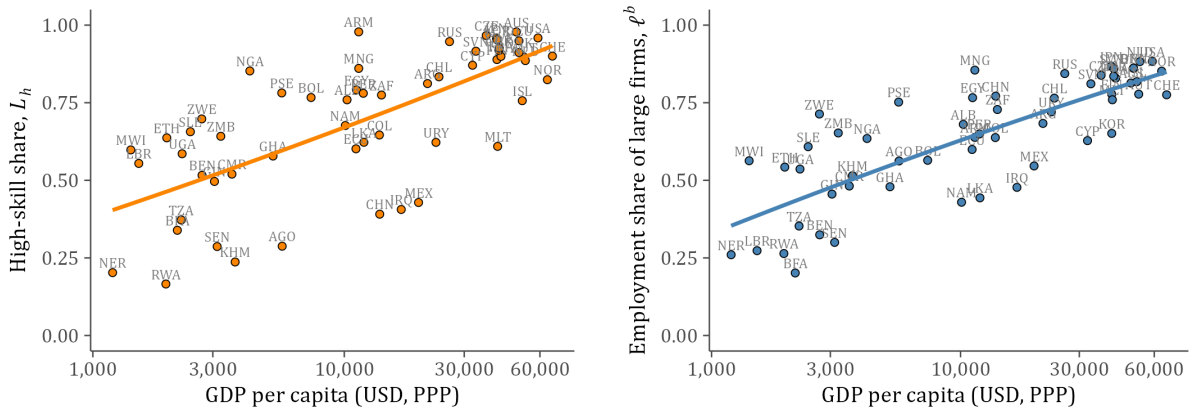
2.1. Skill endowments and the size structure of employment

The size structure of employment tracks the skill endowment closely across countries. Figure I above plots the share of wage employment in large firms against the share of high-skill workers for the 57 countries in our data. Where around half of workers are high-skill, about one half of employment is in large firms; where more than three quarters of workers are high-skill, as in most high-income countries, the share of employment in large firms is likewise close to three quarters. The correlation between the two shares is 0.81, and 0.56 conditional on log GDP per capita, so the relationship is not a by-product of both shares rising with income.[18]

Both shares are also increasing with country income level. Figure II plots them against GDP per capita and Table I reports their averages by country income group, with the United States shown separately.[19] In low-income countries, 51.3% of wage workers have at least ten years of schooling, compared to 89.3% in high-income countries and 95.9% in the United States. The high-skill share in high-income countries is thus 1.74 times that in low-income countries, and 2.25 times under the thirteen-year threshold. The share of wage employment in large firms against GDP per capita rises just as sharply, from 47.8% in low-income countries to 81.0% in high-income countries, in line with Poschke [2018] and Bento and Restuccia [2017, 2021], who document that average firm size rises with country income levels.

[18]The correlation is 0.74 when "high-skill" is defined by the thirteen-year threshold, 0.87 when the skill endowment is measured by average years of schooling in the labor force rather than wage employment only, and 0.71 when large firms are those with at least 50 employees (Appendix Figures C.II, D.Ib and C.VI).

[19]Table C.II and Table C.IX report the same statistics for a different definition of skill and when including the self-employed, respectively.



(A) Share of high-skill workers.

(B) Share of employment in large firms.

FIGURE II
Skill endowments and employment by firm size across countries

Notes. In both panels, each marker is one country, labeled with its ISO3 country code. Panel (a) plots the share of high-skill wage employees, and Panel (b) the share of wage employment in large firms, both against GDP per capita (PPP, log scale). The solid line in each panel is a quadratic (in log GDP per capita) least-squares fit across countries. High-skill denotes at least ten years of schooling and large firms have at least ten employees. Sample: wage employees aged 25–55 with non-missing education and firm-size information. GDP per capita is PPP-adjusted [Feenstra et al., 2015]. The surveys used for each country are listed in Table A.I.

We summarize these findings as

Fact 1. *Countries with lower skill endowments have less employment in large firms.*

Robustness. A first concern is that the pattern reflects sectoral composition, since low-income countries have more employment in agriculture and agricultural production occurs in small firms. Appendix D.2 shows otherwise. Although the share of employment in large firms is indeed lower in agriculture than in industry and services, it is lower in low-income countries within each of the three broad sectors, and the positive relationship between skills and large-firm employment holds within sector. A second concern is the high level of self-employment in low-income countries, which our baseline sample excludes. Appendix C.4 treats the self-employed as employed in small firms and shows steeper gradients with GDP per capita for both skills and large-firm employment (Figure C.X). As the self-employed in low-income countries are predominantly low-skill, including them raises the correlation between the share of employment in large firms and the share of high-skill workers to 0.89. A third concern is where the firm-size threshold is drawn. Appendix C.3 redefines large firms as those with at least 50 employees and reproduces every exhibit of this section; the correlation between the two shares is then 0.71 (Table C.VI and Figure C.VII).

A fourth concern is that the cross-country correlation reflects fixed country characteristics rather than skills as such. To address it, we use the countries observed over multiple years and build a balanced panel of 16 countries observed annually from 2007

TABLE I
Skill endowment, skill intensity of employment by firm size and the skill premium

	Country income			
	Low-income	Middle-income	High-income	US
Employment composition				
Years of schooling	9.1	11.4	13.6	14.0
Share of high-skilled workers L_h	51.3%	70.2%	89.3%	95.9%
Share of employment in large firms ℓ^b	47.8%	66.1%	81.0%	88.3%
Skill premium				
Unconditional: mean wage ratio $\omega = w_h/w_l$	1.84	1.69	1.39	1.79
Conditional: returns to skill (log points)	0.69	0.45	0.32	0.57
Probability of working for a large firm by skill group				
High skill L_h^b/L_h	61.4%	73.9%	81.7%	88.9%
Low skill L_l^b/L_l	34.4%	49.3%	74.4%	73.9%
Skill intensity				
Large firms h^b	64.6%	78.0%	90.1%	96.5%
Small firms h^s	40.1%	57.0%	86.4%	90.7%
Ratio (large / small)† h^b/h^s	2.03	1.50	1.05	1.06
Number of countries	21	17	19	1

Notes. Each column reports the average of the relevant country-level statistic within an income group, with the United States shown separately for reference. Countries are grouped by GDP per capita (PPP): low-income (LIC, below $10,000), middle-income (MIC, $10,000–$30,000) and high-income (HIC, above $30,000). The *Employment composition* block reports average years of schooling, the share of high-skill wage employees and the share of wage employment in large firms. The *Skill premium* block reports two measures, both averaged across countries: the unconditional ratio of average high- to low-skill wages, and the conditional Mincer return, i.e. the country-level coefficient on a high-skill indicator in a log-wage regression controlling for age, age^2, sex, marital status and the wage-reporting frequency. Both are computed on the wage sample after trimming the top and bottom 1% of log wages within each country, and the Mincer return is in log points. The *Probability of working for a large firm by skill group* block reports the share of high-skill, and of low-skill, workers who are employed in a large firm. The *Skill intensity* block reports the high-skill share within large and within small firms and their ratio. † The ratio is the average of the country-level ratios, and hence differs from the ratio of the average shares reported in the two rows above it. The final row gives the number of countries in each group. High-skill denotes at least ten years of schooling and large firms have at least ten employees. Country-level statistics use survey weights; countries are weighted equally within an income group. Sample: wage employees aged 25–55 with non-missing education and firm-size information. The surveys used for each country are listed in Table A.I; the skill-premium rows use the 56 countries that also report wages.

to 2017.[20] In a panel regression with country and year fixed effects, a one-percentage-point increase in the high-skill share is associated with an increase of 0.339 percentage points in the share of employment in large firms, as shown in Table II. The estimate is statistically significant at the 5% level and economically large. Appendix Figure D.V shows the same relationship as a partial-regression scatter of the two shares, each purged of country and year fixed effects. The within-country relationship is mostly driven by low- and middle-income countries, since in high-income countries both shares are already high and change little over the sample period.

[20]The panel comprises 9 high-income and 7 low- and middle-income countries – see Appendix Table A.I.

TABLE II
Changes in skills and large-firm employment over time

	Large-firm share ℓ^b
	(1)
High-skill share L_h	0.339**
	(0.119)
Country fixed effects	✓
Year fixed effects	✓
No. of countries	16
Observations	176
Dependent variable mean	0.748
R^2	0.986
Within R^2	0.165

Notes. Within-country panel regression of the large-firm employment share on the high-skill share, including country and year fixed effects. The sample is a balanced panel of 16 countries observed in every year from 2007 to 2017 (176 country–year observations), built from annually repeated household and labour force surveys. The table also reports the number of countries, the mean of the dependent variable, and the R^2 and within R^2. Appendix Figure D.V shows the same estimate as a partial-regression scatter. Standard errors are clustered by country. High-skill denotes at least ten years of schooling and large firms have at least ten employees. Sample: wage employees aged 25–55 with non-missing education and firm-size information; survey weights are rescaled to one within each country–year cell, so that each country counts the same regardless of its population or survey sample size. The panel countries are listed in Table A.I. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

2.2. The skill premium

Figure IIIa plots the skill premium against GDP per capita. It averages 1.39 in high-income countries, 1.69 in middle-income countries, and 1.84 in low-income countries. High-skill workers thus earn 84% more than low-skill workers in low-income countries, against 39% in high-income countries, a gap 2.14 times as large. Fifteen of the twenty low-income countries in our data have a higher skill premium than the median high-income country. The United States, at 1.79, lies at the high end of the high-income group. A regression of the log skill premium on log GDP per capita implies an elasticity of -0.09, significant at the 1% level.

The gradient does not reflect differences in the composition of the population. Figure IIIc plots, for each country, the coefficient on the high-skill dummy in a Mincer regression of log hourly wages that controls for age, sex, marital status and payment frequency. The conditional return to being high-skill averages 0.32 log points in high-income countries, 0.45 in middle-income countries, and 0.69 in low-income countries, 2.13 times the high-income level. Its elasticity with respect to GDP per capita is -0.14, significant at the 1% level. If anything, the gap between low- and high-income countries

widens once observables are controlled for.[21]

Figure IIIb and Figure IIId plot each country's raw skill premium and Mincer coefficient against the country's skill endowment L_h, rather than against GDP per capita. Both measures decline with L_h. The price of skill is highest where skilled workers are scarcest. We summarize this as

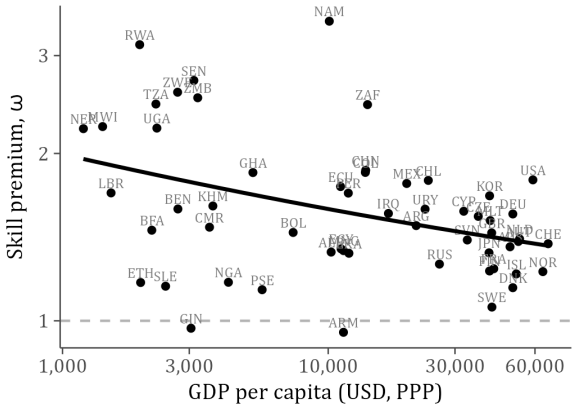
Fact 2. *The skill premium is higher in countries with lower incomes or lower skill endowments.*

Robustness. Under the thirteen-year skill threshold, the gradient of the skill premium with respect to income per capita is even steeper. The premium averages 1.30 in high-income countries, 1.91 in middle-income countries, and 2.20 in low-income countries. Its elasticity with respect to GDP per capita is -0.16 (Appendix C.2, Figure C.IV and Table C.III). The relationship between the skill premium and GDP per capita is present for eight different measures of the premium that vary in terms of the trimming of wages and the treatment of payment frequency.[22] It is present for both skill thresholds (Figure C.Ia and Table C.I), and it survives widening the age window to 16–64 (Appendix C.5).

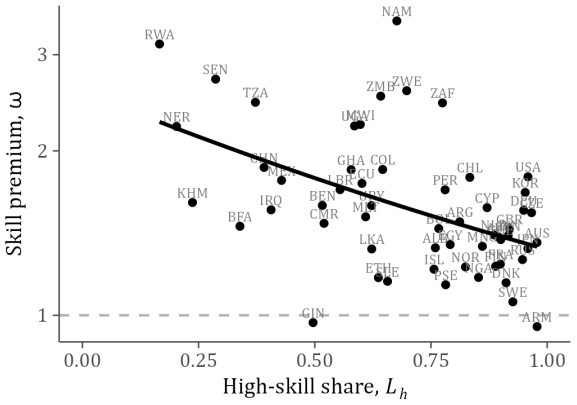
[21] The pattern is the same in Mincer regressions on years of schooling and under the thirteen-year threshold (Appendix Figures C.Ib and C.IVc).

[22] We vary trimming and sample selection along two dimensions. First, we apply four different sample selection rules: no trimming; restricting the sample to workers paid at the country's most common frequency; trimming the top and bottom 1% of wages (our main measure); and the last two combined. Second, we consider both the mean and the median wage for each skill group. This implies eight different measures overall.

Panel A: Skill premium

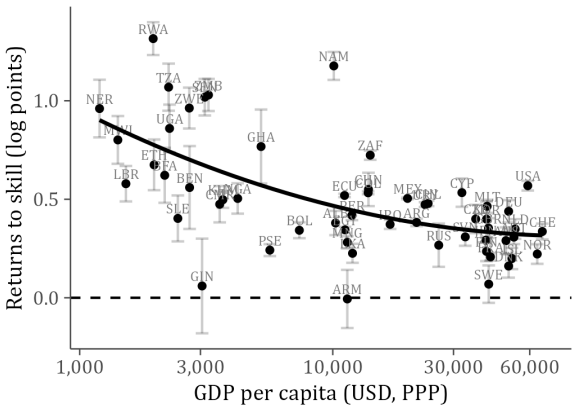


(A) Against GDP per capita.

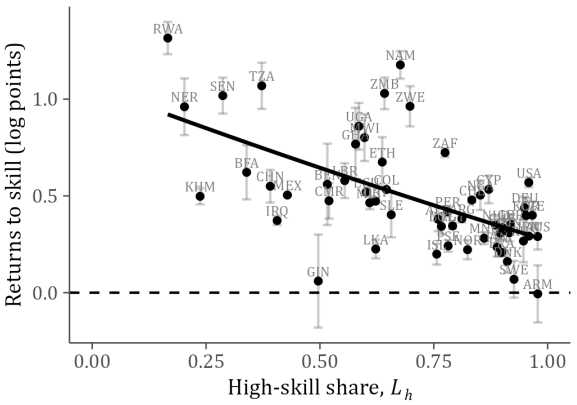


(B) Against the skill endowment L_h.

Panel B: Mincer coefficient



(C) Against GDP per capita.



(D) Against the skill endowment L_h.

FIGURE III
Skill prices across countries

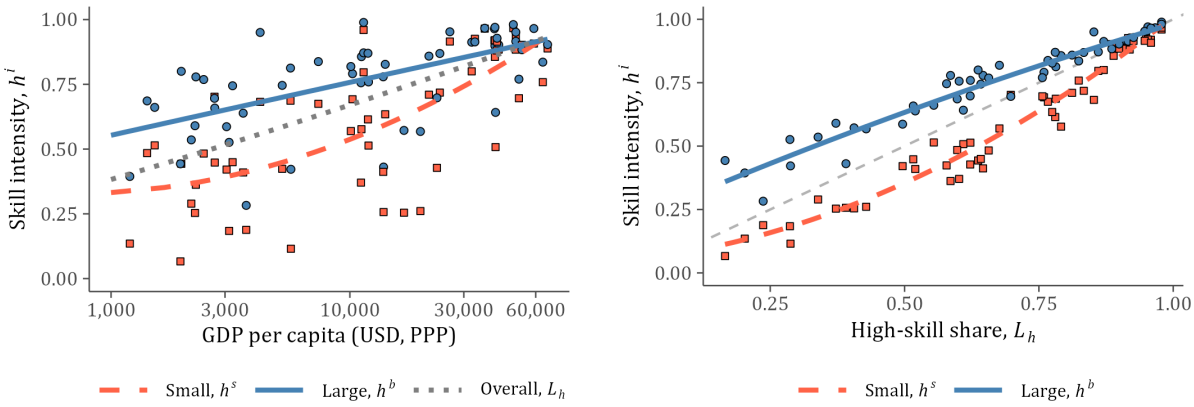
Notes. Each marker is one country, labeled with its ISO3 country code. The figure crosses two measures of the price of skill with two horizontal axes. *Panel A*, Figures (a) and (b), shows the skill premium: the ratio of average wages of high-skill to low-skill wage employees, with the top and bottom 1% of log wages trimmed within each country. Its vertical axis is on a log scale, so the fitted line is fitted to the log premium, and the dashed horizontal line marks parity (no skill premium). *Panel B*, Figures (c) and (d), shows the country-level coefficient on a high-skill indicator from a Mincer regression of log wages on the indicator and individual controls (age, age^2, sex, marital status and the wage-reporting frequency), estimated separately for each country; whiskers are 95% confidence intervals and the dashed horizontal line marks a coefficient of zero. Within each panel, the left-hand figure plots the measure against GDP per capita (PPP, log scale) and the right-hand one against the country's high-skill share L_h, the share of high-skill workers among wage employees. The solid line is in every case a quadratic least-squares fit across countries — in log GDP per capita on the left, and in the high-skill share, which is not logged, on the right. All four figures use the same trimmed wage sample, so the conditional and unconditional measures of the price of skill rest on identical wages. High-skill denotes at least ten years of schooling. Sample: wage employees aged 25–55 reporting positive wages and hours, one survey per country (the survey year closest to 2015); survey weights are rescaled within each country–year cell. The surveys used for each country are listed in Table A.I. GDP per capita is PPP-adjusted [Feenstra et al., 2015].

2.3. Skill intensity of large and small firms

The joint measurement of worker skill and employer size allows us to compare the skill intensity of large and small firms within countries. Figure IVa plots the share of high-skill employees in large firms, in small firms, and in wage employment as a whole against GDP per capita, and Figure IVb plots the same three series against the skill endowment L_h.

Figure IVa shows that large firms are more skill-intensive than small firms in all but three countries.[23] As a result, the line of best fit for large firms (blue) lies above that for small firms (red) across the entire income range. Column (4) in Table III shows that on average, the share of high-skill workers in large firms is 16.6 percentage points greater. Thus,

Fact 3. *Large firms are more skill-intensive than small firms.*



(A) Against GDP per capita (B) Against the skill endowment

FIGURE IV
The skill intensity of large and small firms

Notes. Each marker is one country. Panel (a) plots the high-skill share of small firms, h^s, and of large firms, h^b, against GDP per capita; the lines are least-squares fits on a quadratic in log GDP per capita, and the dotted line is the same fit for the economy-wide share L_h. Panel (b) plots the same shares against the country-wide high-skill share L_h, with least-squares fits on a quadratic in L_h; the dashed line is the 45-degree line, which the economy-wide share follows by construction. High-skill denotes at least ten years of schooling and large firms have at least ten employees. Sample: wage employees aged 25–55 with non-missing education and firm-size information. GDP per capita is PPP-adjusted [Feenstra et al., 2015]. The surveys used for each country are listed in Table A.I.

[23] The exceptions are the Netherlands, Slovenia and Zimbabwe.

TABLE III

The skill gap in large-firm employment, from the worker and the firm side

	Working in large firm			Skill intensity h^i		
	(1)	(2)	(3)	(4)	(5)	(6)
Unit of observation	Worker			Country × firm size		
1{High-skill}	0.231***	0.213***	0.210***			
	(0.018)	(0.016)	(0.017)			
1{High-skill} × log GDP p.c.		-0.046***				
		(0.014)				
1{High-skill} × high-skill share			-0.275***			
			(0.076)			
1{Large firm}				0.166***	0.166***	0.166***
				(0.017)	(0.013)	(0.012)
1{Large firm} × log GDP p.c.					-0.068***	
					(0.010)	
1{Large firm} × high-skill share						-0.396***
						(0.054)
Controls	✓	✓	✓			
Country fixed effects	✓	✓	✓	✓	✓	✓
No. of countries	57	57	57	57	57	57
Observations	1,162,815	1,162,815	1,162,815	114	114	114
R^2	0.189	0.191	0.191	0.928	0.958	0.964
Within R^2	0.044	0.047	0.047	0.631	0.786	0.813

Notes. The table measures the same relationship from two sides, with country fixed effects throughout. In columns (1)–(3) the unit of observation is a worker: these are linear probability regressions of an indicator for working in a large firm on an indicator for being high-skill, controlling for age, age^2, sex and marital status, with observations weighted by the rescaled survey weights. In columns (4)–(6) the unit of observation is a country–firm-size cell, so there are two per country: the outcome is the share of high-skill workers in the cell and the regressor is an indicator for the large-firm cell, with cells weighted equally so that every country counts alike. In both blocks the coefficient on the first row is therefore the gap between large and small firms, read either as the extra probability that a skilled worker is employed in a large firm or as the extra share of skilled workers that large firms employ. Columns (2) and (5) interact that gap with log GDP per capita, and columns (3) and (6) with the country's share of high-skill workers. Both moderators are constant within a country, so the fixed effects absorb their levels, and only the interactions are identified; both enter centered on their means, which leaves the interaction coefficients and their standard errors unchanged and makes the first row the gap at average income and at average skill supply, respectively. The two blocks are in different units and their coefficients are not comparable in magnitude; what is comparable is the sign of the interactions, which is negative on both sides and under both moderators. Standard errors are clustered by country, with $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. High-skill denotes at least ten years of schooling and large firms have at least ten employees. Sample: wage employees aged 25–55 with non-missing education and firm-size information. GDP per capita is PPP-adjusted [Feenstra et al., 2015]. The surveys used for each country are listed in Table A.I.

The same pattern holds for workers. Pooling the 1,162,815 individual observations in our data, the share of wage workers employed in a large firms rises with education from around 28.6% for workers with primary education or less to around 82.3% for those with a university degree (Appendix Figure D.VI). Within countries, and conditional on age, sex and marital status, a high-skill worker is 23.1 percentage points more likely than a low-skill worker to be employed in a large firm (column (1) of Table III).[24]

Robustness. The ranking is unchanged under the thirteen-year skill threshold (Appendix C.2, Table C.II and Figure C.Va), under the fifty-employee threshold for large firms (Appendix C.3, Table C.VI and Figure C.VIIIa), and when the self-employed are counted as small-firm employment (Appendix C.4, Table C.IX and Figure C.XIa).

Fact 4. *The skill-intensity gap between large and small firms is greater in low-income and in low-skill countries.*

Where skills are abundant, small firms employ a workforce almost as skill-intensive as large firms. In the United States, 96.5% of workers in large firms and 90.7% of workers in small firms are high-skill. In countries where skills are scarce, the gap is wide. In low-income countries, the two shares are 64.6% and 40.1%. Figure IVb relates the gap directly to a country's skill endowment. As L_h falls, the high-skill share of small firms falls faster than that of large firms, which remain comparatively skill-intensive while small firms fall behind.

Table III quantifies the gradient. Columns (5) and (6) interact the large-firm indicator variable with log GDP per capita and the country's high-skill share. Both interactions are negative and significant at the 1% level, so the skill intensity gap between large and small firms shrinks as countries become richer and as high-skill workers become more abundant. Columns (2) and (3) show a similar pattern for workers: In richer or more skill-abundant countries, a worker's probability of being employed in a large firm varies less with the worker's skill.

Robustness. The gradient of the skill intensity gap with respect to income per capita and the skill endowment is steeper under the thirteen-year skill threshold, where large firms in the lowest-income countries are on average more than 3.41 times as skill-intensive as small firms (Appendix C.2, Table C.II and Figure C.Vb). The same pattern holds for the fifty-employee threshold for large firms (Appendix C.3) and when the self-employed are counted as small-firm employment (Appendix C.4, Table C.IX and Figure C.XIb). The interactions of Table III keep their sign and significance in all three exercises (Table C.IV, Table C.VII and Table C.X).

Taken together, these facts suggest that a scarce supply of skills — and the high

[24]Without country fixed effects, the gap is 32.6 percentage points; Appendix Table D.I reports that specification and the coefficients on the controls. Under the thirteen-year threshold, the coefficient is 0.269 (0.195 with country fixed effects), significant at the 1% level (Appendix Table C.V).

price of skills it implies — is a likely reason why low-income countries have few large firms. Importantly, our findings are *not* consistent with the reverse channel, namely the hypothesis that the absence of large firms leads to a lack of incentives to acquire skills. Such a channel would imply a low skill premium in low-income countries, while we find a very high one. The remainder of the paper formalizes this argument in a model and quantifies how much of the cross-country variation in large-firm employment and aggregate output skill supply can account for.

3. Model

We develop a static model of heterogeneous firms in the spirit of Hopenhayn [1992], in which a workforce's skill shapes firms' technology choices, the firm size distribution and aggregate output. As we present the model, we establish a series of results that mirror the facts shown in Section 2. We begin by showing how a greater skill premium, which will be the equilibrium consequence of a lower skill endowment (Fact 2), affects firms' choices of skill mix (Facts 3 and 4) and technology. Two final results show why, and under what conditions, a higher skill premium implies less employment in large firms (Fact 1). A quantitative analysis follows in the next section.

3.1. Environment

Consider a static economy populated by a measure one of workers and an endogenous measure M of firms. There are L_h high-skill and $L_l = 1 - L_h$ low-skill workers. Workers supply labor inelastically and consume all their income. Active firms incur a fixed overhead cost and produce a homogeneous final good, which serves as the numeraire. Firms are created by potential entrants, who pay an entry cost, draw their productivity level z, and then decide whether to operate and if so, with which technology. All markets are competitive.

3.2. Technology

Firms have access to two technologies, a modern and a traditional one, which we index by superscript $i \in \{m, t\}$. Each firm produces the final good using high- and low-skill labor, L_h^i and L_l^i. We abstract from capital. The two inputs are combined in a CES aggregator with weight μ^i for the high-skill, $1 - \mu^i$ for the low-skill, and elasticity of substitution ρ. Production has decreasing returns to scale, governed by the parameter γ^i, where $\mu^i, \gamma^i \in (0, 1), \rho > 0$. The output of a firm with productivity z that uses

technology i is then given by

$$y^i(z) = z\left[(1-\mu^i)L_l^{i\frac{\rho-1}{\rho}} + \mu^i L_h^{i\frac{\rho-1}{\rho}}\right]^{\frac{\rho}{\rho-1}\gamma^i} = z\bar{L}^{\gamma^i}, \tag{1}$$

where $\bar{L}$ is the composite labor input. We assume that the two technologies differ in two ways. First, the modern technology has greater returns to scale, i.e. $\gamma^m > \gamma^t$. This will imply that the firms that use it are optimally larger. Second, modern technology uses skilled labor more productively, i.e. $\mu^m > \mu^t$. This will imply that firms using it are more skill-intensive. Because firms choose their technology as well as their size, note that "modern" does not coincide with "large" as measured in the data. We make this mapping explicit in Section 3.6.

3.3. Labor demand

Each firm chooses high- and low-skill labor inputs to maximize profits, taking the low- and high-skill wage rates w_l and w_h as given. We denote the skill premium by $\omega \equiv w_h/w_l$. It is convenient to solve this problem in two steps: first the optimal input mix, then optimal scale.

The input mix. Solving the dual cost minimization problem,[25] the unit cost of the composite labor input $\bar{L}^i$ for a firm of type i is

$$\bar{w}^i = \left[(1-\mu^i)^\rho w_l^{1-\rho} + (\mu^i)^\rho w_h^{1-\rho}\right]^{\frac{1}{1-\rho}}. \tag{2}$$

This is common for firms with the same technology, but varies between technologies with μ^i. Its elasticity with respect to the skill premium ω equals the cost share of skilled labor, Λ_h^i.[26] Since $\mu^m > \mu^t$ implies $\Lambda_h^m > \Lambda_h^t$,

Lemma 1. *The unit cost of modern relative to traditional firms $\bar{w}^m/\bar{w}^t$ increases in the skill premium ω.*

Optimal relative skill use satisfies the familiar condition

$$\frac{L_h^i(z)}{L_l^i(z)} = \left(\frac{\mu^i}{1-\mu^i}\frac{w_l}{w_h}\right)^\rho = \left(\frac{\mu^i}{1-\mu^i}\frac{1}{\omega}\right)^\rho, \tag{3}$$

[25] That is, the firm minimizes the wage bill $w_l L_l + w_h L_h$ subject to attaining a given quantity of composite labor $\bar{L}$. Since the CES aggregator exhibits constant returns to scale, optimal skill intensity and the unit cost are independent of scale. They only depend on parameters and prices.

[26] This is given by $\Lambda_h^i = \left[1 + \left(\frac{1-\mu^i}{\mu^i}\right)^\rho \left(\frac{w_l}{w_h}\right)^{1-\rho}\right]^{-1}$. The statement is for a change in w_h at fixed w_l.

which is independent of scale and declines in the skill premium with elasticity ρ. We denote the implied skill intensity of production employment by $h^i \equiv L_h^i/(L_h^i + L_l^i)$.

Lemma 2. *Since the skill premium is common across firms and $\mu^m > \mu^t$, modern firms optimally choose a higher skill intensity $h^m > h^t$.*

Lemma 3. *The elasticity of skill intensity h^i with respect to the skill premium ω equals (minus) $\rho(1-h^i)$, which is smaller in absolute value for modern, more skill-intensive firms.*[27]

We show below that modern firms are larger than traditional ones. It follows that Lemma 2 aligns with Fact 3: large firms are more skill-intensive than small ones. Lemma 3 speaks to Fact 4, the wider skill-intensity gap between large and small firms in low-skill countries: when skilled labor becomes more expensive, skill-intensive modern firms reduce their skill intensity by less than traditional firms do.

Note that the facts are about large and small firms, and the lemmas about modern and traditional ones. The lemmas can be extended to firm size, and we make this explicit in Section 3.6.

Scale. Given the optimal input mix, the optimal level of composite labor input for a firm with technology i and productivity z is[28]

$$\bar{L}^i(z) = \left(\frac{z\gamma^i}{\bar{w}^i}\right)^{\frac{1}{1-\gamma^i}}. \tag{4}$$

For any level of productivity z, the optimal size of a firm using the modern relative to one using the traditional technology is

$$\frac{\bar{L}^m(z)}{\bar{L}^t(z)} = \hat{\hat{\gamma}}\, z^{\frac{1}{1-\gamma^m}-\frac{1}{1-\gamma^t}}\, \frac{(\bar{w}^t)^{\frac{1}{1-\gamma^t}}}{(\bar{w}^m)^{\frac{1}{1-\gamma^m}}}, \tag{5}$$

where $\hat{\hat{\gamma}} \equiv (\gamma^m)^{\frac{1}{1-\gamma^m}}/(\gamma^t)^{\frac{1}{1-\gamma^t}}$ is a constant. Because $\gamma^m > \gamma^t$, the exponent on z is positive, so the relative size of modern firms grows with productivity. The term in z plays no role in the comparative statics with respect to the skill premium that follow.

Lemma 4. *Optimal composite employment of modern relative to traditional firms declines in the skill premium.*

This occurs for two reasons. First, a higher skill premium raises the unit cost of modern firms by more (Lemma 1). Second, the greater returns to scale of modern firms makes their scale more responsive to unit costs. Combining the two, the elasticity of relative size with respect to the skill premium is $\Lambda_h^t/(1-\gamma^t) - \Lambda_h^m/(1-\gamma^m) < 0$.

[27] This arises because the elasticity of h^i with respect to L_h^i/L_l^i is $1-h^i$, and the elasticity of L_h^i/L_l^i with respect to ω is $-\rho$.

[28] This follows from maximizing $\pi^i(z) = z\bar{L}^{\gamma^i} - \bar{w}^i\bar{L}$ with respect to $\bar{L}$.

Headcount employment. Employment in the data is measured in workers, not in units of the labor composite. Optimal headcount production employment is[29]

$$L^i(z) = L^i_h(z) + L^i_l(z) = \left[\left(\frac{\bar{w}^i \mu^i}{w_h}\right)^\rho + \left(\frac{\bar{w}^i(1-\mu^i)}{w_l}\right)^\rho\right]\left(\frac{z\gamma^i}{\bar{w}^i}\right)^{\frac{1}{1-\gamma^i}} \equiv \phi^i \bar{L}^i(z), \quad (6)$$

where the factor ϕ^i converts composite units of labor $\bar{L}^i(z)$ into headcount employment $L^i(z)$. It is common for firms with the same technology. It increases in the skill premium because more workers are required per unit of the composite as firms substitute towards low-skill workers.

Combining these results yields (derivation in Appendix E.1):

Lemma 5. *At a given productivity z, optimal headcount production employment of a firm using the modern technology relative to one using the traditional technology declines in the skill premium if $\rho < 1/(1-\gamma^m)$.*

This result extends Lemma 4 from composite employment to headcounts. Compared to Lemma 4, the additional restriction on parameters is required due to a countervailing substitution effect: Since modern firms face a larger increase in $\bar{w}^i$, ϕ^m/ϕ^t rises with the skill premium, and modern firms' relative headcount falls by less than their relative composite input. However, this effect overturns the decline in modern firms' relative size with the skill premium only if ρ exceeds a threshold $\hat{\rho}$. In the calibration below, $\hat{\rho}$ exceeds 10 in all country groups. In general, $\hat{\rho} > 1/(1-\gamma^m)$, which equals 6.7 in our calibration. These numbers far exceed typical estimates of ρ. Hence, relative employment of modern firms falls with the skill premium for all plausible parameter values.

3.4. Entry, exit and technology choice

There is an unlimited mass of potential entrants. To start a firm, an entrant hires c^e workers in the population skill mix — $c^e L_l$ low-skill and $c^e L_h$ high-skill workers — at total cost $c^e \hat{w}$, where $\hat{w} \equiv L_h w_h + L_l w_l$ is the average wage in the economy.[30] These workers perform set-up tasks and produce no output. After paying the cost, the entrant draws a productivity z from a Pareto distribution with tail parameter α, lower bound $z_{\min}$, and thus *cdf* $G(z) = 1 - (z_{\min}/z)^\alpha$. Because employment rises with productivity with elasticity $1/\left(1-\gamma^i\right)$ by (4), we impose $\alpha > 1/(1-\gamma^m)$ so that mean firm size is finite.

[29] This follows from combining the cost-minimization FOC $w_h = \bar{w}^i \partial \bar{L}^i / \partial L^i_h$ with the definition of $\bar{L}$.

[30] This assumption follows Klenow and Li [2025]. It implies that firm sizes do not vary with population size, as a greater population is absorbed by a proportional increase in the number of each type of firm.

Similarly, each active firm must hire c^f overhead workers in the population skill mix. This implies a fixed operating cost of $c^f \hat{w}$. When mapping the model to the data, we only count workers in actively operating firms, so that overhead workers count towards measured firm employment, while entry workers do not. This choice simplifies the algebra and does not affect results.

Technology choice. At the optimal input choice, profits are

$$\pi^i(z) = z^{\frac{1}{1-\gamma^i}} \left(\bar{w}^i\right)^{-\frac{\gamma^i}{1-\gamma^i}} \tilde{\gamma}^i - c^f \hat{w}, \tag{7}$$

where $\tilde{\gamma}^i \equiv \left(\gamma^i\right)^{\frac{\gamma^i}{1-\gamma^i}} - \left(\gamma^i\right)^{\frac{1}{1-\gamma^i}}$. Because $\gamma^m > \gamma^t$, $\pi^m(z)$ is less than $\pi^t(z)$ for low z, and is greater for high z. Hence, low-productivity firms prefer the traditional technology and high-productivity firms the modern technology. Denote by z^m the cutoff where $\pi^t(z) = \pi^m(z)$. Equating the two profit functions yields

$$z^m = \left(\frac{\bar{w}^{m\,\frac{\gamma^m}{1-\gamma^m}}}{\bar{w}^{t\,\frac{\gamma^t}{1-\gamma^t}}} \frac{\tilde{\gamma}^t}{\tilde{\gamma}^m} \right)^{1/\left(\frac{1}{1-\gamma^m} - \frac{1}{1-\gamma^t}\right)}. \tag{8}$$

The exit threshold. Firms only choose to operate if they can earn non-negative profits. This is the case if productivity is at least z^x, where $\pi^t(z^x) = 0$. This implies

$$z^x = \left(\frac{c^f}{\tilde{\gamma}^t} \hat{w}\right)^{1-\gamma^t} \left(\bar{w}^t\right)^{\gamma^t} = \left(\frac{c^f}{\tilde{\gamma}^t}\right)^{1-\gamma^t} \left(\frac{\bar{w}^t}{\hat{w}}\right)^{\gamma^t} \hat{w}. \tag{9}$$

Throughout, we consider the case $z^x < z^m$, in which both technologies are in use. Firms then choose the modern technology if $z \geq z^m$, the traditional one if $z \in [z^x, z^m)$, and do not operate if $z < z^x$, as illustrated in Figure V. This is the case we encounter in the calibration for all countries.

The following results follow from the expressions for z^m and z^x:

Lemma 6. *An increase in the skill premium increases the cutoff z^m.*

Lemma 7. *An increase in the skill premium that does not affect the mean wage reduces z^x.*

These results reflect that modern firms are more skill-intensive than traditional ones, so that a higher skill premium raises $\bar{w}^m / \bar{w}^t$ and with it z^m. Traditional firms, in turn, are less skill-intensive than the economy average, so $\bar{w}^t / \hat{w}$ and with it z^x fall.

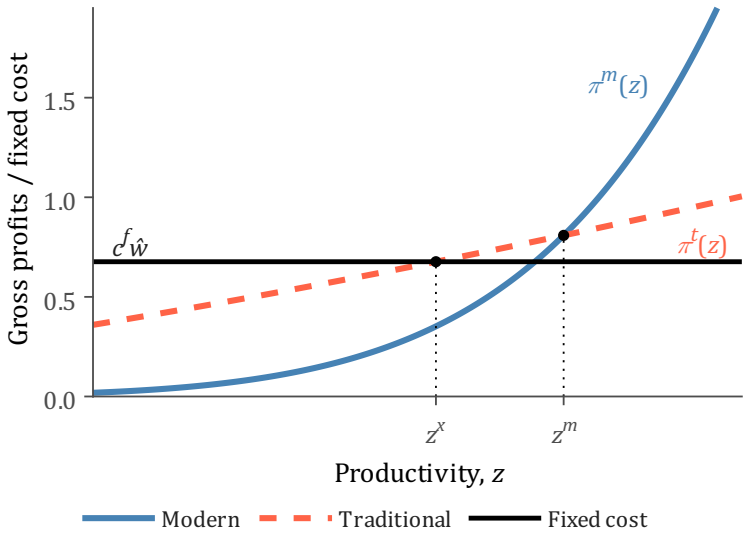


FIGURE V
Profits, exit, and technology choice.

Notes. The figure plots firms' profits under each of the two technologies against their productivity z. A firm operates only if profits cover the fixed cost, which defines the exit threshold z^x. Among operating firms, it chooses the modern technology if $\pi^m(z) \geq \pi^t(z)$, which defines the technology threshold z^m. Profits are shown gross of the fixed cost drawn at the US calibration of Table IV.

3.5. The share of modern firms and of modern employment

The share of modern firms. Entrants with $z \geq z^m$ choose the modern technology, and those with $z < z^x$ do not operate. The share of modern firms among operating firms is therefore

$$m^m \equiv \frac{M^m}{M} = \frac{1 - G(z^m)}{1 - G(z^x)} = (z^x / z^m)^\alpha,$$

where M is the total mass of firms and M^m that of modern ones. Because z^m and z^x are both homogeneous of degree one in wages, their ratio depends on wages only through the skill premium. Combining (8) and (9) yields

$$\frac{z^m}{z^x} = \left(\frac{\bar{w}^m}{\bar{w}^t}\right)^{\gamma^t \frac{1-\gamma^m}{\gamma^m - \gamma^t}} \frac{\bar{w}^m}{\hat{w}} \left(\frac{\hat{w}}{\bar{w}^t}\right)^{\gamma^t} \left(\frac{\tilde{\gamma}^t}{\tilde{\gamma}^m}\right)^{1/\left(\frac{1}{1-\gamma^m} - \frac{1}{1-\gamma^t}\right)} \left(\frac{c^f}{\tilde{\gamma}^t}\right)^{\gamma^t - 1}. \tag{10}$$

The last two factors depend only on parameters. Each of the three wage ratios increases in the skill premium: the first by Lemma 1, and the latter two because modern firms are more, and traditional firms less, skill-intensive than the economy-wide average. Hence:

Lemma 8. *The share of modern firms declines in the skill premium.*

A higher skill premium makes the modern technology relatively less attractive, raising z^m (Lemma 6). At the same time, it lowers the cost of traditional firms relative to the average wage, pushing the exit threshold z^x down (Lemma 7). The range of

productivities for which the traditional technology is chosen therefore expands at both ends, while that of the modern technology shrinks.

The share of employment in modern firms. We next turn from firms' choices to the aggregate share of employment, first in modern and then in large firms.

Proposition 1. *Suppose $\rho < 1/(1-\gamma^m)$. Then the share of production employment in modern firms, ℓ^m, decreases in the skill premium with elasticity $-(1-\ell^m)\,\eta_L(\omega)$, where η_L is a positive elasticity. The share of total employment in modern firms decreases as well.*

See Appendix E.2 for the proof. Modern firms' employment share falls because both their share (Lemma 8) and their relative size (Lemma 5) decline in the skill premium. The elasticity of the share of employment in modern firms with respect to the skill premium in turn depends on its level. The lower ℓ^m, the more the modern employment share changes with the skill premium. Conversely, if traditional firms account for a small share of employment, ℓ^m is close to one and barely responds to relative wages. This is typically the case in high-income countries.

3.6. The share of employment in large firms

In the data, we do not observe firms' technology choice. We only observe whether a worker is employed in a small or a large firm. We therefore analyze the model's implications for the share of employment in large firms, the object we measure in Section 2. To do so, we set the headcount employment threshold b (for "big") to 10, as in our data, and define z^b as the productivity level at which a firm's total headcount employment equals b. We refer to large firms with a superscript b, and to small ones with a superscript s.

Since firms choose both size and technology, the two thresholds z^m (for technology) and z^b (for large vs small size) do not coincide. In our quantitative analysis, z^b exceeds z^m in all countries. As a consequence, traditional firms ($z \in [z^x, z^m)$) are all small, the most productive modern firms ($z \geq z^b$) are large, and less productive modern firms are small ($z \in [z^m, z^b)$). We focus on this case throughout.

The share of production employment in large firms is then the product of (1) the share of production employment in modern firms ℓ^m, and (2) the share of large firms' employment within modern-firm employment. A higher skill premium moves the two factors in opposite directions. The first falls, by Proposition 1. The second rises, since a higher skill premium raises the technology threshold z^m by more than the size threshold z^b, so the range of small modern firms narrows and employment within the modern sector concentrates in large firms. The next proposition gives conditions under which the first force dominates.

Proposition 2. *The share of production employment in large firms, ℓ^b, decreases in the skill premium if $\rho < 1/(1-\gamma^m)$ and if the share of production employment in modern firms ℓ^m does not exceed a threshold $\hat{\ell}^m$. Under these conditions, the share of total employment in large firms, ℓ^b_{tot}, is also guaranteed to decrease in the skill premium if $\gamma^t > \hat{\gamma}^t$, where $\hat{\gamma}^t$ depends on data.*

See Appendix E.3 for the proof. The condition on ρ is the same as in Proposition 1. The threshold $\hat{\ell}^m$ exists because when almost all employment occurs in modern firms, its elasticity with respect to the skill premium ω is low (by Proposition 1), and the within-modern shift in employment towards large firms could in principle dominate. In practice, this is not the case. For the US calibration, $\hat{\ell}^m = 0.987$, and $\hat{\ell}^m$ exceeds 0.95 in all country income groups. It is so close to one because modern firms are much larger than traditional ones, so employment shifts between the modern and traditional technologies dominate employment shifts between small and large modern firms unless almost all employment is modern already.[31]

The empirical share of employment in large firms counts all workers, including overhead workers. Its model counterpart is therefore the share of total employment in large firms, ℓ^b_{tot}, the object in the second part of Proposition 2 and the target of our calibration below. It inherits the decline in large firms' share of production employment, but also depends on their share of overhead workers, which decreases in the skill premium unless γ^t is very small. The bound $\hat{\gamma}^t$ can be computed for each country. In our calibration it lies below 0.2 for all country income groups, so that the condition $\gamma^t > \hat{\gamma}^t$ holds for any plausible γ^t.[32]

In sum, a higher skill premium implies a smaller share of employment in large firms for all plausible parameter values. The two reasons are intuitive. A greater cost of skill raises the relative cost of the modern technology, which both shrinks modern firms relative to traditional ones and reduces the share of firms that adopt it. The countervailing shift of modern employment towards large firms is dominated under very weak conditions on parameters.

[31] The condition $\ell^m < \hat{\ell}^m$ therefore holds in all country income groups and in 48 of the 56 country calibrations, with a particularly large margin in low- and middle-income countries, where ℓ^m is low. Once ℓ^m exceeds 0.95, as in the US or other high-income countries, the effect of the skill premium on large-firm employment becomes small. The condition fails only in five high-income countries in which almost all employment is already modern ($\ell^m \geq 0.93$; Austria, the Czech Republic, France, Germany and Switzerland), and in the three countries in which small firms are more skill-intensive than large ones (see footnote 35). In these countries the effect of the skill premium on large-firm employment is positive but negligible.

[32] Both cutoffs in Proposition 2 are sufficient conditions, not sharp ones. The exact requirement on ρ is $\rho < \hat{\rho} = \left[(\Lambda_h^m - \Lambda_h) + \gamma^t(\Lambda_h - \Lambda_h^t)\right] / \left[(1-\gamma^m)(\Lambda_h^m - h_{\text{tot}}^m)\right]$, which ranges from 11.0 to 52.3 across country groups, well above the transparent cutoff $1/(1-\gamma^m) = 6.7$. The exact requirement on γ^t, $\gamma^t > \hat{\gamma}^t = (\Lambda_h - h_{\text{tot}}^m)/(\Lambda_h - \Lambda_h^t)$, never exceeds 0.17.

3.7. Equilibrium

An equilibrium of the model consists of values of w_h, w_l, M, z^m and z^x such that five equilibrium conditions hold: *(i)* optimal technology choice defined in equation (8), *(ii)* optimal exit defined in equation (9), *(iii-iv)* clearing of the labor market for low- and high-skill workers, and *(v)* free entry. See Appendix E.5 for details.

Free entry. Free entry drives entrants' expected net payoff to zero. The resulting free entry condition is

$$c^e \hat{w} = \tilde{\gamma}^t \, \bar{w}^{t\,-\frac{\gamma^t}{1-\gamma^t}} Z^t + \tilde{\gamma}^m \, \bar{w}^{m\,-\frac{\gamma^m}{1-\gamma^m}} Z^m - c^f \hat{w} \left(1 - G(z^x)\right), \tag{11}$$

where $Z^m \equiv \tilde{\alpha}^m z^{m\,\frac{1}{1-\gamma^m}-\alpha}$, $Z^t \equiv \tilde{\alpha}^t \left(z^{x\,\frac{1}{1-\gamma^t}-\alpha} - z^{m\,\frac{1}{1-\gamma^t}-\alpha} \right)$ and $\tilde{\alpha}^i \equiv \alpha z_{\min}^{\alpha} / \left(\alpha - \frac{1}{1-\gamma^i}\right)$. The first two terms on the right-hand side of (11) are expected variable profit, and the last is the expected fixed-cost outlay. Given ω, this is a non-linear equation in w_l and thus pins down the low-skill wage.

Market clearing. The clearing conditions for the two labor markets then determine the number of firms M and the skill premium ω. See Appendix E.4 for details.

The skill premium is the price through which the skill endowment shapes the firm size distribution. A scarcer supply of skills raises ω through the market clearing conditions, which by Propositions 1 and 2 reduces the share of employment in modern and in large firms. The model thus jointly generates a higher skill premium (Fact 2) and a smaller large-firm employment share (Fact 1) in countries where skills are scarce, as we see in the data. The next section quantifies this mechanism.

4. Calibration

We calibrate the five production parameters (μ^i, γ^i, ρ), the entry and fixed costs (c^e, c^f), and the productivity distribution (α and $z_{\min}$), and take aggregate skill endowments (L_h, L_l) directly from the data.

4.1. Calibration strategy

We calibrate the model first to the United States ("US calibration"), and then separately to all countries in our data set ("country-specific calibration"). In each case, we set some parameters to values known from the literature or to direct data counterparts, and calibrate the remaining ones internally to fit a set of data moments.

Externally set parameters. We set the scale parameter of the modern technology and the elasticity of substitution, γ^m and ρ, to standard values in the literature, 0.85 [Atkeson and Kehoe, 2007] and 4, respectively. We normalize $z_{\min}$ to 1 and the entry cost to 0.1.[33] Equation (6) implies that the tail index of the firm size distribution is equal to $\alpha(1-\gamma^m)$. The value of γ^m and the observed tail index directly determine α. These parameter values are common for all countries.

Our choice of the elasticity of substitution, ρ, is based on recent estimates of the long-run elasticity of substitution between high- and low-skill workers [Hendricks and Schoellman, 2023, Bils et al., 2024]. This literature shows that the long-run elasticity is substantially higher than traditional estimates of the short-run elasticity from annual data, such as those of Katz and Murphy [1992]. We show in Section 5.3 that the choice of ρ is not central to our quantitative results. This is consistent with the fact that the model results ultimately reflect the gap in skill intensity between the two technologies, $\mu^m - \mu^t$.

Internally calibrated parameters. In the US calibration, we set the remaining four parameters μ^m, μ^t, γ^t and c^f to match four moments: (i) the skill intensity of large firms h^b_{tot}, (ii) the share of employment in large firms ℓ^b_{tot}, (iii) the skill premium ω, and (iv) the share of large firms m^b.[34]

Consider first the skill weights μ^i. The firm's first-order condition (3) combined with the observed skill premium, allow us to map the skill intensities of the two types of firms into the skill weights μ^i for both technologies. The difference in returns to scale between the two technologies then governs the share of employment in large firms. As a result, the employment share and skill intensity of large firms plus the skill premium pin down the skill weights μ^m and μ^t and returns to scale of the traditional technology γ^t. Finally, the fixed cost c^f pins down the exit threshold and thereby the share of firms that are large, such that the observed m^b determines c^f.

In the cross-country calibration, we observe h^b_{tot}, ℓ^b_{tot} and ω for all countries, but not the share of large firms m^b. We are not aware of a cross-country dataset that provides this information. We thus set γ^t to its US value for all countries, and set the other three parameters to match each country's three target moments.

Mapping the model to the data. Two features of the data determine how we map the model to it. First, the data classify firms by size, whereas the model distinguishes them

[33] Recall that the entry cost is in units of labor and therefore must be below the total labor supply of 1. Its normalization only affects the wage level.

[34] It would seem natural to also target the skill intensity of small firms h^s_{tot}. However, it does not provide independent information, since it is already implied by the employment share of large firms ℓ^b_{tot} and their skill intensity h^b_{tot}, as the accounting identity $L_h = \ell^b_{\text{tot}} h^b_{\text{tot}} + (1-\ell^b_{\text{tot}}) h^s_{\text{tot}}$ implies $h^s_{\text{tot}} = (L_h - \ell^b_{\text{tot}} h^b_{\text{tot}})/(1-\ell^b_{\text{tot}})$.

by technology. As discussed above, optimal employment at the productivity threshold for the modern technology, z^m, need not equal the ten-employee threshold that defines large firms in the data. Indeed, optimal employment at z^m is below ten in all our calibrations, so the threshold ordering of Section 3.6 applies: firms with $z \in [z^x, z^m)$ are traditional and small, firms with $z \in [z^m, z^b)$ are modern and small, and firms with $z \geq z^b$ are modern and large. Data moments on large firms therefore describe modern firms, while moments on small firms reflect a mix of modern and traditional firms. Second, data moments on employment include overhead workers. The observed skill intensity h^i_{tot} therefore differs from production skill intensity h^i defined after equation (3), so that the equation cannot be used directly to back out μ^i analytically.

4.2. The US calibration

In our data, the US skill premium is 1.79, and large firms account for 88.3% of employment. High-skill workers make up 96.5% of employment in large firms, 90.7% in small firms, and 95.9% of wage employment overall (see Table I). Data from the US Census Business Dynamics Statistics (BDS) show that 21.4% of firms have at least ten employees, and imply a tail index of 1.13 for firm employment.

The model matches all these moments exactly. Table IV reports the implied parameter values. We find that both technologies place a higher weight on high-skill workers ($\mu^i > 1/2$), which is what generates a skill premium. The weight is 63% in the traditional technology and 80% in the modern one, reflecting the greater skill intensity of large firms. The gap in weights may seem moderate, but the relative weight $\mu^i/(1-\mu^i)$, which governs a technology's relative demand for high-skill labor, differs far more: it is 1.7 for the traditional and 4.1 for the modern technology.

Almost all workers in the US have at least secondary education (our definition of high-skill), and almost all work in large firms. Matching this requires a large gap between the scale parameters of the two technologies, such that the calibrated γ^t is 0.301, far below $\gamma^m = 0.85$. Finally, the calibrated fixed cost of 0.748units of labor implies that each firm employs 0.748 overhead workers. With average employment of almost twenty per firm, overhead employment accounts for only about 3.3% of total employment.

We now turn to the model's implications for untargeted moments. The smallest operating firm employs 1.2 workers and the smallest modern firm 5.70 workers, both including overhead. The share of modern firms, at 43.3%, is therefore much larger than the share of large firms of 21.4%. Traditional firms make up the majority of firms, but because they are so small, they account for only 1.3% of employment.

TABLE IV
Calibration to the US economy

Parameter / moment	Value	Source / target
Pre-set parameters		
ρ	4	Hendricks and Schoellman (2023); Bils et al. (2024)
γ^m	0.85	Atkeson and Kehoe (2007)
$z_{\min}$	1	normalization
Calibrated parameters		
α	7.533	Tail index of employment = 1.13 (BDS)
μ^m	0.805	Skill intensity of large firms, $h^b_{\text{tot}} = 0.97$
μ^t	0.632	Employment share of large firms, $\ell^b_{\text{tot}} = 0.88$
γ^t	0.301	Skill premium, $\omega = 1.79$
c^f	0.748	Count share of large firms, $m^b = 0.214$
c^e	0.100	normalization
Supply-side input		
L_h	0.959	Skilled share of the US labor force, L_h
Implied equilibrium objects		
m^m	0.433	Share of modern firms
ℓ^m	0.987	Employment share in modern firms
$L^t(z^x)$	1.17	Employment of smallest active firm (incl. overhead)
$L^m(z^m)$	5.70	Employment of smallest modern firm (incl. overhead)
$c^f M$	0.033	Fixed overhead share of employment
$1/M$	19.67	Average firm size (incl. overhead)

Notes. The table lists each model parameter with its value and the target or source used to set it.

4.3. Country-specific calibrations

We next calibrate the model separately to each country in our dataset. We set the skill endowment L_h to the country's observed value and target the employment share of large firms ℓ^b_{tot}, their skill intensity h^b_{tot}, and the skill premium ω. Since the share of firms that are large is not observed outside the US, we keep γ^t at its US value and choose μ^m, μ^t, and c^f to match these three targets. All other parameters take the values in Table IV. To summarize the results, we also calibrate the model to each of the three country income groups.

Table V reports the targets and calibrated parameters for the US and the three country income groups, and Table F.I in the appendix reports them for each country.[35]

[35]The calibration matches all countries' targets exactly except for Rwanda, where the share of high-skill workers among small-firm employees is only 40% of the population share, by far the lowest in the sample. Because overhead workers are hired at the population skill mix, the model cannot generate such strong sorting even when the traditional technology uses no high-skill labor; it thus misses Rwanda's targets by about 5% on average. In three countries (the Netherlands, Slovenia and Zimbabwe), small firms are slightly more skill-intensive than large firms in the data, which requires a higher skill weight in the traditional than in the modern technology, $\mu^t > \mu^m$. The results below are not sensitive to these countries.

TABLE V
Calibration to country income groups

	Data				Calibrated parameters			Firms			Firm shares		Optimal employment	
	L_h	ω	ℓ^b_{tot}	h^b_{tot}	μ^m	μ^t	c^f	nr firms	avg size	$c^f M$	large	modern	$L^t(z^x)$	$L^m(z^m)$
LIC	0.513	1.84	0.478	0.646	0.682	0.560	0.650	0.42	2.08	0.272	0.012	0.015	1.00	8.20
MIC	0.702	1.69	0.661	0.780	0.700	0.565	0.669	0.25	3.43	0.169	0.028	0.040	1.03	7.40
HIC	0.893	1.39	0.810	0.901	0.708	0.645	0.710	0.13	6.80	0.091	0.068	0.108	1.03	6.83
US	0.959	1.79	0.883	0.965	0.805	0.632	0.748	0.04	19.67	0.033	0.214	0.433	1.17	5.70

Notes. The first two groups of columns of the table report the calibration targets and the internally calibrated parameters for each country income group, with the United States shown separately for reference. Countries are grouped by GDP per capita (PPP) as in Table I. Each target is the average over all countries in the group with data on that moment. Quantities reported to the right of the calibrated parameters are model implications and are not targeted. $L^t(z^x)$ and $L^m(z^m)$ denote the optimal levels of employment of the smallest traditional- and modern-technology firm, respectively.

Three patterns stand out. The skill weight of the traditional technology, μ^t, is slightly higher in HICs than in LICs and MICs. In contrast, the skill weight of the modern technology, μ^m, is similar across the three income groups, at about 0.70, but substantially higher in the US, at 0.80. The fixed cost c^f rises with income, from 0.65 in LICs to 0.75 in the US.

These patterns follow directly from the data. The much lower skill endowment of LICs implies a lower employment share in large firms, and a greater skill premium. As this implied skill premium exceeds observed levels, matching ω requires lower μ^i. (The US stands out among HICs because its high skill premium combined with a large skill endowment requires a high μ^m.) Any remaining gaps in ℓ^b_{tot} are closed by the fixed cost c^f, since lower fixed costs allow smaller, less productive firms to operate, which reduces the employment share of large firms. Overall, the systematic variation in ℓ^b_{tot} across country income groups comes from skill endowments. The fixed cost varies much more within income groups than across them, and effectively acts as a country-specific residual.

The calibrated differences in μ^i allow for two interpretations. First, they may reflect differences in the relative efficiency of high-skill labor stemming from the skill bias of technology or the quality of schooling. Second, they are equivalent to a setting in which μ^m and μ^t take their US values in all countries, but firms face revenue-neutral distortions to their skill mix.[36] Under this second interpretation, lower μ^i in LICs corresponds to a distortion that raises the cost of high-skill relative to low-skill labor, compared to the US, by 93% in modern and 35% in traditional firms.

The calibration also has implications for absolute firm size, which we did not target.

[36]The equivalence requires a tax on high-skill labor that is exactly offset by a subsidy to low-skill labor, so that the distortion raises no revenue. It also requires a small adjustment to the productivity of each technology, because a change in μ^i also shifts productivity. In LICs, this adjustment amounts to about -5% for the modern and 3.3% for the traditional technology.

In LICs, the higher share of employment in small firms comes with a much lower average firm size of two to three workers. Fewer than 2% of firms are large or modern, and the smallest active firm has a single worker, including overhead. These model-implied figures align closely with Bento and Restuccia [2021], who find that average employment of non-agricultural firms is close to three in LICs.

5. The determinants of large-firm employment across countries

With the US and the country-specific calibrations in hand, we now quantify how much skill endowments contribute to cross-country differences in the large-firm employment. The first exercise takes the US calibration and varies only the skill endowment. This alone closely reproduces the cross-country patterns in Facts 1 and 4, which the calibration did not target. The second exercise uses the country-specific calibrations to decompose the cross-country variance of large-firm employment into the contributions of skill endowments, skill weights, and fixed costs.[37]

5.1. The role of skill endowments alone

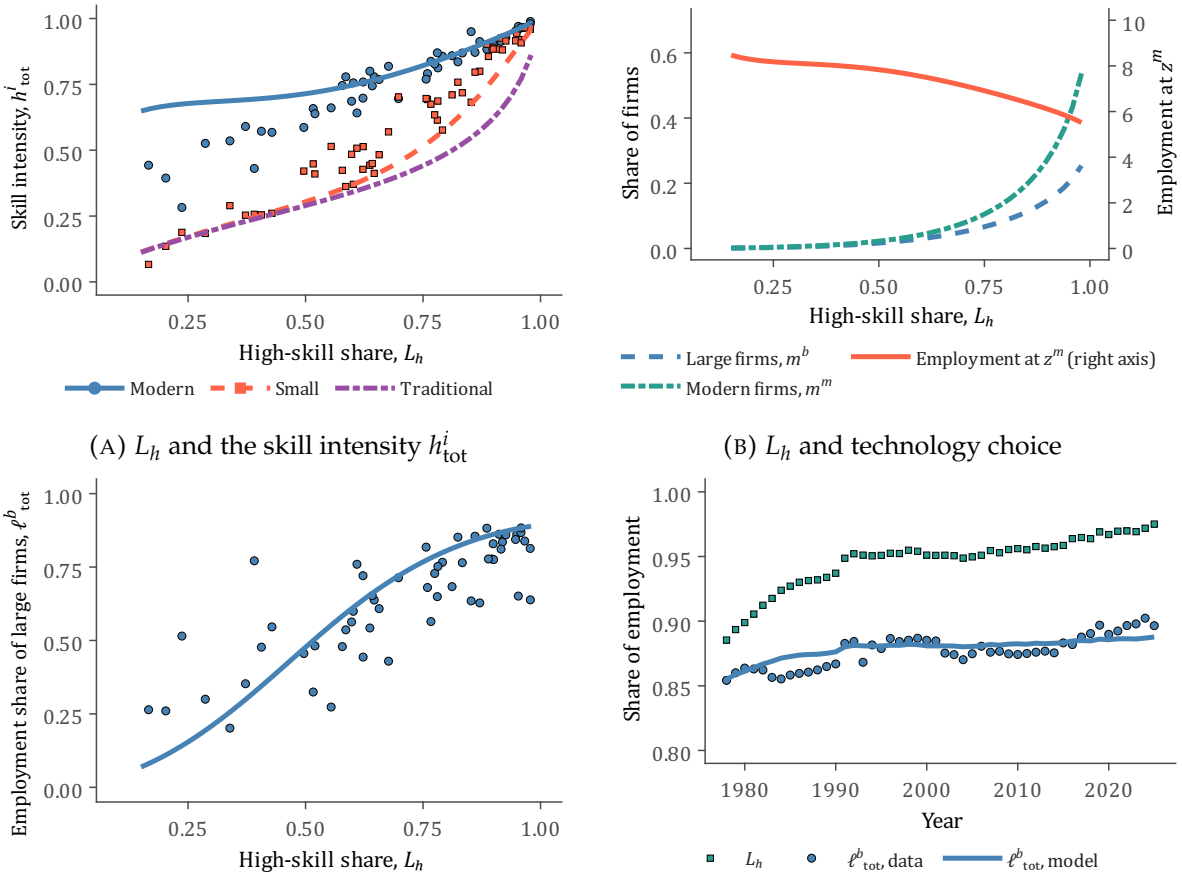
We vary L_h in the US calibration while holding all other parameters at their calibrated values, and compute the equilibrium for each value of L_h. We then trace how firms' input and technology choices and the structure of employment respond.

We begin with the effect of skill endowment L_h on technology choice, the key mechanism. In Figure VIb the red line shows optimal employment at the technology choice threshold z^m (right axis), which is easier to interpret than z^m itself. In the US calibration, the smallest modern firm employs just under six workers. In a typical LIC, where L_h is only about half its US value, the modern technology is much less attractive. The threshold z^m rises, and optimal employment at the threshold increases to about eight workers. As a result, the share of firms using the modern technology falls sharply, from just over 40% to less than 5% (green dash-dot line).[38] The share of large firms also falls sharply, from just over 20% to around 2% (dashed blue line).

These shifts in technology choice translate into large changes in the structure of employment. Figure VIc plots our main object of interest, the share of employment in large firms, against L_h, in the model and in the cross-country data. The model matches the US data point in the upper right by construction. As L_h falls, the model predicts a steep decline in ℓ^b that closely tracks the cross-country data. Even though all other

[37] Throughout, for readability, we write ℓ^b for ℓ^b_{tot}, the share of total employment in large firms.

[38] The share declines so strongly for a change in z^m because of the Pareto shape of the productivity distribution.



(A) L_h and the skill intensity h^i_{tot}

(B) L_h and technology choice

(C) L_h and ℓ^b_{tot}

(D) L_h and ℓ^b_{tot}: model and US data, 1978–2025

FIGURE VI
Varying the skill endowment in the US calibration.

Notes. The figure varies the skill endowment L_h in the US calibration of Table IV, recomputing the equilibrium at each level and holding every other model input at its US value. Lines show model outcomes and markers the data, one marker per country. Panel (a) shows the skill intensity of modern (equivalently, large) firms, of traditional firms and of small firms; traditional firms have no data counterpart, since technology is not observed. Panel (b) shows technology choice: employment of the smallest firm using the modern technology, i.e. at the threshold z^m, together with the share of firms that are modern and the share that are large. Panel (c) shows the share of employment in large firms, ℓ^b, against L_h. The US data point at the top right is matched by construction. Data for other countries is not targeted. Panel (d) shows the data and model-implied time series of ℓ^b in the U.S. when L_h is set to its observed value in each year. Large firms have at least ten employees and high-skill workers at least ten years of schooling.

parameters stay at their US values, the model thus replicates Fact 1, the strong positive relationship between the skill endowment and the structure of employment not only qualitatively but also quantitatively.

Figure VIa turns to firms' the skill intensity. It plots the skill intensity of large firms (which are all modern, see solid blue line and circles), traditional firms (purple dash-dotted line), and small firms (dashed red line and squares) against L_h. As L_h decreases, the average skill intensity of firms must decrease as well. The skill intensity of modern firms, however, declines much less than that of traditional firms, reflecting

the greater skill intensity of modern firms (Lemma 3). Since all large firms are modern, this appears in both the data and the model as a much smaller decline for large firms than for small firms. Small firms are a combination of modern and traditional firms, so their skill intensity declines more than that of modern firms, but less than that of traditional ones. At very low levels of L_h, almost all small firms use the traditional technology and their skill intensity converges to that of traditional firms.

Changes in the skill endowment alone thus nearly reproduce Fact 4, the larger gap in skill intensity between large and small firms in low-income countries. Because we use the US calibration, this pattern is not targeted. This close fit lends credibility to the model as a framework for studying firms' skill intensity across countries.[39]

Skill endowments also account well for variation in large-firm employment over time. We set L_h to its observed value in the US in each year from 1978-2025. Figure VId compares the resulting model-implied ℓ^b (line) with the data (markers).[40] Although the model is calibrated to a single year, it closely tracks the time-series of ℓ^b.[41] With the observed variation in L_h as its only input, the model thus captures both cross-country and time-series variation in large-firm employment.

The same holds within other countries. Figure VII repeats the exercise for each country in the balanced panel, this time starting at that country's own calibration, and varying only L_h. For 12 of the 16 countries we observe from 2007 to 2017, the resulting model line closely follows the observed trajectory.

With observed variation in L_h as its only input, the model thus captures both the cross-country and the time-series variation in large-firm employment. The next subsection lets all calibrated parameters vary across countries and asks how much of this variation skill endowments account for.

5.2. Skill endowments versus other factors

Skill endowments are not the only determinant of the employment structure that varies across countries. In the country-specific calibrations of Section 4.3, countries also differ in the skill weights μ^m and μ^t and the fixed cost c^f. Since these calibrations match each country's employment share of large firms, they let us measure how much each of the three factors contributes to its variation across countries. In the data, the variance of $\ln \ell^b$ across countries is 0.1. We ask what share of this variance each factor accounts for.

[39] Section 5.3 shows that this fit is much worse for a lower value of ρ.

[40] The high-skill share is from the CPS. The employment share of large firms is from the BDS up to 1990 and from the CPS from 1991 onward.

[41] Figure F.I shows the result when we also vary μ^m and μ^t by a common increment each year to match the skill premium in each year. The model then matches the time series of ℓ^b similarly well. The implied skill weights rise strongly from the 1970s to the early 1990s and slowly afterwards, in line with the growth of the skill premium.

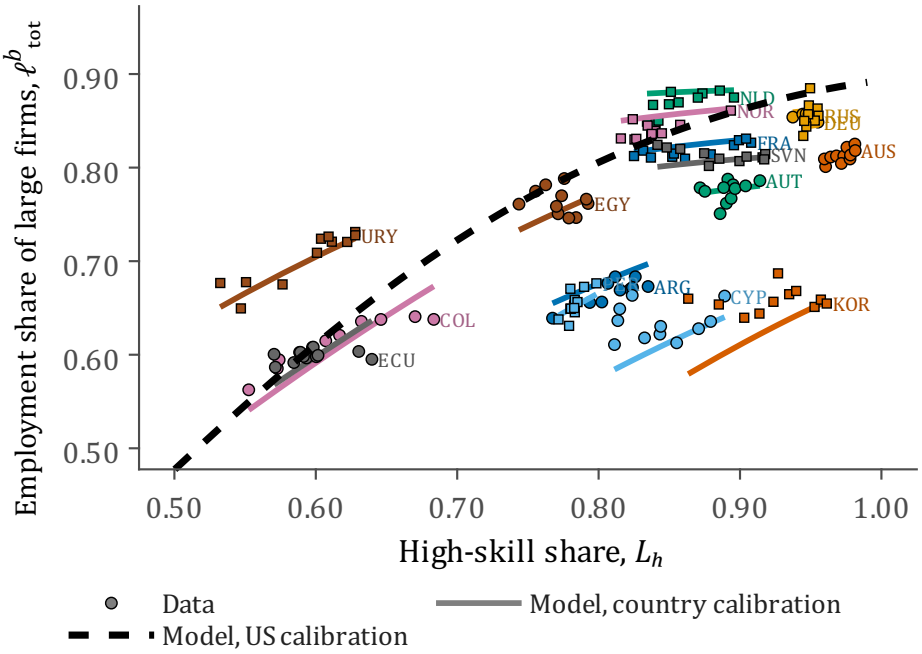


FIGURE VII
Skill endowments and large-firm employment within countries, 2007–2017

Notes. Markers show the share of employment in large firms ℓ^b_{tot} against the high-skill share L_h for each country in the balanced panel of Table II. Solid lines show ℓ^b_{tot} in the model when L_h takes each year's value and all other parameters stay at the country's own calibration. The dashed line shows the US calibration of Table IV, varying only L_h.

We answer this question by adding or removing variation in each factor and measuring the effect on the model-implied variance of $\ln \ell^b$. Because the model is non-linear, each factor's contribution depends on the values of the other factors. Table VI therefore reports three decompositions. The "alone" column starts with all factors at their cross-country averages and sets one factor to its country-specific values. The "drop" column starts from the country-specific calibrations and sets one factor to its cross-country average. The "Shapley" column averages each factor's marginal contribution over all orders in which the factors can be introduced.[42]

Averaging over all orderings ("Shapley"), skill endowments account for 88% of the cross-country variation in large-firm employment. The skill weight of the traditional technology accounts for 12% and fixed costs for 7%, while the contribution of μ^m is

[42] Formally, the Shapley value of input k out of the set of inputs N for a given outcome of interest $Q(\cdot)$ is

$$s_k = \sum_{S \subseteq N \setminus \{k\}} \underbrace{\frac{|S|!\,(|N| - |S| - 1)!}{|N|!}}_{\text{Shapley weight } w(|S|)} \underbrace{[\,Q(S \cup \{k\}) - Q(S)\,]}_{\text{marginal contribution of } k}. \tag{12}$$

Intuitively, s_k is the average marginal change in Q from adding input k when there already is a random subset S of the other inputs; the combinatorial weight $w(|S|)$ is the probability that exactly the blocks in S precede k in a uniformly random permutation of N. By construction, the contributions of all inputs sum to the total, unlike the "drop" or "alone" measures.

TABLE VI
Determinants of the variation in employment in large firms across countries

	$\log \ell^b_{\text{tot}}$			$\log \omega$		
Parameter	alone	drop	Shapley	alone	drop	Shapley
L_h	101.6%	63.2%	87.7%	103.8%	-14.0%	44.6%
μ^m	1.1%	-14.6%	-6.8%	74.7%	13.6%	42.8%
μ^t	3.3%	17.5%	12.4%	10.2%	1.7%	6.2%
c^f	36.1%	-28.9%	6.7%	4.1%	12.1%	6.3%

Notes. The table decomposes the cross-country variance of $\ln \ell^b$, the log share of employment in large firms, into the contributions of the three model inputs that vary across countries: the skill endowment L_h, the skill weights of the two technologies μ^i, and the fixed operating cost c^f. Because the model is non-linear, three decompositions are reported. Column *alone* sets every input to its cross-country average and then restores one input to each country's value. Column *drop* does the reverse: it starts from the country-specific calibrations and sets one input to its cross-country average. Column *Shapley* reports each input's average marginal contribution over all orders in which the inputs can be introduced. The first two columns need not sum to the total variance; the Shapley column does so by construction.

slightly negative. The other two decompositions confirm that skill endowments are by far the most important factor. Their contribution is smallest in the "drop" column, where the other factors take their country-specific values, including the lower μ^t of low-income countries. Even there, it exceeds half of the total. In the "alone" column, it exceeds 100%, in line with the close fit in Figure VIc.

Perhaps surprisingly, variation in relative skill efficiency μ^m, or equivalently in distortions to the optimal skill mix, does not help explain variation in the share of large firms in employment. Its contribution is even slightly negative. The reason is that μ^m mainly affects the skill mix within modern firms, rather than whether firms adopt the modern technology. A lower μ^m makes the modern technology less productive, but also lowers the skill premium. In equilibrium, the two effects largely offset each other, leaving technology choices almost unchanged. The $\log \omega$ columns of Table VI confirm this. Variation in μ^m accounts for 43% of the cross-country variation in the skill premium, compared with 45% for skill endowments. The skill weights are thus an important determinant of the skill premium but not of employment in large firms, whereas skill endowments determine both.[43]

Whichever decomposition we use, skill endowments thus account for most of the cross-country variation in large-firm employment. The skill weights matter mainly for the skill premium, and fixed costs act largely as a country-specific residual.

[43]Similarly, in the US time series, letting μ^i vary year by year in the US calibration is essential for matching the skill premium, but hardly affects the model-implied series for the share of employment in large firms (Figure F.I and footnote 41).

5.3. Robustness

We now examine the robustness of our findings to a key model parameter, the elasticity of substitution between low- and high-skill workers, and to an alternative definition of skill.

Lower elasticity of substitution (ρ). Our main analysis uses an elasticity of substitution of 4. With a lower elasticity, modern firms, which are more skill-intensive, find it harder to substitute high-skill by low-skill labor when high-skill labor is scarce. Skill scarcity then raises their unit cost by more relative to that of traditional firms. Fewer firms adopt the modern technology and modern firms shrink relative to traditional ones. Skill endowments therefore have a stronger effect on the employment structure.

Table F.II and Table F.III in the appendix report the US and country-income group calibrations when ρ is set to 1.5. With this lower elasticity, matching the data requires a larger gap between μ^m and μ^t in the US. This makes firms' input choices and large-firm employment more responsive to the skill premium, as discussed in Section 3. Matching the data also requires larger differences in μ^i across country income groups. This reflects the familiar result that a lower elasticity of substitution implies larger cross-country differences in the relative efficiency of skill.

Figure F.II repeats the exercise of Section 5.1 with $\rho = 1.5$. As the argument above suggests, modern firms now remain highly skill-intensive even when skills are scarce. The model therefore predicts a skill intensity of large firms in skill-scarce countries that is far too high – in contrast to the close fit when $\rho = 4$. At the same time, skill endowments have a stronger effect on large-firm employment. With US parameters and L_h of 50%, the employment share of large firms is around 50% in the baseline but just under 30% with $\rho = 1.5$. In a decomposition paralleling the exercise in Section 5.2, the variation in L_h now accounts for more than the entire cross-country variation in $\ln \ell^b$ (see Table F.IV in the appendix). A lower ρ thus strengthens our main result, but yields a much worse fit to the skill intensity of large firms across countries. Our baseline value of $\rho = 4$ is therefore both conservative and better supported by the data.

An alternative definition of skill. In our benchmark analysis, we classify workers with at least ten years of schooling as high-skill. As an alternative, we raise the threshold to thirteen years. This lowers the share of high-skill workers everywhere (Table C.II). In the US, for example, it falls from 95.9% to 67.2%.

Tables F.V to F.VII in the appendix reproduce the main results under this definition. In the US calibration, both skill weights μ^m and μ^t are lower, reflecting the lower L_h. The scale parameter of the traditional technology, γ^t, is higher than in the benchmark but still below that of the modern technology. The share of firms using the modern

technology is lower, at 35.1%, but these firms still account for most employment, and average firm size remains in line with the data. Across income groups (Table F.VI), both μ^t and μ^m vary little. Fixed costs are again lower in low-income countries. There, firms again devote much of their labor to overhead and the smallest traditional firm employs about one worker.

Skill endowments remain the main determinant of large-firm employment (Table F.VII). They account for 74.7% of its cross-country variance, compared with 88% in the benchmark, while the skill weight of technology plays a secondary role. Skill endowments and skill weights also account for most of the variation in the skill premium, but in more equal proportions than in the baseline. Our conclusions therefore do not depend on the definition of skill.

6. Skill endowments, the firm size distribution, output and measured distortions

Skill endowments shape firms' technology choices and the firm size distribution. This section examines what this implies for aggregate output (Section 6.1) and for measured size-dependent distortions (Section 6.2).

6.1. Aggregate output

We first ask how the organization of production – firms' technology choices and the resulting firm size distribution – shapes the aggregate output gains from a larger skill endowment. Starting from the LIC calibration in Table V, we conduct counterfactual experiments in which we raise the skill endowment L_h, the skill weights μ^m and μ^t and the fixed cost c^f to their US value, both separately and in combination. Table VII reports the effects on output. In the "GE" columns, we recompute the full equilibrium. In the "PE" columns, the thresholds z^x and z^m and the mass of firms M stay at their LIC values, so that wages and firm sizes adjust but technology choices and entry do not. The difference between the two isolates the contribution of firms' technology choices to the output gains.

Raising the skill endowment from the LIC level ($L_h = 0.513$) to the US level ($L_h = 0.959$) increases output by 10.6% and the share of employment in large firms from 48% to 78%. With the firm structure held fixed, the output gain is only 8.2%. The adjustment of the firm structure thus raises the output gain from skill accumulation by almost a third, a factor of 1.296. The reason is that with a fixed firm structure, large firms absorb few of the additional high-skill workers, and their employment share rises only to 56% rather than 78%. Many of the additional high-skill workers remain in traditional firms,

where they are less productive. Once the firm structure adjusts, they move to modern firms, and the full output gain is realized. Changes in the firm structure are thus an important channel through which skill accumulation raises output.

Skill endowments and skill weights are strongly complementary. Raising the skill weights together with the skill endowment almost triples the output gain, to 29% (row "L_h, μ^m, μ^t to US"). Raising the skill weights alone, in contrast, hardly affects output and even lowers the employment share of large firms, consistent with Section 5.2. The return to skill accumulation is thus much larger when the skill weights rise at the same time, and higher skill weights raise output only when there are high-skill workers to use them.

TABLE VII
Skill endowments, technology choices and output

	GE		PE (fixed M & thresholds)	
	Y/Y_{LIC}	ℓ^b_{tot}	Y/Y_{LIC}	ℓ^b_{tot}
Baseline	1.000	0.478	1.000	0.478
L_h to US	1.106	0.777	1.082	0.563
μ^m to US	1.048	0.368	1.047	0.383
μ^m, μ^t to US	1.023	0.394	1.023	0.397
μ^m, μ^t, c^f to US	0.988	0.495	1.023	0.380
L_h, μ^m, μ^t to US (not c^f)	1.289	0.841	1.224	0.578
L_h and (μ^m, μ^t, c^f) to US	1.282	0.883	1.224	0.553

Notes. The table reports the effect on aggregate output of raising model inputs from the average low-income country to their US values, starting from the LIC calibration of Table V. Rows differ in what is raised: the skill endowment L_h, the skill weights μ^i, or both jointly. Columns labeled *GE* recompute the full equilibrium, so firms' technology choices, the entry and technology thresholds z^x and z^m, and the mass of firms M all adjust. Columns labeled *PE* hold technology choices and entry – z^x, z^m and M – fixed. Comparing the two isolates how much of the output gain from skill accumulation runs through the reorganization of production.

Raising the fixed cost to its US value reduces output, even though it raises the share of employment in large firms to 50%. A larger share of employment in large firms thus raises output when it results from a larger skill endowment, but not when it results from higher fixed costs.

Discussion. The output gains reported in this section are smaller than those found in related exercises in the literature, for two reasons. First, because L_h is a share, raising it also lowers L_l. A higher L_h thus converts low-skill into high-skill workers rather than making all workers more productive, as human capital often does in the literature. The output gain from this conversion is limited by the skill premium, and it is smaller still

when the two types of labor are imperfect substitutes.[44]

Second, higher skill weights μ^i raise output only when the skill endowment is large enough. Even then, they raise output by less than than the skill-augmenting productivity growth considered by, for example, Rossi [2022]. Skill augmenting productivity growth combines a change in effective skill weights with an increase in neutral productivity, whereas a change in skill weights alone leaves neutral productivity unchanged.[45]

Our focus, however, is not the overall size of these gains but how technology choices amplify them, and our parsimonious formulation fits the cross-country data on firm structure well.

6.2. Measured size-dependent distortions

A large literature argues that firms in low-income countries are small because of misallocation. The main argument brought forward is that there are distortions that increase with firm size or productivity, which prevent firms from growing and hiring [Restuccia and Rogerson, 2008, Hsieh and Klenow, 2009, Guner et al., 2008]. Our findings offer an alternative explanation in which small firms are an efficient response to skill scarcity. Our analysis also implies a need for caution in the measurement of distortions.

Inference in this literature rests on two conditions. First, in the absence of distortions, the marginal product of each input is equated across firms. Second, marginal products, which are not observed, are proportional to average products, which are observed. The second condition holds when firms share a common Cobb-Douglas technology without fixed costs. When both conditions hold, dispersion in measured average products reveals dispersion in marginal products and hence distortions, or wedges in first order conditions.

In our model, the first condition holds, but the second does not. All firms hire workers in the same competitive labor market and equate each factor's marginal product to its price, so marginal products are equated across firms and the first condition holds. The average product of labor of a firm with productivity z and technology i,

[44] Converting a low-skill worker into a high-skill worker raises output by the difference in marginal products, $w_h - w_l$. The implied proportional increase in output is $LIS_h/L_h - LIS_l/L_l$, where LIS_j is the labor income share of skill group j. (To see this: $(w_h - w_l)/Y = w_h L_h/(L_h Y) - w_l L_l/(L_l Y) \equiv LIS_h/L_h - LIS_l/L_l$.) In the LIC calibration, this term equals 0.587, so the change in L_h of 0.446 could raise output by at most 26%, if marginal products did not diminish. In even simpler terms, suppose that the two types of labor were perfect substitutes, and high-skill workers 84% more productive, implying the LIC skill premium of 1.84. Then, raising L_h from 0.513 to 0.959 also implies an output gain of 26%. In the model, imperfect substitutability reduces the output gain to 10.6%, about two-fifths of this bound.

[45] Suppose high-skill labor L_h is replaced by AL_h in the production function. For $\rho > 1$, raising A from 1 to $\bar{A} > 1$ is then equivalent to two changes. The first is a change in the skill weights from μ and $1-\mu$ to $\tilde{\mu}$ and $1-\tilde{\mu}$, where $\tilde{\mu} = \mu \bar{A}^{\frac{\rho-1}{\rho}} / (\mu \bar{A}^{\frac{\rho-1}{\rho}} + 1 - \mu)$. The second is a multiplication of the labor bundle $\bar{L}$ by $\left(\mu \bar{A}^{\frac{\rho-1}{\rho}} + 1 - \mu\right)^{\frac{\rho}{\rho-1}} > 1$, and hence of output by this term raised to the power γ. Higher A thus not only augments high-skill labor, like a higher μ, but also raises neutral productivity, unlike a higher μ.

including overhead workers, is

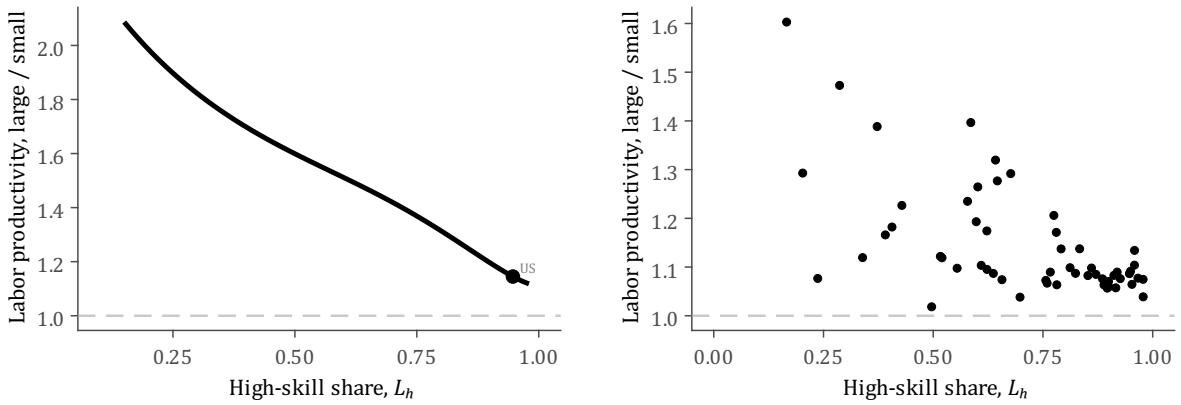
$$APL^i(z) = \frac{y^i(z)}{L^i(z) + c^f} = \frac{y^i(z)/L^i(z)}{1 + c^f/L^i(z)} = \frac{\bar{w}^i}{\gamma^i \phi^i} \frac{1}{1 + c^f/L^i(z)} \tag{13}$$

where the last equality uses $y^i(z)/\bar{L}^i(z) = \bar{w}^i/\gamma^i$ and $L^i(z) = \phi^i \bar{L}^i(z)$.[46] Even though marginal products are equated, the average product of labor thus differs across technologies and firm sizes, for three reasons. First, the unit cost $\bar{w}^i$ and the headcount conversion factor ϕ^i differ across technologies. Second, the traditional technology has lower returns to scale, which raises its average product relative to its marginal product, and thus the average product of traditional relative to modern firms. Third, small firms spread the fixed cost over fewer workers, which lowers their average product relative to large firms. Average products are therefore not proportional to marginal products, and their dispersion is not informative about distortions in our economy.

The second and third channels work in opposite directions, so the relation between average product and firm size cannot be signed in general. In the US calibration, the average product of labor in large firms is about 1.13 times that in small firms, because the fixed-cost channel dominates the effect of differences in returns to scale. Figure VIII shows how this ratio varies with the skill endowment. In Figure VIIIa, we vary L_h in the US calibration. A lower L_h raises the APL of large relative to small firms, because it shifts the composition of small firms towards traditional firms, for whom overhead makes up a larger share of employment. Figure VIIIb shows the ratio for each country at its own calibration. It is around 1.1 in high-income countries, and around 1.2 on average in low-income countries. The dispersion around the relationship with L_h mainly reflects cross-country variation in the calibrated fixed cost.

In both exercises, the average product of large relative to small firms is substantially higher in countries with low skill endowments. A measured productivity gap between large and small firms therefore does not by itself indicate distortions. In our framework it arises because firms operate different technologies and face fixed costs, and it is larger where skill scarcity pushes more firms into the traditional technology. Recent work shows that heterogeneity in technology affects measured markups [Foster et al., 2024] and measured factor shares [Hubmer and Restrepo, 2026]. Our results show that it also affects measured distortions.

[46] The first relation holds because the wage bill $\bar{w}^i \bar{L}^i(z)$ accounts for a share γ^i of a firm's revenue. The second converts the labor bundle $\bar{L}^i(z)$ into headcount production employment $L^i(z)$.



(A) Vary L_h in the US calibration

(B) Country-specific calibrations

FIGURE VIII
The average product of labor of large relative to small firms

Notes. Both panels plot the model-implied ratio of the average product of labor in large firms to that in small firms against the skill endowment L_h. Labor input includes overhead. Panel (a) varies L_h in the US calibration of Table IV, holding every other input at its US value. Panel (b) shows one marker per country, each evaluated at that country's own calibration — summarized by income group in Table V — so its spread around the Panel (a) relationship reflects cross-country variation in the skill weights μ^i and the fixed cost c^f. The model has no distortions and equates each factor's marginal product across firms, so a ratio above one arises without distortions.

7. Conclusion

This paper documents four facts on skills and the size structure of employment across countries. Countries with lower skill endowments have less employment in large firms. Large firms are more skill-intensive than small firms, and the gap is far wider in low-income countries. The premium for basic skills is highest where these skills are scarce. A heterogeneous-firm model with endogenous technology choice reproduces these patterns from differences in skill endowments alone, and calibrated country by country attributes 88% of the cross-country variation in large-firm employment to them.

Two implications follow. First, skill scarcity constrains not only individual productivity, but also the scale of production. The technologies that operate at scale rely more heavily on basic skills of the workforce. A larger skill endowment therefore raises output both directly and by moving firms into these technologies. This second channel adds about a third to the output gain from a larger skill endowment, a component that wage-based measures of the return to education do not capture. Analyses that hold the organization of production fixed miss this part of the gain. Second, a firm-size distribution dominated by small firms is not by itself evidence of distortions, nor are size-related differences in the average product across firms. Both can be an efficient response to skill scarcity. Where the workforce lacks the skills that scale requires, reforms of the business environment may therefore do less to change the organization of production. Investments in basic education, in turn, raise output in part by enabling

firms to adopt the technologies that operate at scale.

References

Daron Acemoglu and Fabrizio Zilibotti. Productivity differences. *Quarterly Journal of Economics*, 116(2):563–606, 2001.

Ufuk Akcigit, Harun Alp, and Michael Peters. Lack of selection and limits to delegation: Firm dynamics in developing countries. *American Economic Review*, 111(1):231–75, 2021.

Andrew Atkeson and Patrick J Kehoe. Modeling the transition to a new economy: lessons from two technological revolutions. *American Economic Review*, 97(1):64–88, 2007.

Abhijit V. Banerjee and Esther Duflo. Growth theory through the lens of development economics. In Philippe Aghion and Steven N. Durlauf, editors, *Handbook of Economic Growth*, volume 1A, chapter 7, pages 473–552. Elsevier, 2005.

Robert Barro and Jong-Wha Lee. A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics*, 104:184–198, 2013.

Eric Bartelsman, John Haltiwanger, and Stefano Scarpetta. Cross-country differences in productivity: The role of allocation and selection. *American Economic Review*, 103(1): 305–34, 2013.

Susanto Basu and David N. Weil. Appropriate technology and growth. *Quarterly Journal of Economics*, 113(4):1025–1054, 1998.

Pedro Bento and Diego Restuccia. Misallocation, establishment size, and productivity. *American Economic Journal: Macroeconomics*, 9(3):267–303, 2017.

Pedro Bento and Diego Restuccia. On average establishment size across sectors and countries. *Journal of Monetary Economics*, 117:220–242, 2021.

Mark Bils, Bariş Kaymak, and Kai-Jie Wu. Labor substitutability among schooling groups. *American Economic Journal: Macroeconomics*, 16(4):1–34, 2024.

Johannes Boehm and Ezra Oberfield. Misallocation in the market for inputs: Enforcement and the organization of production. *The Quarterly Journal of Economics*, 135(4): 2007–2058, 2020.

Charles Brown and James Medoff. The employer size-wage effect. *Journal of Political Economy*, 97(5):1027–1059, 1989.

Francisco J. Buera, Joseph P. Kaboski, and Yongseok Shin. Finance and development: A tale of two sectors. *American Economic Review*, 101(5):1964–2002, 2011.

Francesco Caselli and Antonio Ciccone. The contribution of schooling in development accounting: Results from a nonparametric upper bound. *Journal of Development Economics*, 104:199–211, 2013.

Francesco Caselli and Wilbur John Coleman. The world technology frontier. *American Economic Review*, 96(3):499–522, 2006.

Alvaro Cox. Fostering development through higher education: College attainment, firms and economic growth. *mimeo, Universidad Carlos III de Madrid*, 2025.

Austin Davis, Eric Hsu, and Mitchell VanVuren. Self employment, micro-entrepreneurship, and development. STEG Pathfinding Paper 12, Structural Transformation and Economic Growth (STEG), CEPR, 2023.

Niklas Engbom, Hannes Malmberg, Tommaso Porzio, Federico Rossi, and Todd Schoellman. Economic development according to Chandler. NBER Working Paper 34483, National Bureau of Economic Research, 2025.

Robert C. Feenstra, Robert Inklaar, and Marcel P. Timmer. The next generation of the Penn World Table. *American Economic Review*, 105(10):3150–3182, 2015.

Lucia S. Foster, John C. Haltiwanger, and Cody Tuttle. Rising markups or changing technology? NBER Working Paper 30491, National Bureau of Economic Research, 2024.

Wei Gao and Matthias Kehrig. Returns to Scale, Productivity and Competition: Empirical Evidence from U.S. Manufacturing and Construction Establishments. *mimeo, Duke University*, 2025. ISSN 1556-5068. doi: 10.2139/ssrn.2731596.

Johannes Goensch. Non-convergence of East German wages: Effects of skill shortage on firm organization. *mimeo, Goethe Universität Frankfurt*, 2025.

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

Charles Gottlieb, Jan Grobovšek, and Alex Monge-Naranjo. Occupational choices, human capital and cross-country income differences. *mimeo*, 2026.

Jan Grobovšek. Managerial delegation, law enforcement, and aggregate productivity. *The Review of Economic Studies*, 87(5):2256–2289, 2020.

Nezih Guner, Gustavo Ventura, and Yi Xu. Macroeconomic implications of size-dependent policies. *Review of Economic Dynamics*, 11(4):721–744, 2008.

Nezih Guner, Andrii Parkhomenko, and Gustavo Ventura. Managers and productivity differences. *Review of Economic Dynamics*, 29:256–282, 2018.

Daniel Haanwinckel and Rodrigo R Soares. Workforce composition, productivity, and labour regulations in a compensating differentials theory of informality. *The Review of Economic Studies*, 88(6):2970–3010, 2021.

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

Ricardo Hausmann and César A. Hidalgo. The network structure of economic output. *Journal of Economic Growth*, 16(4):309–342, 2011.

Lutz Hendricks and Todd Schoellman. Skilled labor productivity and cross-country income differences. *American Economic Journal: Macroeconomics*, 15(1):240–68, 2023.

Jonas Hjort, Hannes Malmberg, and Todd Schoellman. The missing middle managers:

Labor costs, firm structure, and development. *American Economic Review: Insights*, 8 (3):341–353, 2026.

Hugo Hopenhayn. Entry, exit, and firm dynamics in long run equilibrium. *Econometrica*, 60(5):1127–1150, 1992.

Hugo Hopenhayn and Richard Rogerson. Job turnover and policy evaluation: A general equilibrium analysis. *Journal of Political Economy*, 101(5):915–938, 1993.

Chang-Tai Hsieh and Peter J. Klenow. Misallocation and manufacturing TFP in China and India. *The Quarterly Journal of Economics*, 124(4):1403–1448, 2009.

Joachim Hubmer and Pascual Restrepo. Not a typical firm: Capital–labor substitution and firms' labor shares. *American Economic Journal: Macroeconomics*, 18(2):34–71, 2026.

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

Lawrence F. Katz and Kevin M. Murphy. Changes in relative wages, 1963-1987: Supply and demand factors. *The Quarterly Journal of Economics*, 107(1):35–76, 1992.

Peter J Klenow and Huiyu Li. Entry costs rise with growth. *Journal of Political Economy Macroeconomics*, 3(1):43–74, 2025.

Alexandr Kopytov, Mathieu Taschereau-Dumouchel, and Zebang Xu. Endogenous Returns to Scale. *mimeo, Cornell University*, 2026.

Michael Kremer. The O-ring theory of economic development. *The Quarterly Journal of Economics*, 108(3):551–575, 1993.

Virgiliu Midrigan and Yi Xu. Finance and misallocation: Evidence from plant-level data. *American Economic Review*, 104(2):422–58, 2014.

Hernan Moscoso Boedo and Toshihiko Mukoyama. Evaluating the effects of entry regulations and firing costs on international income differences. *Journal of Economic Growth*, 17(2):143–170, 2012.

Walter Y. Oi and Todd L. Idson. Firm size and wages. In Orley Ashenfelter and David Card, editors, *Handbook of Labor Economics*, volume 3, chapter 33, pages 2165–2214. Elsevier, 1999.

Serdar Ozkan, Sergio Salgado, Joachim Hubmer, Guangbin Hong, and Mons Chan. Scalable versus Productive Technologies. Technical report, 2026.

Markus Poschke. Employment protection, firm selection, and growth. *Journal of Monetary Economics*, 56(8):1074–1085, 2009.

Markus Poschke. The regulation of entry and aggregate productivity. *The Economic Journal*, 120(549):1175–1200, 2010.

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

George Psacharopoulos. Returns to investment in education: A global update. *World*

Development, 22(9):1325–1343, 1994.

George Psacharopoulos and Harry Anthony Patrinos. Returns to investment in education: a further update. *Education Economics*, 12(2):111–134, 2004.

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

Federico Rossi. The relative efficiency of skilled labor across countries: Measurement and interpretation. *American Economic Review*, 112(1):235–266, 2022.

Kenneth R. Troske. Evidence on the employer size-wage premium from worker-establishment matched data. *The Review of Economics and Statistics*, 81(1):15–26, 1999.

Gabriel Ulyssea. Regulation of entry, labor market institutions and the informal sector. *Journal of Development Economics*, 91(1):87–99, 2010.

A. Data

Sources: The primary data sources for the household and labor force surveys used to assemble our final dataset are listed in Table A.I. The table includes the country, the survey name, and the earliest and latest survey-year observation that satisfy our selection criteria.

TABLE A.I
Household and Labor Force Surveys

Country	Survey name	Year	Wage sample	Balanced panel
Albania	Labour Force Survey	2013	✓	-
Angola	Inquerito Integrado Sobre O Bem-Estar Da Populacao	2008	-	-
Argentina	Encuesta Permanente de Hogares	2015	✓	✓
Armenia	Labour Force Survey	2016	✓	-
Australia	Household, Income And Labour Dynamics In Australia	2015	✓	✓
Austria	European Union Statistics On Income And Living Conditions	2015	✓	✓
Benin	Enquête Modulaire Intégrée sur les Conditions de Vie des Ménages	2015	✓	-
Bolivia	Encuesta de Hogares	2015	✓	-
Burkina Faso	Enquête Nationale sur Le Bien-Être et les Inégalités Sociales	2023	✓	-
Cambodia	Cambodia Labor Force And Child Labor Survey	2012	✓	-
Cameroon	Fourth Cameroon Household Survey	2014	✓	-
Chile	Encuesta de Caracterización Socioeconómica Nacional	2015	✓	-
China	Family Panel Studies	2016	✓	-
Colombia	Gran Encuesta Integrada de Hogares	2015	✓	✓
Cyprus	European Union Statistics On Income And Living Conditions	2015	✓	✓
Czechia	European Union Statistics On Income And Living Conditions	2015	✓	-
Denmark	European Union Statistics On Income And Living Conditions	2015	✓	-
Ecuador	Encuesta Nacional de Empleo, Desempleo y Subempleo	2015	✓	✓
Egypt	Harmonized Labor Force Survey	2015	✓	✓
Ethiopia	Socioeconomic Survey	2018	✓	-
Finland	European Union Statistics On Income And Living Conditions	2006	✓	-
France	Enquête Emploi en Continu	2015	✓	✓
Germany	European Union Statistics On Income And Living Conditions	2015	✓	✓
Ghana	Labour Force Survey	2015	✓	-
Guinea	Enquête Nationale sur L'emploi et Le Secteur Informel	2019	✓	-
Iceland	European Union Statistics On Income And Living Conditions	2015	✓	-
Iraq	Household Socio-Economic Survey	2012	✓	-
Japan	Employment Status Survey	2017	✓	-
Liberia	Household Income And Expenditure Survey	2016	✓	-
Malawi	Integrated Household Survey	2019	✓	-
Malta	European Union Statistics On Income And Living Conditions	2015	✓	-
Mexico	Encuesta Nacional de Ocupación y Empleo	2015	✓	-

continued on next page

Table A.I – continued from previous page

Country	Survey name	Year	Wage sample	Balanced panel
Mongolia	Labor Force Survey	2015	✓	-
Namibia	Labor Force Survey	2018	✓	-
Netherlands	European Union Statistics On Income And Living Conditions	2015	✓	✓
Niger	Enquete Nationale sur L'emploi et Le Secteur Informel	2012	✓	-
Nigeria	National Labour Force Survey	2023	✓	-
Norway	European Union Statistics On Income And Living Conditions	2015	✓	✓
Palestinian Territories	Harmonized Labor Force Survey	2014	✓	-
Peru	Encuesta Nacional de Hogares	2015	✓	✓
Russia	Russia Longitudinal Monitoring Survey	2015	✓	✓
Rwanda	Labor Force Survey	2017	✓	-
Senegal	Enquête Nationale sur L'emploi Au Sénégal	2017	✓	-
Sierra Leone	Integrated Household Survey	2018	✓	-
Slovenia	European Union Statistics On Income And Living Conditions	2015	✓	✓
South Africa	Labor Market Dynamics	2015	✓	-
South Korea	Korean Labor And Income Panel Study	2015	✓	✓
Sri Lanka	Labor Force Survey	2015	✓	-
Sweden	European Union Statistics On Income And Living Conditions	2005	✓	-
Switzerland	European Union Statistics On Income And Living Conditions	2015	✓	-
Tanzania	Living Standards Measurement Survey	2014	✓	-
Uganda	Labor Force Survey	2017	✓	-
United Kingdom	European Union Statistics On Income And Living Conditions	2011	✓	-
United States	Current Population Survey	2015	✓	-
Uruguay	Encuesta Continua de Hogares	2015	✓	✓
Zambia	Labour Force Survey	2017	✓	-
Zimbabwe	Labour Force And Child Labour Survey	2014	✓	-

Notes. The table lists the household and labour force surveys underlying the analysis, one row for each of the 57 countries in the cross-section. For each country it gives the survey name and the survey year used. A check mark in "Wage sample" indicates that the survey also enters the subsample with reliable wage information (the basis for the skill-premium and Mincer results); a check mark in "Balanced panel" indicates that the country is part of the balanced panel used in the time-series analysis. The cross-section analysis uses one survey per country: the survey carrying the required information (both firm size and education) whose year is closest to 2015. For a few countries, the only such survey in the assembled data predates 2015, so an earlier year is used even where wage-less surveys exist for later years.

B. Dataset validation

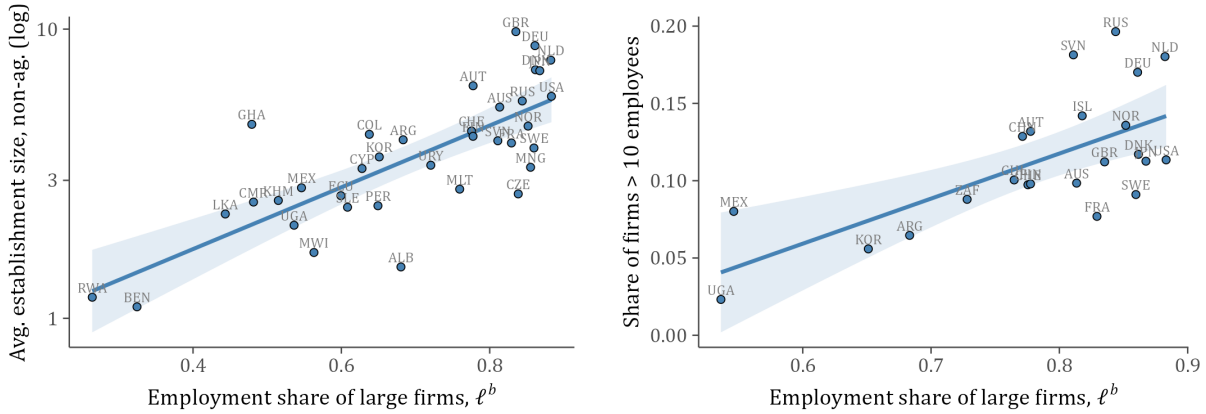
This appendix checks two of our measures against other cross-country datasets. The share of employment in large firms has no like-for-like benchmark, since no other source records employer size for a representative sample of workers across the development spectrum. To validate our dataset, we check whether it orders countries the same way as the established firm-side evidence on firm size. The skill endowment is checked

against the standard source on educational attainment.

Firm size. We use two firm-side sources on the size structure of production. Bento and Restuccia [2021] compile average establishment size in manufacturing, services, and non-agriculture from economic censuses and nationally representative surveys; 42 of our 57 countries appear in their data, 36 with the non-agricultural measure. Poschke [2018] computes mean firm size and the share of firms with more than ten employees for 44 countries from the Global Entrepreneurship Monitor (GEM), 23 of which are in our data.

Neither source measures the same statistic as ours, the purpose of our comparison is to evaluate their correlation across countries. Their establishment averages include non-employers, the GEM universe is businesses owned by respondents rather than all employers, and our sample is wage employees who report the size of their establishment. Both sources also predate our surveys by about a decade, the establishment data being centered on 2000–2012 and the GEM moments on 2003–2005, against a median survey year of 2015 in our data. We report the cross-country correlation between each published measure and our share of employment in large firms, ℓ^b, together with the slope from a regression of the former on the latter.

Figure B.I and Table B.I report the results. Our measure correlates at 0.77 with average establishment size in non-agriculture and at 0.72 with that in services in Bento and Restuccia [2021], and at 0.67 with the share of firms above ten employees and at 0.60 with mean business size in Poschke [2018]. The estimated coefficients are significant at the 1% level. A ten percentage point higher ℓ^b corresponds to an average establishment size about 25 log points higher, a share of firms above ten employees 2.9 percentage points higher, and a mean business size 35 log points higher. The correlation is weakest for manufacturing (0.65), as expected since our sample is predominantly services, and replacing ℓ^b by its non-agricultural counterpart leaves the slope nearly unchanged (column 4 of Panel A). Our measure of the size structure of employment is thus strongly correlated with firm-side measures of firm size, across sources with different sampling frames, units of observation, and vintages.



(A) Average establishment size in non-agriculture (B) Share of firms with more than ten employees

FIGURE B.I
Our employment share in large firms against published firm-size measures

Notes. Each marker is one country, labelled with its ISO3 country code; the line is a least-squares fit with a 95% confidence band. The horizontal axis in both panels is ℓ^b, our share of wage employment in firms with ten or more employees, from the baseline cross-section. Panel (a) plots average establishment size in non-agriculture from Bento and Restuccia [2021], on a log scale for 36 countries. Panel (b) plots the share of businesses with more than ten employees from Poschke [2018], computed from the Global Entrepreneurship Monitor for 23 countries. Note that the vertical axis in Panel (b) is a share of *firms*, not of employment, and so is not comparable in level to ℓ^b. The establishment-size data are averaged over 2000–2012 for manufacturing and taken at the year closest to 2007 for services; the GEM moments pool the 2003–2005 waves; our surveys span 2005–2023 with a median of 2015.

TABLE B.I
Published firm-size measures regressed on our employment share in large firms

Panel A: average establishment size [Bento and Restuccia, 2021]

	Non-agriculture (1)	Manufacturing (2)	Services (3)	Non-agriculture (4)
Employment share in large firms	2.46***	2.77***	2.21***	
	(0.323)	(0.620)	(0.309)	
Employment share in large firms, non-agriculture				2.65***
				(0.500)
Constant	-0.433*	0.215	-0.440**	-0.618*
	(0.226)	(0.456)	(0.206)	(0.354)
Correlation	0.77	0.65	0.72	0.72
Observations	36	40	39	31
R^2	0.591	0.416	0.513	0.521

Panel B: size of GEM businesses [Poschke, 2018]

	Mean size (log) (1)	Share $>$ 10 employees (2)
Employment share in large firms	3.49***	0.292***
	(0.990)	(0.066)
Constant	-0.648	-0.116**
	(0.811)	(0.050)
Correlation	0.60	0.67
Observations	23	23
R^2	0.358	0.451

Notes. Each column is a separate bivariate OLS regression of the published measure named in the column head on ℓ^b, our share of wage employment in firms with ten or more employees, with one observation per country. Heteroskedasticity-robust standard errors in parentheses. The *Correlation* row reports the simple cross-country correlation between the two variables on the estimation sample. In Panel A the dependent variable is the log of average establishment size in the sector named in the column head, from Bento and Restuccia [2021]; columns (1)–(3) use our all-sector ℓ^b as the regressor and column (4) uses its non-agricultural counterpart, so that both sides refer to the same part of the economy. In Panel B the dependent variables are the log of mean business size and the share of businesses with more than ten employees, from Poschke [2018] and computed from the Global Entrepreneurship Monitor; that share counts *firms* rather than employment and so is not comparable in level to ℓ^b. Samples differ across columns because coverage of the published sources differs by sector and by measure. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Skills. We compare average years of schooling, the variable from which we compute the high-skill share, with the educational attainment series of Barro and Lee [2013] in its latest release (version 3.0), which covers 146 countries, 51 of them in our data. We compare years of schooling rather than the high-skill share because their secondary category spans lower and upper secondary and cannot be mapped onto our nine-year threshold. Two differences between the series are known in advance. Theirs covers the whole population and ours wage employees, who are better schooled, so our levels are higher by construction, and by more in low-income countries, where the gap between the employed and the population is largest. Their age window is 15–64, so we run the comparison on the 16–64 sample of Appendix C.5 and repeat it on the baseline 25–55 sample. Figure B.II and Table B.II report the results. The two series correlate at 0.89 across countries, with a slope of 1.04. For the baseline sample the correlation is 0.89. Our measure of the skill endowment thus tracks closely the standard data source on educational attainment.

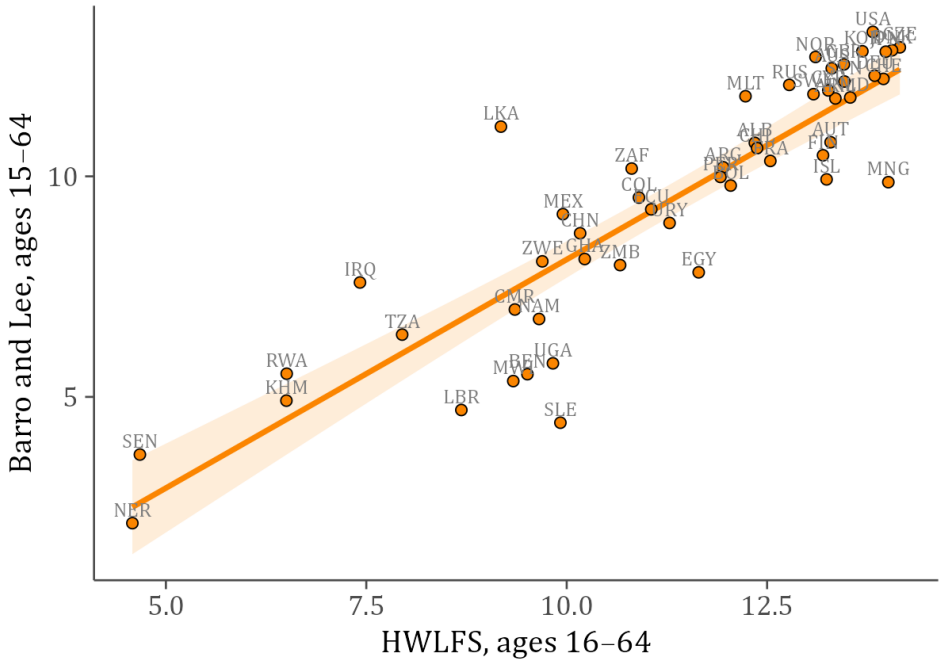


FIGURE B.II
Average years of schooling : our dataset versus Barro and Lee [2013]

Notes. Each marker is one country, labelled with its ISO3 country code; the line is a least-squares fit with a 95% confidence band. The vertical axis is average years of schooling from Barro and Lee [2013], for the population aged 15–64. The horizontal axis is the same statistic from our data, computed on the 16–64 sample of Appendix C.5 so that the age windows are comparable; 51 countries. Our measure covers wage employees and theirs the whole population, so our levels are higher throughout, by more where schooling is lower. Barro and Lee [2013] report quinquennial data ending in 2015, which we interpolate linearly to each survey year; surveys after 2015 take the 2015 value.

TABLE B.II
Average years of schooling : our dataset versus Barro and Lee [2013]

	Avg. years of schooling	
	(1)	(2)
Our age window	16–64	25–55
Barro–Lee age window	15–64	25–64
Avg. years of schooling, ours	1.04***	1.18***
	(0.068)	(0.072)
Constant	-2.25**	-4.23***
	(0.864)	(0.940)
Correlation	0.89	0.89
Observations	51	51
R^2	0.799	0.801

Notes. Each column is a separate bivariate OLS regression, with one observation per country and heteroskedasticity-robust standard errors in parentheses. The *Correlation* row reports the simple cross-country correlation between the two variables. The dependent variable is average years of schooling from Barro and Lee [2013], for the population aged 15–64 in column (1) and 25–64 in column (2). The regressor is the same statistic from our data, computed on the 16–64 sample of Appendix C.5 in column (1), so that the age windows are comparable, and on the baseline 25–55 sample in column (2). Their series covers the whole population and ours wage employees, so the intercepts are negative by construction and the slope and the correlation carry the content. Barro and Lee [2013] report quinquennial data ending in 2015; each country is matched to its nearest available year. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

C. Robustness of the cross-country evidence

This appendix replicates the exhibits of Section 2 under a series of alternative measurement choices. Each subsection covers one exercise and carries that exercise's version of the main-text exhibits, in the order the paper presents them: the cross-country scatters of Figure I and Figure II, the summary statistics of Table I, the skill-intensity figure Figure IV, the two-perspective regressions of Table III, and the skill-price figure Figure III. Exhibits that are not robustness variants of anything in the main text, including the evidence by broad sector, are collected in Online Appendix D instead.

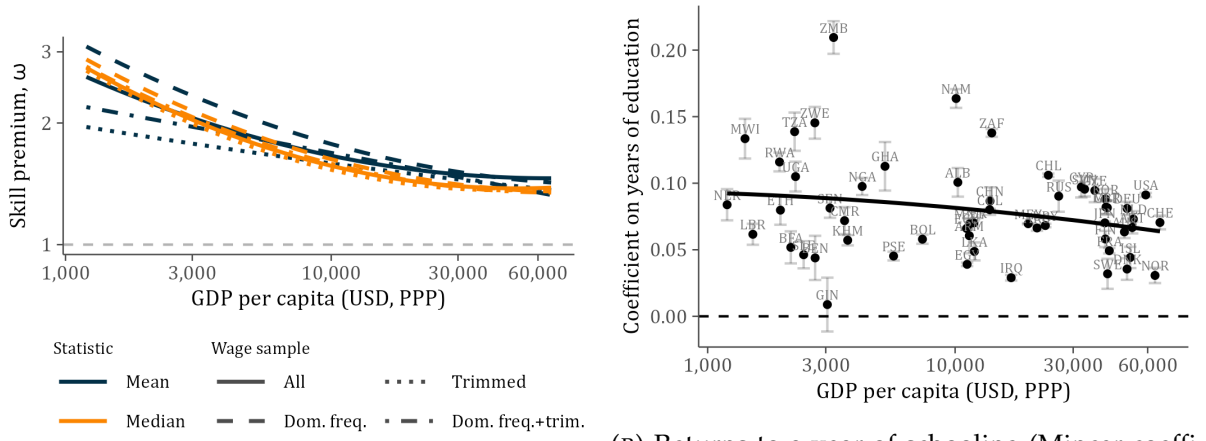
C.1. Alternative measures of the skill premium

We measure the skill premium in ten ways, which differ along three choices. The first is the statistic: the ratio of mean or of median hourly wages of high- to low-skill workers, or the conditional Mincer return to high skill, controlling for age, age^2, sex, marital status and the wage-reporting frequency. The second is the treatment of outliers: whether the top and bottom 1% of log wages are trimmed within each country. The third, for the two wage ratios, is the treatment of pay frequency: whether the sample is restricted to workers paid at their country's dominant reporting frequency, since higher-frequency pay carries a premium and low-skill workers are more often paid by the day or week. Table C.I reports all ten measures side by side, and Figure C.I plots the gradients of the eight wage ratios. The premium declines with GDP per capita under every measure, with elasticities between -0.09 and -0.18, and the measure we use in the main text, the trimmed ratio of mean wages, has the flattest gradient. The two remaining subsections that touch the wage sample, Section C.2 and Section C.5, report the same table under their own sample definition.

TABLE C.I
The skill premium across countries, alternative measures

	Country income				
	Low-income	Middle-income	High-income	US	Elasticity
Ratio of mean wages					
No restriction	2.35	1.84	1.45	1.94	-0.14^{**}
Trimmed tails (1%)	1.84	1.69	1.39	1.79	-0.09^{***}
Dominant pay frequency	2.54	1.89	1.45	1.94	-0.18^{***}
Dominant frequency + trimmed	2.05	1.72	1.39	1.79	-0.12^{***}
Ratio of median wages					
No restriction	2.21	1.60	1.38	1.71	-0.16^{***}
Trimmed tails (1%)	2.17	1.55	1.37	1.68	-0.16^{***}
Dominant pay frequency	2.35	1.61	1.38	1.71	-0.18^{***}
Dominant frequency + trimmed	2.26	1.57	1.37	1.68	-0.17^{***}
Mincer return on high skill (log points)					
No restriction	0.76	0.50	0.37	0.60	-0.15^{***}
Trimmed tails (1%)	0.69	0.45	0.32	0.57	-0.14^{***}
Number of countries	20	17	19	1	

Notes. Each row is one measure of the price of skill. The first four columns report the average of the country-level measure within an income group, with the United States shown separately for reference; the grouping is the one Table I uses (LIC below \$10,000 PPP, MIC \$10,000–\$30,000, HIC above \$30,000), and countries are weighted equally within a group. The first two blocks report the ratio of average high- to low-skill wages, using mean and median wages respectively; within each, the four rows vary the wage sample — all workers; only workers paid at their country's dominant wage-reporting frequency; the top and bottom 1% of log wages trimmed within each country; and both restrictions together. The third block reports the country-level coefficient on a high-skill indicator from a Mincer regression of log wages on the indicator, age, age^2, sex, marital status and the wage-reporting frequency, estimated separately for each country, on the untrimmed and on the trimmed wage sample. Because that regression already controls for the payment frequency, it has no dominant-frequency counterpart. The row the paper reports in Table I and plots in Figure III is the trimmed mean in the first block and the trimmed Mincer return in the third. The *Elasticity* column is the coefficient on log GDP per capita from a country-level regression of the measure on log GDP per capita, with standard errors clustered by country; it is an elasticity for the two ratio blocks, whose dependent variable is the log premium, and a semi-elasticity for the Mincer block, which is already in log points. High-skill denotes at least ten years of schooling. Sample: wage employees aged 25–55 reporting positive wages and hours, one survey per country (the survey year closest to 2015); survey weights are rescaled within each country–year cell. The surveys used for each country are listed in Table A.I. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.



(A) Eight measures of the skill premium

(B) Returns to a year of schooling (Mincer coefficient) by country

FIGURE C.I
The skill premium across countries, alternative measures

Notes. Both panels plot the price of skill against GDP per capita (PPP, log scale). Panel (a) overlays the fitted gradients of eight alternative measures of the skill premium, crossing mean vs. median wages (carried by colour) with four wage samples (all workers; only workers paid at their country's dominant wage-reporting frequency; the top and bottom 1% of log wages trimmed; and both restrictions together, carried by the dash pattern); the country markers are omitted there, since eight measures across 56 countries would bury the comparison the panel is for. In Panel (b) each marker is one country, labeled with its ISO3 country code, the whiskers are 95% confidence intervals, and the solid line is a quadratic (in log GDP per capita) least-squares fit across countries; the plotted measure is the country-level coefficient on years of schooling from a Mincer wage regression that also controls for age, age^2, sex, marital status and the wage-reporting frequency, estimated separately by country on the trimmed wage sample. The premium axis is on a log scale in Panel (a), where the dashed horizontal line marks parity (no skill premium); in Panel (b), which is already in log-wage points, it marks a coefficient of zero. High-skill denotes at least ten years of schooling. Sample: wage employees aged 25–55 reporting positive wages and hours, one survey per country (the survey year closest to 2015); survey weights are rescaled within each country–year cell. The surveys used for each country are listed in Table A.I. GDP per capita is PPP-adjusted [Feenstra et al., 2015].

C.2. Alternative definition of skilled: at least thirteen years of schooling

Throughout this subsection a worker counts as high-skill only with at least thirteen years of schooling, instead of at least ten. Nothing else changes: the sample, the firm-size threshold and every denominator are those of Section 2. All four facts survive, and Facts 2 and 4 are if anything stronger. The share of wage employment in large firms carries no skill information, so Panel (b) of Figure C.III is identical to Panel (b) of Figure II; it is reproduced here only to keep the figure self-contained.

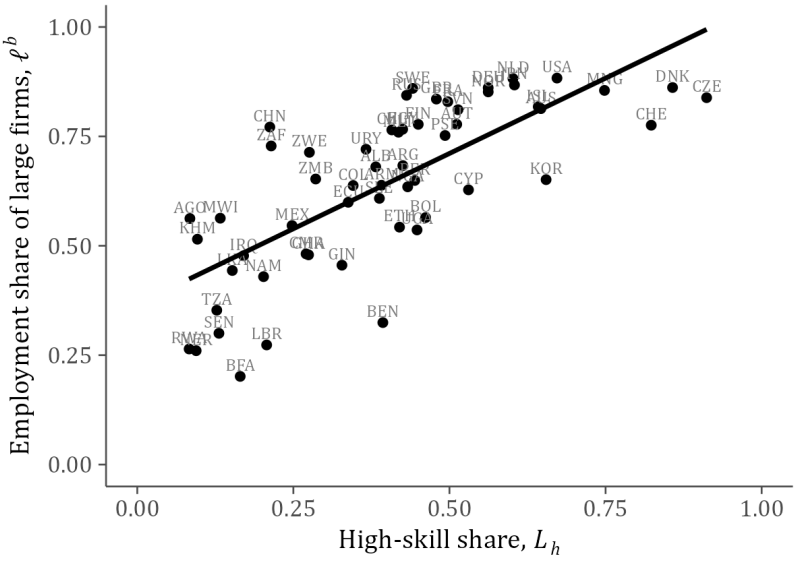
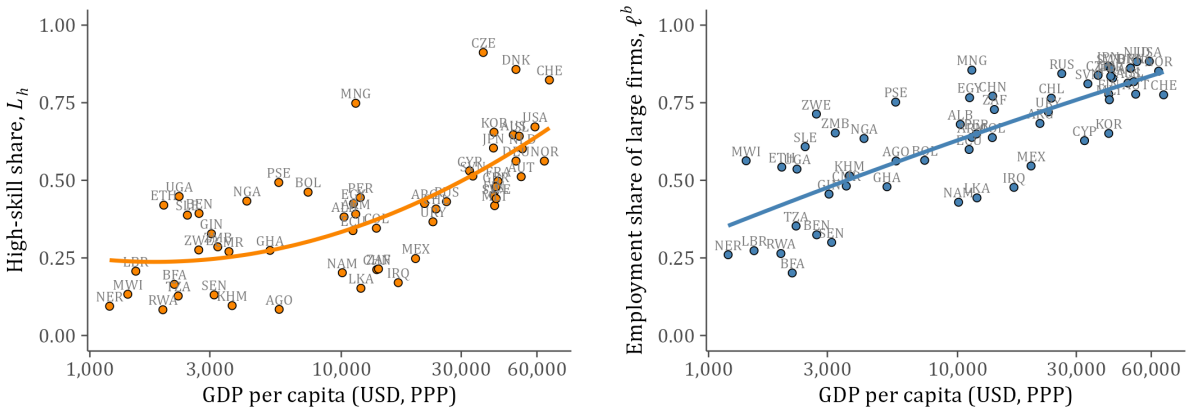


FIGURE C.II
Skills and employment in large firms, alternative definition of skilled

Notes. As in Figure I, but high-skill workers are those with at least thirteen years of schooling instead of at least ten. Each marker is one country, labeled with its ISO3 country code, and the solid line is a straight-line least-squares fit across countries. Large firms have at least ten employees. Sample: wage employees aged 25–55 with non-missing education and firm-size information, one survey per country (the survey year closest to 2015); survey weights are rescaled within each country–year cell. The surveys used for each country are listed in Table A.I.



(A) Share of skilled workers.

(B) Share of employment in large firms.

FIGURE C.III
Skill endowments and employment by firm size across countries, alternative definition of skilled

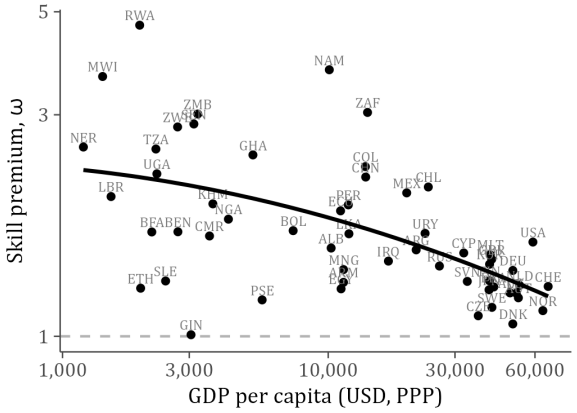
Notes. As in Figure II, but high-skill workers are those with at least thirteen years of schooling instead of at least ten. In both panels each marker is one country, labeled with its ISO3 country code. Panel (a) plots the high-skill share and Panel (b) the share of wage employment in large firms, both against GDP per capita (PPP, log scale). The solid line in each panel is a quadratic (in log GDP per capita) least-squares fit across countries. Panel (b) carries no skill information and is identical to Panel (b) of Figure II; it is reproduced here only to keep the figure self-contained. Large firms have at least ten employees. Sample: wage employees aged 25–55 with non-missing education and firm-size information, one survey per country (the survey year closest to 2015); survey weights are rescaled within each country–year cell. The surveys used for each country are listed in Table A.I. GDP per capita is PPP-adjusted [Feenstra et al., 2015].

TABLE C.II
Skills, the organization of production and the skill premium by country income group, alternative definition of skilled

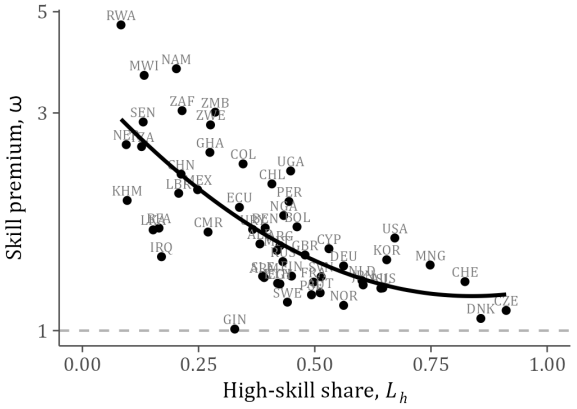
	Country income			
	Low-income	Middle-income	High-income	US
Employment composition				
Years of schooling	9.1	11.4	13.6	14.0
Share of high-skilled workers L_h	26.6%	34.7%	59.9%	67.2%
Share of employment in large firms ℓ^b	47.8%	66.1%	81.0%	88.3%
Skill premium				
Unconditional: mean wage ratio $\omega = w_h/w_l$	2.20	1.91	1.30	1.60
Conditional: returns to skill (log points)	0.85	0.62	0.30	0.45
Probability of working for a large firm by skill group				
High skill L_h^b/L_h	69.8%	82.4%	84.0%	90.9%
Low skill L_l^b/L_l	40.4%	58.5%	77.0%	83.1%
Skill intensity				
Large firms h^b	37.5%	42.6%	62.0%	69.2%
Small firms h^s	16.1%	20.0%	51.4%	52.5%
Ratio (large / small)† h^b/h^s	3.41	2.59	1.25	1.32
Number of countries	21	17	19	1

Notes. As in Table I, but workers are classified as high-skill only if they have at least thirteen years of schooling, instead of at least ten. Only the skill label changes: the sample, the firm-size threshold and every denominator are those of Table I. Each column is the unweighted average of the country-level statistic within an income group (LIC: GDP per capita below \$10,000 PPP; MIC: \$10,000–\$30,000; HIC: above \$30,000), with the United States shown separately. See Table I for the definitions of the individual rows. The surveys used for each country are listed in Table A.I.

Panel A: Skill premium

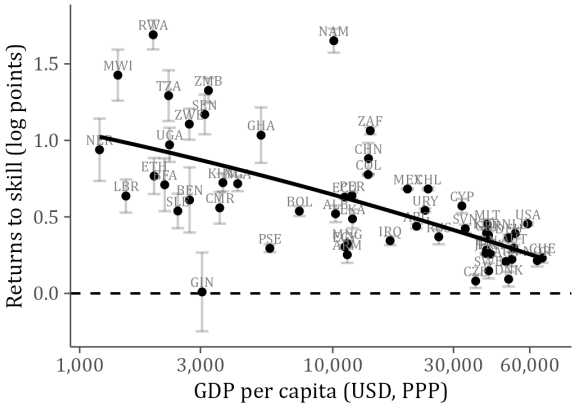


(A) Against GDP per capita.

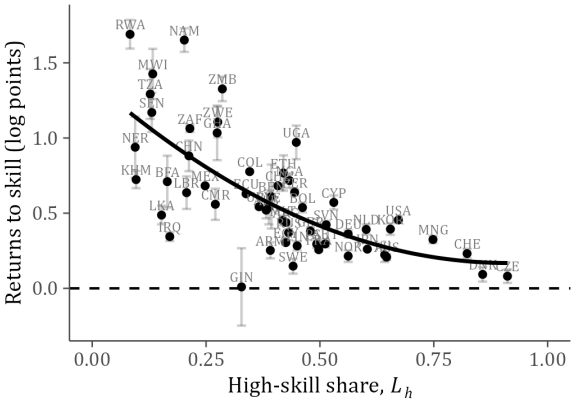


(B) Against the skill endowment L_h.

Panel B: Mincer coefficient



(C) Against GDP per capita.



(D) Against the skill endowment L_h.

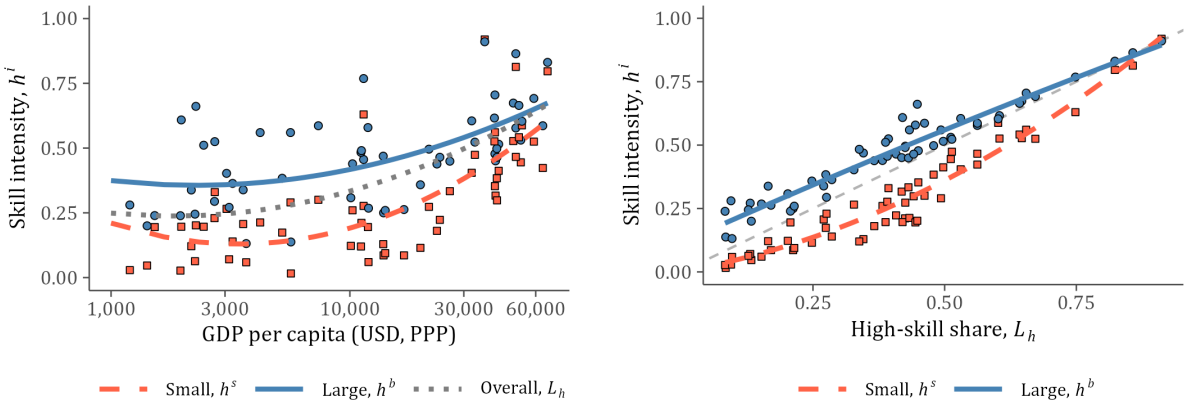
FIGURE C.IV
Skill prices across countries, alternative definition of skilled

Notes. As in Figure III, but high-skill workers are those with at least thirteen years of schooling instead of at least ten. Each marker is one country, labeled with its ISO3 country code. *Panel A*, Figures (a) and (b), shows the skill premium, the ratio of average wages of high-skill to low-skill wage employees, with the top and bottom 1% of log wages trimmed within each country; its vertical axis is on a log scale and the dashed horizontal line marks parity (no skill premium). *Panel B*, Figures (c) and (d), shows the country-level coefficient on a high-skill indicator from a Mincer regression of log wages on the indicator and individual controls (age, age^2, sex, marital status and the wage-reporting frequency), estimated separately for each country; whiskers are 95% confidence intervals and the dashed horizontal line marks a coefficient of zero. Within each panel, the left-hand figure plots the measure against GDP per capita (PPP, log scale) and the right-hand one against the country's high-skill share L_h. The solid line is in every case a quadratic least-squares fit across countries. Sample: wage employees aged 25–55 reporting positive wages and hours, one survey per country (the survey year closest to 2015); survey weights are rescaled within each country–year cell. The surveys used for each country are listed in Table A.I. GDP per capita is PPP-adjusted [Feenstra et al., 2015].

TABLE C.III
The skill premium across countries, alternative measures — alternative definition of skilled

	Country income				
	Low-income	Middle-income	High-income	US	Elasticity
Ratio of mean wages					
No restriction	2.59	2.09	1.35	1.69	-0.19^{***}
Trimmed tails (1%)	2.20	1.91	1.30	1.60	-0.16^{***}
Dominant pay frequency	3.13	2.11	1.35	1.69	-0.25^{***}
Dominant frequency + trimmed	2.31	1.91	1.30	1.60	-0.18^{***}
Ratio of median wages					
No restriction	2.78	2.23	1.31	1.62	-0.24^{***}
Trimmed tails (1%)	2.64	2.17	1.30	1.58	-0.23^{***}
Dominant pay frequency	2.89	2.22	1.31	1.62	-0.25^{***}
Dominant frequency + trimmed	2.76	2.16	1.30	1.58	-0.24^{***}
Mincer return on high skill (log points)					
No restriction	0.92	0.68	0.33	0.48	-0.21^{***}
Trimmed tails (1%)	0.85	0.62	0.30	0.45	-0.20^{***}
Number of countries	20	17	19	1	

Notes. As in Table C.I, but high-skill workers are defined as those with at least thirteen years of schooling instead of at least ten. Only the skill label changes; the wage sample, the four sample restrictions, the Mincer specification and the elasticity column are those of Table C.I, where the individual rows and columns are defined. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.



(A) The skill intensity of large and small firms vs GDP per capita

(B) The skill intensity of large and small firms and the aggregate supply of skills

FIGURE C.V
Skill and firm size, alternative definition of skilled

Notes. As in Figure IV, but high-skill workers are those with at least thirteen years of schooling instead of at least ten. Each marker is one country. Panel (a) plots the high-skill share of small firms, h^s, and of large firms, h^b, against GDP per capita; the lines are least-squares fits on a quadratic in log GDP per capita, and the dotted line is the same fit for the economy-wide share L_h. Panel (b) plots the same shares against the country-wide high-skill share L_h, with least-squares fits on a quadratic in L_h; the dashed line is the 45-degree line, which the economy-wide share follows by construction. Large firms have at least ten employees. Sample: wage employees aged 25–55 with non-missing education and firm-size information, one survey per country (the survey year closest to 2015); survey weights are rescaled within each country–year cell. The surveys used for each country are listed in Table A.I. GDP per capita is PPP-adjusted [Feenstra et al., 2015].

TABLE C.IV
The skill gap in large-firm employment, from the worker and the firm side, alternative definition of skilled

	Working in large firm			Skill intensity h^i		
	(1)	(2)	(3)	(4)	(5)	(6)
Unit of observation	Worker			Country × firm size		
1{High-skill}	0.195***	0.204***	0.205***			
	(0.017)	(0.014)	(0.015)			
1{High-skill} × log GDP p.c.		-0.071***				
		(0.012)				
1{High-skill} × high-skill share			-0.444***			
			(0.058)			
1{Large firm}				0.182***	0.182***	0.182***
				(0.013)	(0.012)	(0.013)
1{Large firm} × log GDP p.c.					-0.033***	
					(0.011)	
1{Large firm} × high-skill share						-0.135**
						(0.053)
Controls	✓	✓	✓			
Country fixed effects	✓	✓	✓	✓	✓	✓
No. of countries	57	57	57	57	57	57
Observations	1,162,815	1,162,815	1,162,815	114	114	114
R^2	0.184	0.190	0.189	0.946	0.955	0.950
Within R^2	0.039	0.046	0.045	0.773	0.811	0.789

Notes. As in Table III, but high-skill workers are those with at least thirteen years of schooling instead of at least ten. Columns (1)–(3) are worker-level linear probability models for working in a large firm, with a high-skill indicator and controls for age, age^2, sex and marital status; columns (4)–(6) are country–firm-size cell regressions of the cell's high-skill share on a large-firm indicator. Country fixed effects throughout. Columns (2) and (5) interact the gap with log GDP per capita and columns (3) and (6) with the country's high-skill share, both entered centered on their means. See Table III for the full description of the two blocks and why their magnitudes are not comparable. Standard errors are clustered by country. Large firms have at least ten employees. Sample: wage employees aged 25–55 with non-missing education and firm-size information, one survey per country (the survey year closest to 2015). $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE C.V
Skills and the probability of working in a large firm, alternative definition of skilled

	Working in large firm		
	(1)	(2)	(3)
Constant	0.329***		
	(0.077)		
1{High-skill}	0.269***	0.195***	0.208***
	(0.021)	(0.017)	(0.017)
Age	0.012***	0.007**	0.008**
	(0.004)	(0.003)	(0.003)
Age^2	-8.05×10^{-5}*	-6.76×10^{-5}*	-7.48×10^{-5}*
	(4.28×10^{-5})	(3.37×10^{-5})	(3.73×10^{-5})
Sex = Male	-0.018	0.021**	0.013
	(0.014)	(0.010)	(0.010)
Marital status = Married	-0.120***	0.012	0.023***
	(0.025)	(0.021)	(0.006)
Marital status = Never married	-0.086***	0.002	0.007
	(0.024)	(0.021)	(0.010)
Marital status = Separated	-0.125***	-0.006	0.008
	(0.023)	(0.023)	(0.013)
Country fixed effects		✓	
Country-Sector fixed effects			✓
No. of countries	57	57	55
Observations	1,162,815	1,162,815	1,125,761
Dependent variable mean	0.714	0.714	0.712
R^2	0.087	0.184	0.195
Within R^2		0.039	0.042

Notes. As in Table D.I, but high-skill workers are defined as those with at least thirteen years of schooling. Linear probability models for working in a large firm, with every estimated coefficient shown — the constant, the high-skill indicator and the controls (age, age^2, sex and marital status, the latter two entering as sets of category indicators) — and country and country-by-sector fixed effects as indicated; the fixed effects absorb the constant. Estimates are weighted by the rescaled survey weights and standard errors clustered at the country level are in parentheses. Large firms have at least ten employees. Sample: wage employees aged 25–55 with non-missing education and firm-size information, one survey per country (the survey year closest to 2015). $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

C.3. Alternative definition of large firms: at least 50 employees

Throughout this subsection a firm counts as large only with at least 50 employees, instead of ten or more. Firms with ten to fifty employees are grouped with small firms, so no worker is dropped by the reclassification and the definition of high-skill (at least ten years of schooling) is unchanged — only the firm-size label moves. Palestine is excluded throughout: no wage employee there is recorded in a firm above fifty employees, so its large-firm statistics have no denominator. Firm size enters no wage regression, so the skill-premium exhibits — Figure III and Table C.I — are unchanged by this exercise and are not reproduced here.

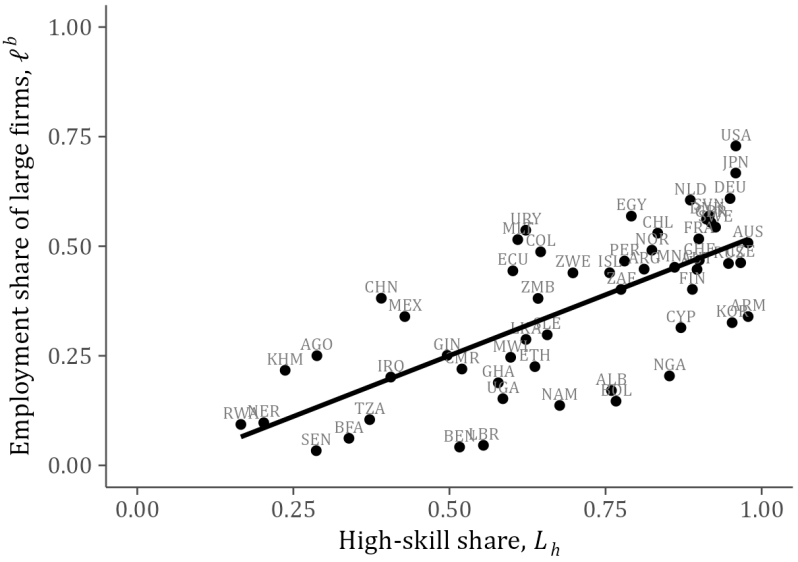
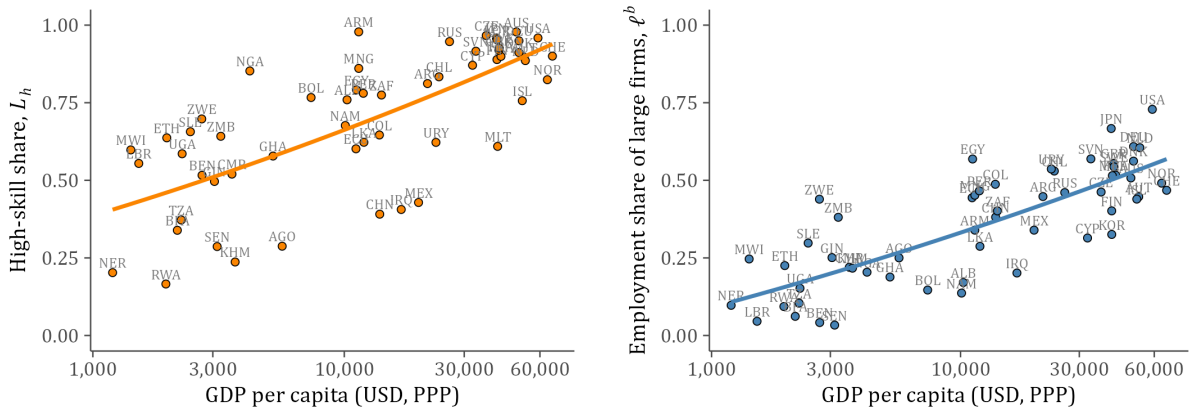


FIGURE C.VI
Skills and employment in large firms, alternative definition of large firms

Notes. As in Figure I, but a firm counts as large only if it has at least 50 employees, instead of ten or more; firms with ten to fifty employees are grouped with small firms, so no worker is dropped by the reclassification. Palestine is excluded: no wage employee there is recorded in a firm above fifty employees, so its large-firm statistics have no denominator. Each marker is one country, labeled with its ISO3 country code, and the solid line is a straight-line least-squares fit across countries. High-skill denotes at least ten years of schooling. Sample: wage employees aged 25–55 with non-missing education and firm-size information, one survey per country (the survey year closest to 2015); survey weights are rescaled within each country–year cell. The surveys used for each country are listed in Table A.I.



(A) Share of skilled workers.

(B) Share of employment in large firms.

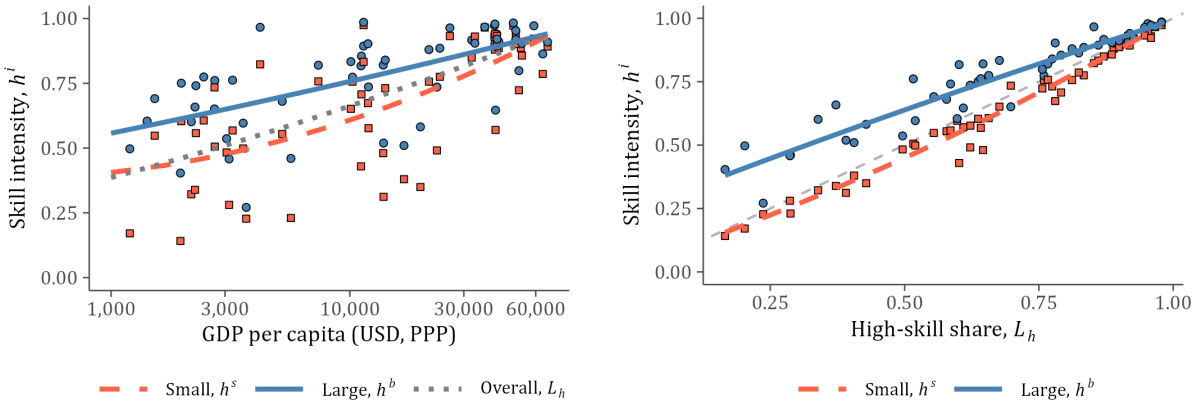
FIGURE C.VII
Skill endowments and employment by firm size across countries, alternative definition of large firms

Notes. As in Figure II, but a firm counts as large only if it has at least 50 employees, instead of ten or more; firms with ten to fifty employees are grouped with small firms, so no worker is dropped by the reclassification. Palestine is excluded: no wage employee there is recorded in a firm above fifty employees, so its large-firm statistics have no denominator. In both panels each marker is one country, labeled with its ISO3 country code. Panel (a) plots the high-skill share and Panel (b) the share of wage employment in large firms, both against GDP per capita (PPP, log scale). The solid line in each panel is a quadratic (in log GDP per capita) least-squares fit across countries. Panel (a) carries no firm-size information and reproduces Figure IIa up to the one dropped country; it is shown only to keep the figure self-contained. High-skill denotes at least ten years of schooling. Sample: wage employees aged 25–55 with non-missing education and firm-size information, one survey per country (the survey year closest to 2015); survey weights are rescaled within each country–year cell. The surveys used for each country are listed in Table A.I. GDP per capita is PPP-adjusted [Feenstra et al., 2015].

TABLE C.VI
Skills, the organization of production and the skill premium by country income group, alternative definition of large firms

	Country income			
	Low-income	Middle-income	High-income	US
Employment composition				
Years of schooling	9.0	11.4	13.6	14.0
Share of high-skilled workers L_h	50.0%	70.2%	89.3%	95.9%
Share of employment in large firms ℓ^b	18.5%	39.1%	51.2%	72.9%
Skill premium				
Unconditional: mean wage ratio $\omega = w_h/w_l$	1.88	1.69	1.39	1.79
Conditional: returns to skill (log points)	0.71	0.45	0.32	0.57
Probability of working for a large firm by skill group				
High skill L_h^b/L_h	23.2%	45.2%	52.1%	73.9%
Low skill L_l^b/L_l	14.3%	25.8%	43.0%	49.1%
Skill intensity				
Large firms h^b	63.4%	79.6%	90.9%	97.2%
Small firms h^s	47.7%	63.6%	87.7%	92.2%
Ratio (large / small)† h^b/h^s	1.50	1.32	1.04	1.05
Number of countries	20	17	19	1

Notes. As in Table I, but a firm counts as large only if it has at least 50 employees, instead of ten or more. Firms with ten to fifty employees are grouped with small firms, so no worker is dropped by the reclassification and the definition of high-skill (at least ten years of schooling) is unchanged — only the firm-size label moves. Palestine is excluded from this table: no wage employee there is recorded in a firm above fifty employees, so its large-firm statistics have no denominator. Each column is the unweighted average of the country-level statistic within an income group (LIC: GDP per capita below $10,000 PPP; MIC: $10,000–$30,000; HIC: above $30,000), with the United States shown separately. See Table I for the definitions of the individual rows. The surveys used for each country are listed in Table A.I.



(A) The skill intensity of large and small firms vs GDP per capita

(B) The skill intensity of large and small firms and the aggregate supply of skills

FIGURE C.VIII
Skill and firm size, alternative definition of large firms

Notes. As in Figure IV, but a firm counts as large only if it has at least 50 employees, instead of ten or more; firms with ten to fifty employees are grouped with small firms, and Palestine is excluded. Each marker is one country. Panel (a) plots the high-skill share of small firms, h^s, and of large firms, h^b, against GDP per capita; the lines are least-squares fits on a quadratic in log GDP per capita, and the dotted line is the same fit for the economy-wide share L_h. Panel (b) plots the same shares against the country-wide high-skill share L_h, with least-squares fits on a quadratic in L_h; the dashed line is the 45-degree line, which the economy-wide share follows by construction. High-skill denotes at least ten years of schooling. Sample: wage employees aged 25–55 with non-missing education and firm-size information, one survey per country (the survey year closest to 2015); survey weights are rescaled within each country–year cell. The surveys used for each country are listed in Table A.I. GDP per capita is PPP-adjusted [Feenstra et al., 2015].

TABLE C.VII
The skill gap in large-firm employment, from the worker and the firm side, alternative definition of large firms

	Working in large firm			Skill intensity h^i		
	(1)	(2)	(3)	(4)	(5)	(6)
Unit of observation	Worker			Country × firm size		
1{High-skill}	0.130***	0.137***	0.130***			
	(0.016)	(0.016)	(0.016)			
1{High-skill} × log GDP p.c.		0.021**				
		(0.010)				
1{High-skill} × high-skill share			0.0007			
			(0.052)			
1{Large firm}				0.113***	0.113***	0.113***
				(0.014)	(0.013)	(0.011)
1{Large firm} × log GDP p.c.					-0.039***	
					(0.010)	
1{Large firm} × high-skill share						-0.299***
						(0.045)
Controls	✓	✓	✓			
Country fixed effects	✓	✓	✓	✓	✓	✓
No. of countries	56	56	56	55	55	55
Observations	1,151,004	1,151,004	1,151,004	110	110	110
R^2	0.150	0.150	0.150	0.942	0.955	0.967
Within R^2	0.015	0.015	0.015	0.542	0.641	0.738

Notes. As in Table III, but a firm counts as large only if it has at least 50 employees, instead of ten or more; Palestine is excluded. Columns (1)–(3) are worker-level linear probability models for working in a large firm, with a high-skill indicator and controls for age, age^2, sex and marital status; columns (4)–(6) are country–firm-size cell regressions of the cell's high-skill share on a large-firm indicator. Country fixed effects throughout. Columns (2) and (5) interact the gap with log GDP per capita and columns (3) and (6) with the country's high-skill share, both entered centered on their means. See Table III for the full description of the two blocks and why their magnitudes are not comparable. Standard errors are clustered by country. High-skill denotes at least ten years of schooling. Sample: wage employees aged 25–55 with non-missing education and firm-size information, one survey per country (the survey year closest to 2015). $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE C.VIII
Skills and the probability of working in a large firm, alternative definition of large firms

	Working in large firm		
	(1)	(2)	(3)
Constant	0.015		
	(0.072)		
1{High-skill}	0.225***	0.130***	0.148***
	(0.020)	(0.016)	(0.014)
Age	0.011***	0.006**	0.005**
	(0.003)	(0.002)	(0.002)
Age^2	-8.35×10^{-5}**	-5.31×10^{-5}*	-4.29×10^{-5}
	(3.58×10^{-5})	(2.7×10^{-5})	(2.68×10^{-5})
Sex = Male	-0.004	0.028***	0.018*
	(0.012)	(0.009)	(0.010)
Marital status = Married	-0.110***	0.048**	0.030***
	(0.031)	(0.019)	(0.004)
Marital status = Never married	-0.080**	0.041*	0.016*
	(0.034)	(0.021)	(0.009)
Marital status = Separated	-0.105***	0.038*	0.021***
	(0.031)	(0.019)	(0.007)
Country fixed effects		✓	
Country-Sector fixed effects			✓
No. of countries	56	56	54
Observations	1,151,004	1,151,004	1,113,950
Dependent variable mean	0.482	0.482	0.482
R^2	0.054	0.150	0.169
Within R^2		0.015	0.018

Notes. As in Table D.I, but a firm counts as large only if it has at least 50 employees, instead of ten or more; Palestine is excluded. Linear probability models for working in a large firm, with every estimated coefficient shown — the constant, the high-skill indicator and the controls (age, age^2, sex and marital status) — and country and country-by-sector fixed effects as indicated; the fixed effects absorb the constant. Estimates are weighted by the rescaled survey weights and standard errors clustered at the country level are in parentheses. High-skill denotes at least ten years of schooling. Sample: wage employees aged 25–55 with non-missing education and firm-size information, one survey per country (the survey year closest to 2015). $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

C.4. Including the self-employed

The baseline sample covers wage employees only. Throughout this subsection it additionally includes the self-employed, who are treated as employed in small firms. This matters most for low-income countries, where self-employment is widespread, and it strengthens rather than weakens the patterns: the gradients of both the skill share and the large-firm employment share with GDP per capita become steeper, and the correlation between the two rises to 0.89. Wages are observed only for wage employees, so the skill-premium exhibits — Figure III and Table C.I — are unchanged by this exercise and are not reproduced here.

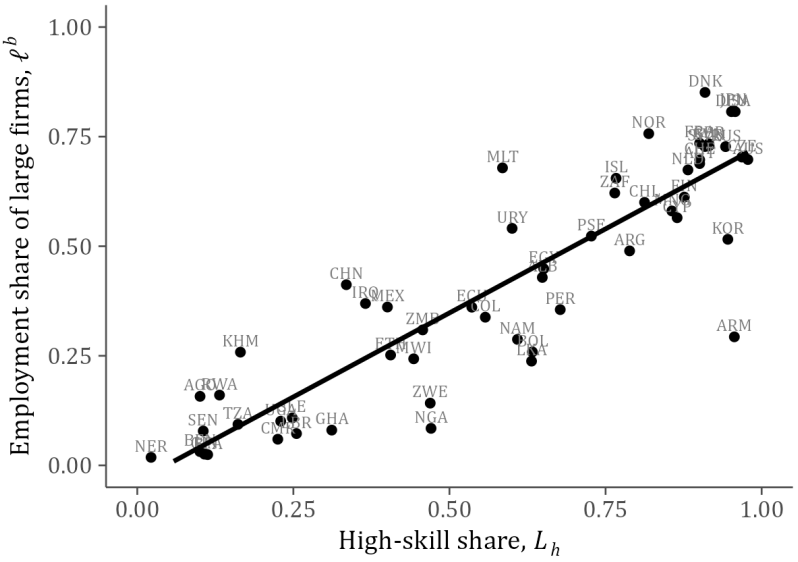
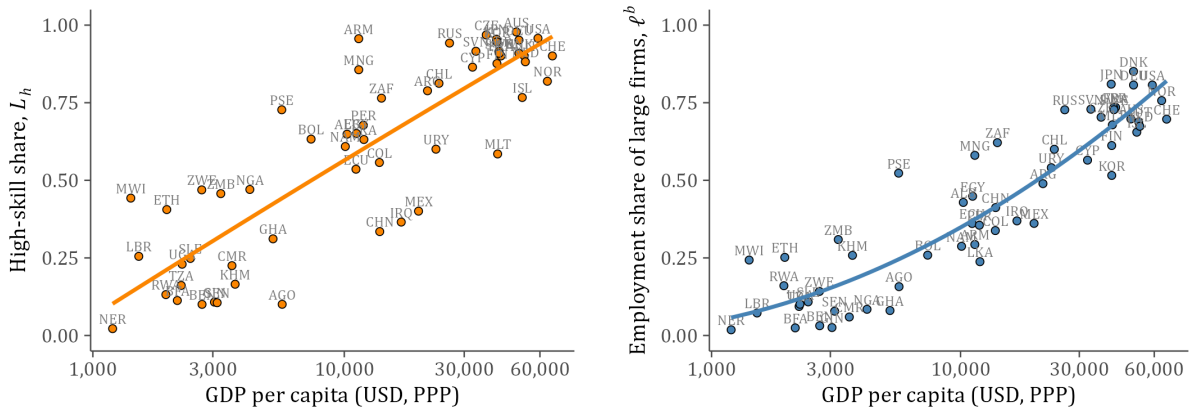


FIGURE C.IX
Skills and employment in large firms, including the self-employed

Notes. As in Figure I, but the sample additionally includes self-employed workers, who are treated as employed in small firms. Each marker is one country, labeled with its ISO3 country code, and the solid line is a straight-line least-squares fit across countries. High-skill denotes at least ten years of schooling and large firms have at least ten employees. Sample: wage and self-employed workers aged 25–55 with non-missing education and firm-size information, one survey per country (the survey year closest to 2015); survey weights are rescaled within each country–year cell. The surveys used for each country are listed in Table A.I.



(A) Share of skilled workers.

(B) Share of employment in large firms.

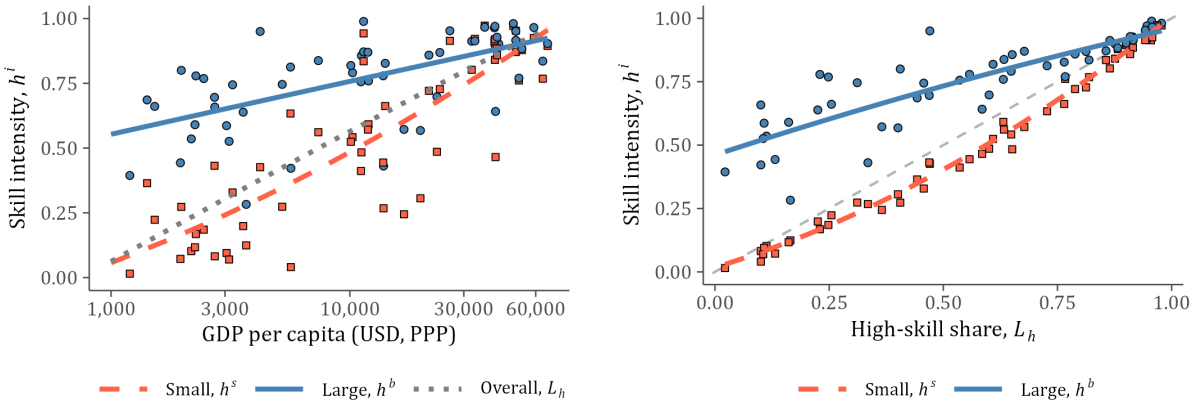
FIGURE C.X
Skill endowments and employment by firm size across countries, including the self-employed

Notes. As in Figure II, but the sample additionally includes self-employed workers, who are treated as employed in small firms. In both panels each marker is one country, labeled with its ISO3 country code. Panel (a) plots the high-skill share and Panel (b) the share of wage employment in large firms, both against GDP per capita (PPP, log scale). The solid line in each panel is a quadratic (in log GDP per capita) least-squares fit across countries. High-skill denotes at least ten years of schooling and large firms have at least ten employees. Sample: wage and self-employed workers aged 25–55 with non-missing education and firm-size information, one survey per country (the survey year closest to 2015); survey weights are rescaled within each country–year cell. The surveys used for each country are listed in Table A.I. GDP per capita is PPP-adjusted [Feenstra et al., 2015].

TABLE C.IX
Skills, the organization of production and the skill premium by country income group, including the self-employed

	Country income			
	Low-income	Middle-income	High-income	US
Employment composition				
Years of schooling	6.1	10.8	13.6	14.0
Share of high-skilled workers L_h	28.0%	65.5%	89.0%	95.8%
Share of employment in large firms ℓ^b	14.7%	43.8%	70.8%	80.7%
Share of self-employed	71.2%	34.3%	12.7%	8.6%
Probability of working for a large firm by skill group				
High skill L_h^b / L_h	33.7%	52.6%	71.8%	81.4%
Low skill L_l^b / L_l	7.8%	28.8%	63.4%	65.8%
Skill intensity				
Large firms h^b	64.6%	78.0%	90.1%	96.5%
Small firms h^s	22.8%	56.9%	86.3%	92.5%
Ratio (large / small)† h^b/h^s	5.15	1.48	1.05	1.04
Number of countries	21	17	19	1

Notes. As in Table I, but the sample additionally includes self-employed workers, who are treated as employed in small firms. The *Employment composition* block therefore carries an extra row giving the share of the self-employed in employment. The skill-premium rows are omitted because wages are observed only for wage employees. Each column is the unweighted average of the country-level statistic within an income group (LIC: below \$10,000 PPP; MIC: \$10,000–\$30,000; HIC: above \$30,000), with the United States shown separately. See Table I for the definitions of the individual rows. The surveys used for each country are listed in Table A.I.



(A) The skill intensity of large and small firms vs GDP per capita

(B) The skill intensity of large and small firms and the aggregate supply of skills

FIGURE C.XI
Skill and firm size, including the self-employed

Notes. As in Figure IV, but the sample additionally includes self-employed workers, who are treated as employed in small firms. Each marker is one country. Panel (a) plots the high-skill share of small firms, h^s, and of large firms, h^b, against GDP per capita; the lines are least-squares fits on a quadratic in log GDP per capita, and the dotted line is the same fit for the economy-wide share L_h. Panel (b) plots the same shares against the country-wide high-skill share L_h, with least-squares fits on a quadratic in L_h; the dashed line is the 45-degree line, which the economy-wide share follows by construction. High-skill denotes at least ten years of schooling and large firms have at least ten employees. Sample: wage and self-employed workers aged 25–55 with non-missing education and firm-size information, one survey per country (the survey year closest to 2015); survey weights are rescaled within each country–year cell. The surveys used for each country are listed in Table A.I. GDP per capita is PPP-adjusted [Feenstra et al., 2015].

TABLE C.X
The skill gap in large-firm employment, from the worker and the firm side, including the self-employed

	Working in large firm			Skill intensity h^i		
	(1)	(2)	(3)	(4)	(5)	(6)
Unit of observation	Worker			Country × firm size		
1{High-skill}	0.212***	0.202***	0.202***			
	(0.015)	(0.014)	(0.014)			
1{High-skill} × log GDP p.c.		-0.046***				
		(0.012)				
1{High-skill} × high-skill share			-0.204***			
			(0.054)			
1{Large firm}				0.230***	0.230***	0.230***
				(0.025)	(0.013)	(0.013)
1{Large firm} × log GDP p.c.					-0.132***	
					(0.011)	
1{Large firm} × high-skill share						-0.534***
						(0.044)
Controls	✓	✓	✓			
Country fixed effects	✓	✓	✓	✓	✓	✓
No. of countries	57	57	57	57	57	57
Observations	1,700,945	1,700,945	1,700,945	114	114	114
R^2	0.309	0.311	0.311	0.886	0.970	0.971
Within R^2	0.042	0.044	0.044	0.602	0.894	0.900

Notes. As in Table III, but the sample additionally includes self-employed workers, who are treated as employed in small firms. Columns (1)–(3) are worker-level linear probability models for working in a large firm, with a high-skill indicator and controls for age, age^2, sex and marital status; columns (4)–(6) are country–firm-size cell regressions of the cell's high-skill share on a large-firm indicator. Country fixed effects throughout. Columns (2) and (5) interact the gap with log GDP per capita and columns (3) and (6) with the country's high-skill share, both entered centered on their means. See Table III for the full description of the two blocks and why their magnitudes are not comparable. Standard errors are clustered by country. High-skill denotes at least ten years of schooling and large firms have at least ten employees. Sample: wage and self-employed workers aged 25–55 with non-missing education and firm-size information, one survey per country (the survey year closest to 2015). $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE C.XI
Skills and the probability of working in a large firm, including the self-employed

	Working in large firm		
	(1)	(2)	(3)
Constant	0.118*		
	(0.070)		
1{High-skill}	0.411***	0.212***	0.197***
	(0.024)	(0.015)	(0.014)
Age	0.007**	-0.001	-0.001
	(0.003)	(0.002)	(0.002)
Age^2	-6.83×10^{-5}**	-7.06×10^{-6}	3.03×10^{-6}
	(3.2×10^{-5})	(2.27×10^{-5})	(2.39×10^{-5})
Sex = Male	-0.003	0.0007	-0.003
	(0.010)	(0.009)	(0.009)
Marital status = Married	-0.147***	-0.024	-0.005
	(0.040)	(0.030)	(0.005)
Marital status = Never married	-0.044	-0.003	0.006
	(0.040)	(0.030)	(0.009)
Marital status = Separated	-0.125***	-0.019	-0.013*
	(0.040)	(0.030)	(0.007)
Country fixed effects		✓	
Country-Sector fixed effects			✓
No. of countries	57	57	55
Observations	1,700,945	1,700,945	1,636,509
Dependent variable mean	0.488	0.488	0.490
R^2	0.187	0.309	0.325
Within R^2		0.042	0.035

Notes. As in Table D.I, but the sample additionally includes self-employed workers, who are treated as employed in small firms. Linear probability models for working in a large firm, with every estimated coefficient shown — the constant, the high-skill indicator and the controls (age, age^2, sex and marital status) — and country and country-by-sector fixed effects as indicated; the fixed effects absorb the constant. Estimates are weighted by the rescaled survey weights and standard errors clustered at the country level are in parentheses. High-skill denotes at least ten years of schooling and large firms have at least ten employees. Sample: wage and self-employed workers aged 25–55 with non-missing education and firm-size information, one survey per country (the survey year closest to 2015). $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

C.5. Wider age window: workers aged 16–64

The baseline restricts attention to workers aged 25–55, the range over which schooling is essentially complete and retirement is not yet a margin. Widening it to 16–64 changes the wage sample and hence the skill premium, which Table C.XII reports. It leaves the descriptive exhibits of Section 2 qualitatively unchanged, and those are not reproduced here.

TABLE C.XII
The skill premium across countries, alternative measures — wider age window

	Country income				
	Low-income	Middle-income	High-income	US	Elasticity
Ratio of mean wages					
No restriction	2.33	1.78	1.52	2.41	-0.13^{**}
Trimmed tails (1%)	1.85	1.65	1.43	1.72	-0.08^{***}
Dominant pay frequency	2.54	1.83	1.52	2.41	-0.17^{***}
Dominant frequency + trimmed	2.07	1.68	1.43	1.72	-0.12^{***}
Ratio of median wages					
No restriction	2.23	1.52	1.41	1.67	-0.16^{***}
Trimmed tails (1%)	2.15	1.49	1.38	1.65	-0.15^{***}
Dominant pay frequency	2.43	1.53	1.41	1.67	-0.18^{***}
Dominant frequency + trimmed	2.34	1.50	1.38	1.65	-0.18^{***}
Mincer return on high skill (log points)					
No restriction	0.73	0.47	0.38	0.56	-0.14^{***}
Trimmed tails (1%)	0.67	0.43	0.34	0.53	-0.13^{***}
Number of countries	20	17	19	1	

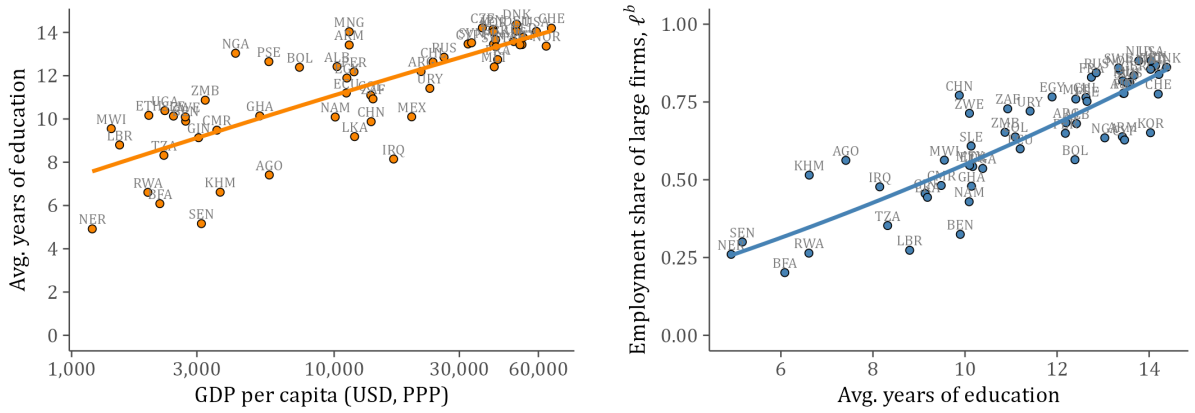
Notes. As in Table C.I, but the sample covers wage employees aged 16–64 rather than 25–55. Everything else — the skill definition, the four sample restrictions, the Mincer specification and the elasticity column — is as in Table C.I, where the individual rows and columns are defined. This is the only other exercise in the paper that changes the wage sample: the alternative large-firm threshold and the inclusion of the self-employed leave it untouched, since firm size enters no wage regression, so their premia are identical to those in Table C.I. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

D. Additional evidence

This appendix collects evidence that supports the four facts but is not a robustness variant of any main-text exhibit: average years of schooling as a continuous alternative to the skill indicator, the same relationships estimated within broad sectors, the partial-regression scatter behind the within-country panel regression of Table II, and the worker-level evidence behind Fact 3.

D.1. Average years of schooling as the skill measure

The paper measures a country's skill endowment by the share of workers above a schooling threshold. Figure D.I replaces that indicator with average years of schooling in the labor force, a continuous measure that uses the whole schooling distribution and does not depend on where a threshold is drawn. Both relationships of Figure II survive: schooling rises with GDP per capita, and large-firm employment rises with schooling, with a correlation of 0.87.



(A) Years of education.

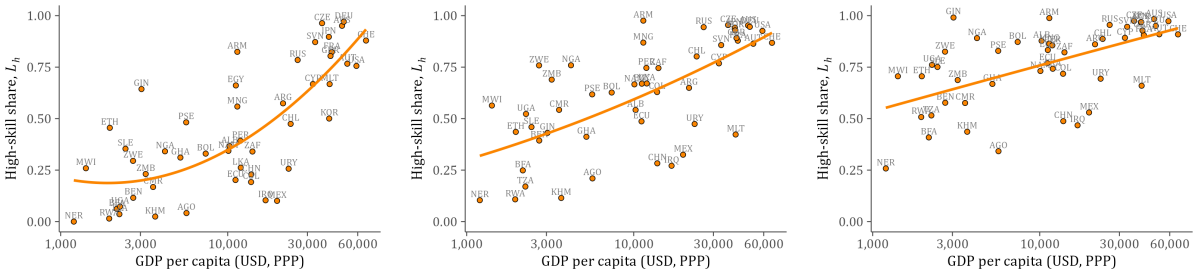
(B) Education and employment in large firms.

FIGURE D.I
Average years of schooling and the size structure of employment across countries

Notes. Each marker is one country, labeled with its ISO3 country code. Panel (a) plots the average years of schooling of wage employees against GDP per capita (PPP, log scale); Panel (b) plots the share of wage employment in large firms against average years of schooling. The solid line is a quadratic least-squares fit across countries — in log GDP per capita in Panel (a) and in average years of schooling in Panel (b). Large firms have at least ten employees. Sample: wage employees aged 25–55 with non-missing education and firm-size information, one survey per country (the survey year closest to 2015); survey weights are rescaled within each country–year cell. The surveys used for each country are listed in Table A.I. GDP per capita is PPP-adjusted [Feenstra et al., 2015].

D.2. The relationship within broad sectors

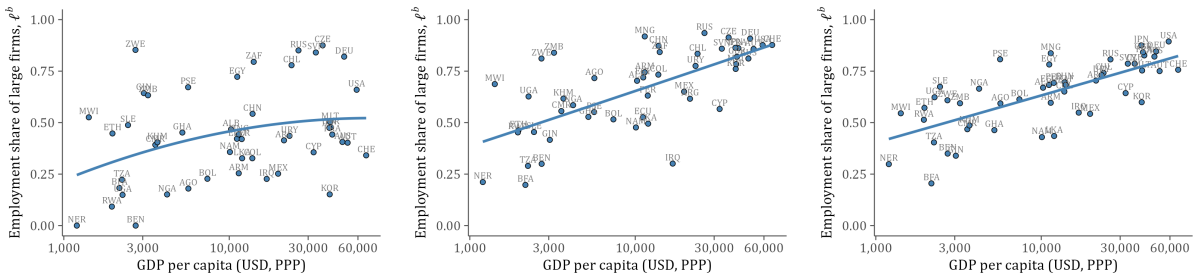
Low-income countries have more employment in agriculture, and agricultural production occurs in small firms, so the cross-country pattern of Fact 1 could in principle be a composition effect. The three figures in this subsection show that it is not. The share of employment in large firms is indeed lower in agriculture than in industry and services, but it is lower in low-income countries *within* each of the three broad sectors, and the positive relationship between skills and large-firm employment holds within sector as well.



(A) Agriculture (B) Industry (C) Services

FIGURE D.II
Skill endowments by sector

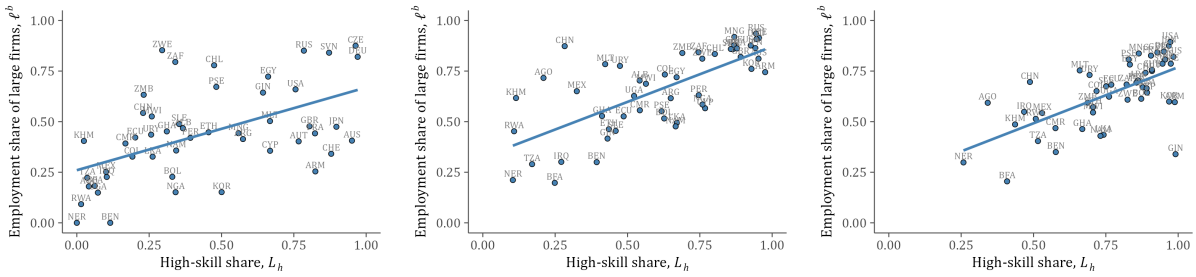
Notes. Each panel reproduces Panel (a) of Figure II — the cross-country relationship between the high-skill share of wage employees and GDP per capita (PPP, log scale) — within one broad sector. Each marker is a country–sector cell, labeled with the ISO3 country code, and the solid line in each panel is a quadratic (in log GDP per capita) least-squares fit across countries. High-skill denotes at least ten years of schooling. Workers are assigned to their main sector of activity and those with no recorded sector are dropped. Sample: wage employees aged 25–55 with non-missing education and firm-size information, one survey per country (the survey year closest to 2015); survey weights are rescaled within each country–year cell. The surveys used for each country are listed in Table A.I. GDP per capita is PPP-adjusted [Feenstra et al., 2015].



(A) Agriculture (B) Industry (C) Services

FIGURE D.III
The share of employment in large firms by sector

Notes. As in Figure D.II, but the vertical axis is the within-sector share of wage employment in large firms (ten or more employees), plotted against GDP per capita (PPP, log scale), so each panel reproduces Panel (b) of Figure II within one broad sector. Each marker is a country–sector cell and the solid line in each panel is a quadratic (in log GDP per capita) least-squares fit across countries. Sample: wage employees aged 25–55 with non-missing education and firm-size information, one survey per country (the survey year closest to 2015); survey weights are rescaled within each country–year cell. The surveys used for each country are listed in Table A.I. GDP per capita is PPP-adjusted [Feenstra et al., 2015].



(A) Agriculture (B) Industry (C) Services

FIGURE D.IV
Skills and the share of employment in large firms by sector

Notes. As in Figure D.II, but plotting the within-sector share of wage employment in large firms (ten or more employees) against the within-sector share of high-skill workers (at least ten years of schooling), so each panel reproduces Figure I within one broad sector. Each marker is a country–sector cell and the solid line in each panel is a straight-line least-squares fit across countries. Sample: wage employees aged 25–55 with non-missing education and firm-size information, one survey per country (the survey year closest to 2015); survey weights are rescaled within each country–year cell. The surveys used for each country are listed in Table A.I.

D.3. The within-country relationship over time

Table II reports the within-country panel regression of the large-firm employment share on the high-skill share. Figure D.V shows that same estimate as a picture. Each marker is a country–year observation with both variables purged of country and year fixed effects, so the slope of the cloud is the regression coefficient and the spread around it is what the standard error is computed from. The two shares move together within countries: when a country's skill share is above its own average, its large-firm employment share tends to be above its own average as well.

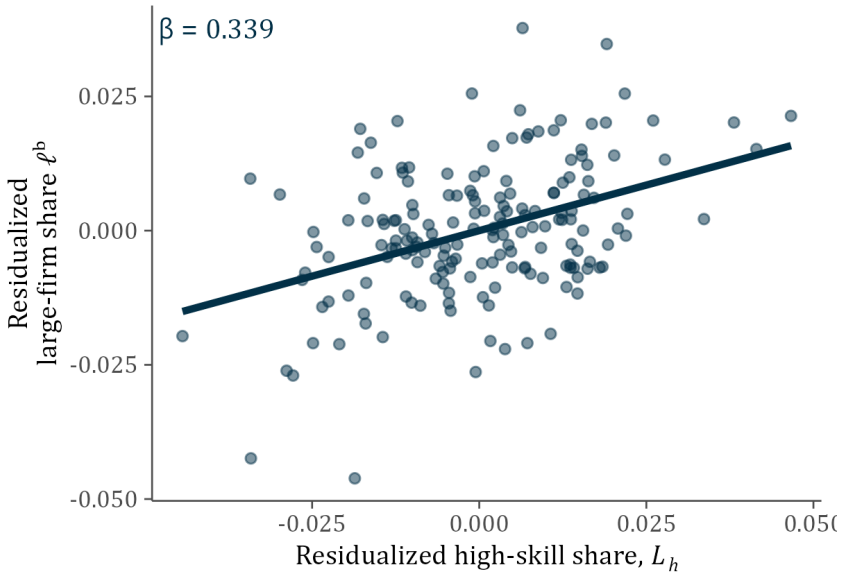


FIGURE D.V
Changes in skills and in large-firm employment over time

Notes. A partial-regression (Frisch–Waugh–Lovell) scatter of the regression in Table II: each marker is a country–year observation after both the large-firm employment share (vertical axis) and the high-skill share (horizontal axis) have been purged of country and year fixed effects. The sample is a balanced panel of 16 countries observed in every year from 2007 to 2017 (176 country–year observations), built from annually repeated household and labour force surveys. High-skill denotes at least ten years of schooling and large firms have at least ten employees. Sample: wage employees aged 25–55 with non-missing education and firm-size information; survey weights are rescaled to one within each country–year cell, so that each country counts the same independent of its population or survey size. The panel countries are listed in Table A.I.

D.4. Skills and large-firm employment at the worker level

The country-level patterns of Section 2 are mirrored at the individual level. Figure D.VI shows the share of wage workers employed in a large firm by years of education, pooled across our 1,162,815 individual-level observations, and Table D.I reports the underlying linear probability models with every estimated coefficient shown. Columns (1)–(3) of Table III in the main text are the country fixed-effects and interaction columns of the same specification. The link between skills and firm size thus holds worker by worker within countries and sectors, and not only in the aggregate.

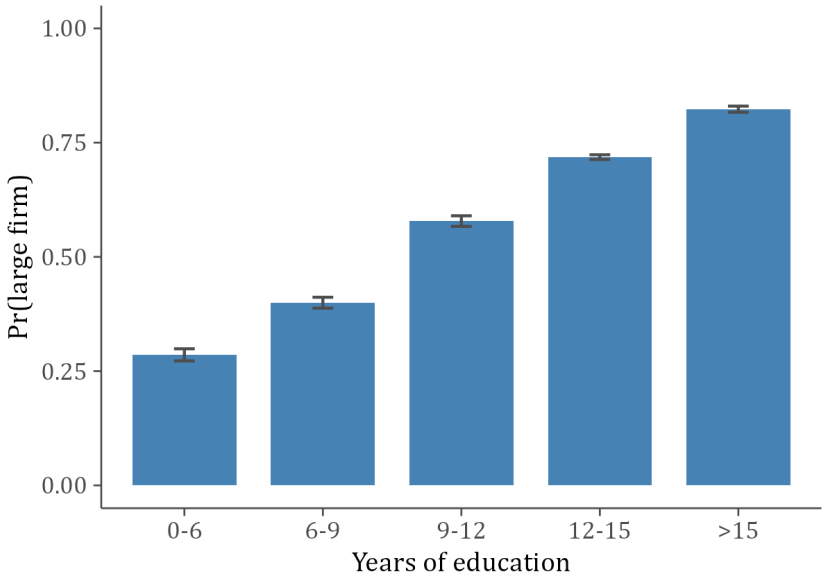


FIGURE D.VI
Share working in a large firm, by years of education

Notes. The figure shows the share of wage employees working in a large firm within bins of completed years of schooling, pooled across countries. Bars are weighted shares and whiskers are 95% confidence intervals. Large firms have at least ten employees. Sample: wage employees aged 25–55 with non-missing education and firm-size information, one survey per country (the survey year closest to 2015); survey weights are rescaled within each country–year cell. The surveys used for each country are listed in Table A.I.

TABLE D.I
Skills and the probability of working in a large firm, full controls

	Working in large firm		
	(1)	(2)	(3)
Constant	0.167**		
	(0.081)		
1{High-skill}	0.326***	0.231***	0.241***
	(0.024)	(0.018)	(0.015)
Age	0.013***	0.009***	0.010***
	(0.004)	(0.003)	(0.003)
Age^2	-0.0001**	-9.2×10^{-5}***	-0.0001***
	(4.09×10^{-5})	(3.41×10^{-5})	(3.66×10^{-5})
Sex = Male	-0.017	0.017	0.009
	(0.014)	(0.010)	(0.010)
Marital status = Married	-0.085***	0.010	0.016***
	(0.027)	(0.015)	(0.006)
Marital status = Never married	-0.062**	-0.0010	0.0004
	(0.026)	(0.016)	(0.010)
Marital status = Separated	-0.094***	-0.010	-0.004
	(0.025)	(0.018)	(0.013)
Country fixed effects		✓	
Country-Sector fixed effects			✓
No. of countries	57	57	55
Observations	1,162,815	1,162,815	1,125,761
Dependent variable mean	0.714	0.714	0.712
R^2	0.108	0.189	0.197
Within R^2		0.044	0.045

Notes. This table shows linear probability models in which the outcome is one if a wage employee works in a large firm, with every estimated coefficient reported: the constant, the high-skill indicator and the controls (age, age^2, sex and marital status, the latter two entering as sets of category indicators). Columns (2) and (3) add country and country-by-sector fixed effects respectively, which absorb the constant; column (3) drops workers with no recorded sector, which is why it covers two countries fewer. Estimates are weighted by the rescaled survey weights and standard errors clustered at the country level are in parentheses. High-skill denotes at least ten years of schooling and large firms have at least ten employees. Sample: wage employees aged 25–55 with non-missing education and firm-size information, one survey per country (the survey year closest to 2015). $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

E. Proofs and derivations

E.1. Derivation of Lemma 5

Optimal headcount employment is[47]

$$L_h^i(z) = \left(\frac{\bar{w}^i \mu^i}{w_h}\right)^\rho \bar{L}^i(z) \qquad \text{and} \quad L_l^i(z) = \left(\frac{\bar{w}^i (1-\mu^i)}{w_l}\right)^\rho \bar{L}^i(z) \tag{14}$$

for high- and low-skill workers, respectively.

Headcount employment of a modern relative to a traditional firm at a given productivity z is a weighted average of skill-specific headcount ratios:

$$\frac{L^m(z)}{L^t(z)} = \frac{L_h^m(z) + L_l^m(z)}{L_h^t(z) + L_l^t(z)} = \frac{L_h^m(z)}{L_h^t(z)} \frac{L_h^t(z)}{L_h^t(z) + L_l^t(z)} + \frac{L_l^m(z)}{L_l^t(z)} \frac{L_l^t(z)}{L_h^t(z) + L_l^t(z)}. \tag{15}$$

Using (14) and log-differentiating implies that the high-skill employment ratio $L_h^m(z)/L_h^t(z)$ has an elasticity with respect to the skill premium ω of

$$\frac{d \ln(L_h^m(z)/L_h^t(z))}{d \ln \omega} = \rho(\Lambda_h^m - \Lambda_h^t) + \frac{\Lambda_h^t}{1-\gamma^t} - \frac{\Lambda_h^m}{1-\gamma^m}. \tag{16}$$

The last two terms in the equation carry over from the result on optimal composite employment above. They again reflect the greater effect of an increase in the skill premium on modern firms' unit cost, amplified by their greater returns to scale. The first term captures a countervailing substitution effect.[48]

With our assumptions on the parameters, the overall elasticity in (16) is strictly negative unless ρ exceeds a threshold that is itself strictly greater than $1/(1-\gamma^m)$.[49] For typical values of γ^m of around 0.85, $1/(1-\gamma^m)$ is about 6.7, which is far above any reasonable estimate of ρ. The precise threshold exceeds 10 in all country income groups.

Thus, a higher skill premium reduces relative high-skill employment in modern firms for any reasonable ρ: the increase in their relative unit cost reduces their relative employment, further amplified by their greater scale parameter. The same holds for low-skill employment. By (3), $L_l^i(z) = L_h^i(z) \left(\frac{1-\mu^i}{\mu^i}\right)^\rho \omega^\rho$, so the low-skill employment

[47] This follows from combining the cost-minimization first-order condition with the definition of $\bar{L}$.

[48] From (14), an increase in $\bar{w}^i/w_j^i$ increases $L_j^i(z)/\bar{L}^i(z)$, with elasticity ρ. $\bar{w}^i/w_j^i$ increases for modern relative to traditional firms, so both $L_l^m(z)/L_l^t(z)$ and $L_h^m(z)/L_h^t(z)$ fall by less than the composite ratio $\bar{L}^m(z)/\bar{L}^t(z)$. However, this effect is weak at plausible parameters.

[49] The threshold is $\frac{\Lambda_h^m/(1-\gamma^m) - \Lambda_h^t/(1-\gamma^t)}{\Lambda_h^m - \Lambda_h^t}$. If $\gamma^t = \gamma^m$, it is equal to $1/(1-\gamma^m)$ and is smaller than that if $\gamma^t < \gamma^m$.

ratio satisfies

$$\frac{L_l^m(z)}{L_l^t(z)} = \frac{L_h^m(z)}{L_h^t(z)} \left(\frac{\mu^t(1-\mu^m)}{\mu^m(1-\mu^t)} \right)^{\rho}.$$

The ω^ρ terms cancel, and the remaining factor depends only on the skill weights. The low-skill employment ratio therefore has the same elasticity with respect to ω as the high-skill ratio, and relative low-skill employment in modern firms also declines in the skill premium.

In addition, from (3), a greater skill premium prompts firms to reduce their skill intensity, implying an increase in the weight on $L_l^m(z)/L_l^t(z)$, the smaller term of the weighted sum in (15).[50] All these factors reduce relative headcount employment of modern firms, $L^m(z)/L^t(z)$.

Summarizing, all these factors reduce relative headcount employment of modern firms, which establishes Lemma 5.

E.2. Proof of Proposition 1

Total production headcount employment in modern-technology firms is

$$L^m = \frac{M}{1-G(z^x)} \int_{z^m}^{\infty} L^m(z)\, \mathrm{d}G(z).$$

Using equations (4), (6) and the Pareto distribution,

$$L^m = \frac{M}{1-G(z^x)} \phi^m \left(\frac{\gamma^m}{\bar{w}^m} \right)^{\frac{1}{1-\gamma^m}} Z^m,$$

where

$$Z^m \equiv \frac{\alpha z_{\min}^{\alpha}}{\alpha - \frac{1}{1-\gamma^m}} z^{m\,\frac{1}{1-\gamma^m} - \alpha}. \tag{17}$$

Similarly, total production headcount employment in traditional-technology firms is

$$L^t = \frac{M}{1-G(z^x)} \phi^t \left(\frac{\gamma^t}{\bar{w}^t} \right)^{\frac{1}{1-\gamma^t}} Z^t,$$

where

$$Z^t \equiv \alpha z_{\min}^{\alpha} \frac{z^{x\,\frac{1}{1-\gamma^t} - \alpha} - z^{m\,\frac{1}{1-\gamma^t} - \alpha}}{\alpha - \frac{1}{1-\gamma^t}}. \tag{18}$$

[50] Greater skill intensity of modern firms directly implies that the ratio is greater for high-skill workers: $L_h^m(z)/L_h^t(z) > L_l^m(z)/L_l^t(z)$.

The production employment share of modern firms ℓ^m is

$$\ell^m = \frac{L^m}{L^m + L^t} = \frac{1}{1 + \frac{L^t}{L^m}}, \tag{19}$$

where

$$\frac{L^t}{L^m} = \frac{\phi^t \left(\frac{\gamma^t}{\bar{w}^t}\right)^{\frac{1}{1-\gamma^t}} Z^t}{\phi^m \left(\frac{\gamma^m}{\bar{w}^m}\right)^{\frac{1}{1-\gamma^m}} Z^m}. \tag{20}$$

Then

$$\frac{d \ln \ell^m}{d \ln \omega} = -(1 - \ell^m) \frac{d \ln(L^t / L^m)}{d \ln \omega} \equiv -(1 - \ell^m) \eta_L, \tag{21}$$

where η_L denotes the elasticity of the ratio L^t / L^m with respect to ω. From Lemma 5 and (6), $\phi^m / \phi^t \cdot \bar{w}^{t\,\frac{1}{1-\gamma^t}} / \bar{w}^{m\,\frac{1}{1-\gamma^m}}$ declines in the skill premium if $\rho < 1/(1-\gamma^m)$, since the remaining factor of $L^m(z)/L^t(z)$ depends only on z and parameters. Applying this to equation (20) implies that $\eta_L > 0$ if $\rho < 1/(1-\gamma^m)$. Hence, ℓ^m falls in ω if $\rho < 1/(1-\gamma^m)$.

Write the share of total employment in modern firms, including overhead, as

$$\ell^m_{\text{tot}} = \frac{L^m + M^m c^f}{L^m + L^t + M c^f} = \ell^m (L^m + L^t) + m^m \cdot M c^f, \tag{22}$$

using the facts that the total labor force is unity, so that $L^m + L^t$ is the share of production employment in total employment, and Mc^f that of overhead employment. This is a weighted sum of modern firms' shares of production employment and overhead employment, with aggregate shares of production and overhead employment as weights. Because modern firms are larger than average, ℓ^m is larger than m^m. From above, ℓ^m and m^m both decline in the skill premium. In addition, a greater skill premium reduces z^x, hence average firm productivity and average size, thus raising the number of firms M. This implies a shift in the weights of the sum towards the second, smaller share m^m. All these factors imply that modern firms' share of total employment declines in the skill premium, by more than their share in production employment.[51] □

E.3. Proof of Proposition 2

A modern firm at productivity z employs $\phi^m (\gamma^m z / \bar{w}^m)^{1/(1-\gamma^m)}$ production workers plus c^f fixed-overhead workers, so its measured employment is $L^m(z) + c^f$. The

[51] This argument analyzes the share of modern employment in production plus overhead employment, excluding entry employment. This parallels our quantitative analysis, since entry employment cannot be explicitly observed in the data. Nevertheless, the argument goes through essentially unchanged when also including entry employment in total employment.

threshold z^b is the productivity at which measured employment equals b, i.e. at which production employment equals $b - c^f$:

$$z^b = \frac{((b - c^f)/\phi^m)^{1-\gamma^m} \bar{w}^m}{\gamma^m}. \tag{23}$$

Aggregate production employment in large firms, L^b, equals

$$L^b = \frac{M}{1 - G(z^x)} \phi^m \left(\frac{\gamma^m}{\bar{w}^m}\right)^{\frac{1}{1-\gamma^m}} \frac{\alpha z_{\min}^{\alpha}}{\alpha - \frac{1}{1-\gamma^m}} \left(z^b\right)^{\frac{1}{1-\gamma^m} - \alpha}. \tag{24}$$

The production employment share of large firms can be written as

$$\ell^b = \frac{L^b}{L^m + L^t} = \frac{L^b}{L^m} \cdot \ell^m = \left(\frac{z^b}{z^m}\right)^{\frac{1}{1-\gamma^m} - \alpha} \ell^m. \tag{25}$$

Since $z^b > z^m$ and $\alpha > 1/(1-\gamma^m)$, $\ell^b < \ell^m$. The share of production employment in large firms thus equals the production employment share of modern firms, times the share of production employment in large firms among modern firms. Define $\zeta = z^b/z^m$. Since $z^b > z^m$, $\zeta > 1$. Using the expressions for z^m and z^b:

$$\frac{1}{\zeta} = \frac{z^m}{z^b} = \left(\frac{\bar{w}^m}{\bar{w}^t}\right)^{\gamma^t \frac{1-\gamma^m}{\gamma^m - \gamma^t}} \left(\frac{\tilde{\gamma}^t}{\tilde{\gamma}^m}\right)^{1/(\frac{1}{1-\gamma^m} - \frac{1}{1-\gamma^t})} \gamma^m \left(\frac{\phi^m}{b - c^f}\right)^{1-\gamma^m} \tag{26}$$

Apart from the size threshold b (which enters through the third factor), ζ depends only on parameters and the skill premium, via $\bar{w}^m/\bar{w}^t$ and ϕ^m. Since both increase in the skill premium, the ratio z^m/z^b also increases in the skill premium. A higher skill premium shifts up the productivity threshold for choosing the modern technology (z^m) relative to the threshold for being large (z^b). As a result, employment among modern firms shifts into large modern firms, since higher z^m implies less employment in small modern firms with $z \in [z^m, z^b)$.

From equation (25),

$$\frac{d \ln \ell^b}{d \ln \omega} = \left(\frac{1}{1-\gamma^m} - \alpha\right) \frac{d \ln \zeta}{d \ln \omega} + \frac{d \ln \ell^m}{d \ln \omega} \equiv \eta_\zeta - (1 - \ell^m)\eta_L, \tag{27}$$

where η_ζ denotes $\frac{1}{1-\gamma^m} - \alpha$ times the elasticity of ζ with respect to ω. From the discussion above, $\eta_\zeta > 0$ (recall that $\alpha > 1/(1-\gamma^m)$). We know from Proposition 1 that $\eta_L > 0$, so the final term is negative.

This expression gathers all four effects of the skill premium on the share of employment in large firms. As the skill premium changes, employment in large firm

changes due to changes in relative unit cost (Lemma 2), changes in technology choice and exit thresholds (Lemma 3 and 4), changes in headcounts vs composites, and shifts between large and small modern firms ($\eta_\zeta > 0$). Proposition 1 states that the first two channels dominate the third one in terms of the employment share in modern firms if $\rho < 1/(1-\gamma^m)$.

Because η_ζ and η_L are both positive, the sign of $\frac{d \ln \ell^b}{d \ln \omega}$ is ambiguous. If $\rho < 1/(1-\gamma^m)$, $\eta_L > 0$, so that the elasticity of production employment in large firms with respect to the skill premium is positive when $\ell^m > \hat{\ell}^m$ and negative if $\ell^m < \hat{\ell}^m$, with $\hat{\ell}^m = (\eta_L - \eta_\zeta)/\eta_L$. This arises because for a modern employment share below $\hat{\ell}^m$, the first two channels dominate. For a modern employment share above this value, they are dominated by the final channel.

Finally, write the share of total employment in large firms, including overhead, as

$$\ell^b_{\text{tot}} = \frac{L^b + M^b c^f}{L^m + L^t + M c^f} = \ell^b (L^m + L^t) + m^b \cdot M c^f, \tag{28}$$

where M^b is the mass of large firms and $m^b = M^b/M = (z^x/z^b)^\alpha$ is their share of operating firms. Since we are dealing with large firms, $\ell^b > m^b$. As in Proposition 1, M rises in the skill premium, since a higher skill premium reduces average firm size. A higher skill premium thus reduces ℓ^b under the conditions just discussed, and increases the weight on the smaller share m^b. The change in m^b itself is theoretically ambiguous. Using the definitions of z^x and z^b, it is given by

$$\frac{d \ln(z^x/z^b)}{d \ln \omega} = \gamma^t (\Lambda^t_h - \Lambda_h) - (\Lambda^m_h - \Lambda_h) + (1-\gamma^m)\rho(\Lambda^m_h - h^m_{\text{tot}}). \tag{29}$$

The first term reflects lower z^x, the second term relatively higher costs of large firms, and the third term the change in large firms' headcount conversion term ϕ^m with ω. Here $\Lambda_h \equiv \ell^m \Lambda^m_h + \ell^t \Lambda^t_h$ denotes the aggregate cost share of high-skill labor. Given our assumptions on skill intensity, the sum of the two first terms, $\gamma^t(\Lambda^t_h - \Lambda_h) - (\Lambda^m_h - \Lambda_h)$, is negative. Because $\omega > 1$, the last term is positive. Under the assumption $\rho < 1/(1-\gamma^m)$, the elasticity is negative if $h^m_{\text{tot}} > (1-\gamma^t)\Lambda_h + \gamma^t \Lambda^t_h$ or $\gamma^t > \frac{\Lambda_h - h^m_{\text{tot}}}{\Lambda_h - \Lambda^t_h} \equiv \hat{\gamma}^t$. All terms in this inequality apart from γ^t and all terms entering the full elasticity apart from γ^i and ρ are directly observable data, so that $\hat{\gamma}^t$ can be computed from data.

Overall, the share of total employment in large firms is guaranteed to decreases in the skill premium if $\rho < 1/(1-\gamma^m)$, $\ell^m < \hat{\ell}^m$ and $\gamma^t > \hat{\gamma}^t$.

□

E.4. Labor market clearing conditions

First, note that $\bar{w}^t$ and $\bar{w}^m$ can be re-written as a function of w_l and the skill premium as

$$\bar{w}^i = w_l \left[(1-\mu^i)^\rho + (\mu^i)^\rho \omega^{1-\rho} \right]^{\frac{1}{1-\rho}}$$

and that the number of firms M and the number of entrants M^e are related by

$$M = (1 - G(z^x))M^e.$$

In equilibrium, the labor supplies L_l, L_h must equal the sum of production-labor demand across firms plus the overhead labor that entrants and active firms hire. Entrants (M^e) hire c^e workers each, and active firms (M) hire c^f workers each, all in the population skill mix. Labor market clearing for low-skill workers then requires

$$L_l = M^e \left[\int_{z^x}^{z^m} L_l^t(z) \mathrm{d}G(z) + \int_{z^m}^{\infty} L_l^m(z) \mathrm{d}G(z) \right] + \left(c^e M^e + c^f M \right) L_l.$$

Using the skill-specific labor demand functions and integrating optimal demands across firms implies

$$\begin{aligned} L_l &= \frac{M}{1-G(z^x)} \left[\left(\frac{(1-\mu^t)\bar{w}^t}{w_l} \right)^\rho \left(\frac{\gamma^t}{\bar{w}^t} \right)^{\frac{1}{1-\gamma^t}} Z^t + \left(\frac{(1-\mu^m)\bar{w}^m}{w_l} \right)^\rho \left(\frac{\gamma^m}{\bar{w}^m} \right)^{\frac{1}{1-\gamma^m}} Z^m \right] \\ &\quad + \left(c^e M^e + c^f M \right) L_l. \\ &= L_l^t + L_l^m + (c^e M^e + c^f M) L_l \end{aligned} \tag{30}$$

Given w_l and ω, this equation pins down the number of firms M. For high-skill workers, analogously,

$$\begin{aligned} L_h &= \frac{M}{1-G(z^x)} \left[\left(\frac{\mu^t \bar{w}^t}{w_h} \right)^\rho \left(\frac{\gamma^t}{\bar{w}^t} \right)^{\frac{1}{1-\gamma^t}} Z^t + \left(\frac{\mu^m \bar{w}^m}{w_h} \right)^\rho \left(\frac{\gamma^m}{\bar{w}^m} \right)^{\frac{1}{1-\gamma^m}} Z^m \right] \\ &\quad + \left(c^e M^e + c^f M \right) L_h. \\ &= L_h^t + L_h^m + (c^e M^e + c^f M) L_h \end{aligned} \tag{31}$$

Combining the two conditions,

$$\frac{L_h - (Mc^f + M^e c^e) L_h}{L_l - (Mc^f + M^e c^e) L_l} = \frac{L_h^t + L_h^m}{L_l^t + L_l^m}. \tag{32}$$

The overhead and entry terms on the left hand side cancels. Further, let $\ell_l^i \equiv L_l^i / (L_l^t +$

L_l^m) denote sector i's share of low-skill production employment and substitute on the right hand side to obtain

$$\frac{L_h}{L_l} = \ell_l^t \frac{L_h^t}{L_l^t} + \ell_l^m \frac{L_h^m}{L_l^m}. \tag{33}$$

The right-hand side of this equation decreases in ω: A higher skill premium reduces the skill intensity L_h^i/L_l^i for each firm type, and also shifts employment from more skill-intensive modern firms to less skill-intensive traditional ones. Hence, the equation pins down the skill premium ω given w_l and M. It also ensures that a greater skill supply (an incrase in L_h/L_l) reduces ω.

E.5. Equilibrium

An equilibrium of the model consists in values of the equilibrium variables w_h, w_l, M, z^m and z^x such that the following five equilibrium conditions hold. Values for the auxiliary variables $\omega, \bar{w}^t, \bar{w}^m, \hat{w}, \Lambda_j^i, \phi^i, m^m, Z^t, Z^m$ and Y and functions $\bar{L}^i(z), L_h^i(z), L_l^i(z)$ and $y^i(z)$ are implied, with definitions given in the following.

Equilibrium conditions:

(i) Technology choice:

$$z^m = \left(\frac{\bar{w}^{m\,\frac{\gamma^m}{1-\gamma^m}}}{\bar{w}^{t\,\frac{\gamma^t}{1-\gamma^t}}} \frac{\tilde{\gamma}^t}{\tilde{\gamma}^m} \right)^{1/\left(\frac{1}{1-\gamma^m} - \frac{1}{1-\gamma^t}\right)}.$$

(ii) Optimal continuation/exit:

$$z^x = \left(\frac{c^f}{\tilde{\gamma}^t} \hat{w} \right)^{1-\gamma^t} \bar{w}^{t\gamma^t}.$$

(iii) Labor market clearing, high skill:

$$L_h = \frac{M}{1-G(z^x)} \left[\left(\frac{\mu^t \bar{w}^t}{w_h} \right)^\rho \left(\frac{\gamma^t}{\bar{w}^t} \right)^{\frac{1}{1-\gamma^t}} Z^t + \left(\frac{\mu^m \bar{w}^m}{w_h} \right)^\rho \left(\frac{\gamma^m}{\bar{w}^m} \right)^{\frac{1}{1-\gamma^m}} Z^m \right] + \left(c^e M^e + c^f M \right) L_h.$$

(iv) Labor market clearing, low skill:

$$L_l = \frac{M}{1-G(z^x)}\left[\left(\frac{(1-\mu^t)\bar{w}^t}{w_l}\right)^{\rho}\left(\frac{\gamma^t}{\bar{w}^t}\right)^{\frac{1}{1-\gamma^t}} Z^t + \left(\frac{(1-\mu^m)\bar{w}^m}{w_l}\right)^{\rho}\left(\frac{\gamma^m}{\bar{w}^m}\right)^{\frac{1}{1-\gamma^m}} Z^m\right] + \left(c^e M^e + c^f M\right) L_l.$$

(v) Free entry:

$$c^e \hat{w} = \tilde{\gamma}^t \, \bar{w}^{t\,-\frac{\gamma^t}{1-\gamma^t}} Z^t + \tilde{\gamma}^m \, \bar{w}^{m\,-\frac{\gamma^m}{1-\gamma^m}} Z^m - c^f \hat{w}\left(1-G(z^x)\right).$$

Auxiliary objects:

- Skill premium: $\omega = w_h / w_l$.
- Unit cost, for each firm type:

$$\bar{w}^i = \left[(1-\mu^i)^{\rho} w_l^{1-\rho} + (\mu^i)^{\rho} w_h^{1-\rho}\right]^{\frac{1}{1-\rho}}.$$

- Cost shares in type-i firms:

$$\Lambda_h^i = \frac{w_h L_h^i}{\bar{w}^i \bar{L}^i} = \frac{1}{1+\left(\frac{1-\mu^i}{\mu^i}\right)^{\rho}\left(\frac{w_l}{w_h}\right)^{1-\rho}}, \qquad \Lambda_l^i = \frac{w_l L_l^i}{\bar{w}^i \bar{L}^i} = \frac{1}{1+\left(\frac{\mu^i}{1-\mu^i}\right)^{\rho}\left(\frac{w_h}{w_l}\right)^{1-\rho}}.$$

- Average wage: $\hat{w} = w_h L_h + w_l L_l$.
- Optimal CES labor composite:

$$\bar{L}^i(z) = \left(\frac{z\gamma^i}{\bar{w}^i}\right)^{\frac{1}{1-\gamma^i}}.$$

- Skill-specific demands:

$$L_h^i(z) = \left(\frac{\mu^i \bar{w}^i}{w_h}\right)^{\rho} \bar{L}^i(z), \qquad L_l^i(z) = \left(\frac{(1-\mu^i)\bar{w}^i}{w_l}\right)^{\rho} \bar{L}^i(z).$$

- Headcount conversion: $L^i(z) = \phi^i \bar{L}^i(z)$ with

$$\phi^i = \left(\frac{\bar{w}^i \mu^i}{w_h}\right)^{\rho} + \left(\frac{\bar{w}^i (1-\mu^i)}{w_l}\right)^{\rho}.$$

- Share of modern firms:

$$m^m = (z^x / z^m)^\alpha.$$

- Pareto-integrated productivity:

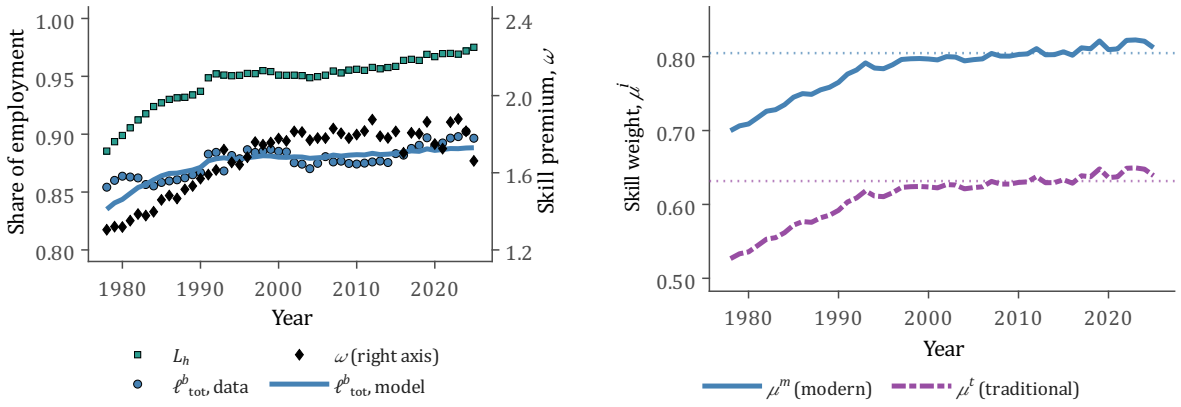
$$Z^m = \frac{\alpha z_{\min}^\alpha}{\alpha - \frac{1}{1-\gamma^m}} z^{m\left(\frac{1}{1-\gamma^m}-\alpha\right)},$$
$$Z^t = \frac{\alpha z_{\min}^\alpha}{\alpha - \frac{1}{1-\gamma^t}} \left(z^{x\frac{1}{1-\gamma^t}-\alpha} - z^{m\frac{1}{1-\gamma^t}-\alpha} \right)$$

- Optimal firm-level output: $y^i(z) = z^{\frac{1}{1-\gamma^i}} \left(\frac{\gamma^i}{\bar{w}^i}\right)^{\frac{\gamma^i}{1-\gamma^i}}$.

- Aggregate output:

$$Y = \frac{M}{1 - G(z^x)} \left[\left(\frac{\gamma^t}{\bar{w}^t}\right)^{\frac{\gamma^t}{1-\gamma^t}} Z^t + \left(\frac{\gamma^m}{\bar{w}^m}\right)^{\frac{\gamma^m}{1-\gamma^m}} Z^m \right].$$

F. Additional tables and figures – quantitative analysis



(A) L_h, ℓ^b_{tot} and the skill premium

(B) Implied μ^m and μ^t

FIGURE F.I
The skill endowment, large-firm employment and the skill premium in the US, 1978–2025, varying skill weights μ^i

Notes. Each year, L_h takes its US value, and μ^m and μ^t shift by a common amount so that the skill premium in the model equals that in the data; all other parameters are at their US values. In panel (a), markers show data (large-firm share: BDS to 1990, CPS from 1991; skill premium on the right axis), and the line shows ℓ^b in the model. In panel (b), dotted lines show the values of μ^m and μ^t in the US calibration. Large firms have at least ten employees and high-skill workers at least ten years of schooling.

TABLE F.I
Calibration to each country separately

	Data				Calibrated parameters			Firms			Firm shares		Optimal employment	
	L_h	ω	ℓ^b_{tot}	h^b_{tot}	μ^m	μ^t	c^f	nr firms	avg size	$c^f M$	large	modern	$L^t(z^x)$	$L^m(z^m)$
BEN	0.516	1.59	0.324	0.658	0.652	0.568	0.536	0.69	1.26	0.368	0.005	0.007	0.79	7.95
BFA	0.339	1.46	0.202	0.535	0.602	0.506	0.512	0.87	1.00	0.443	0.002	0.003	0.75	8.69
BOL	0.767	1.44	0.564	0.838	0.685	0.586	0.569	0.40	2.16	0.229	0.015	0.023	0.84	7.04
CMR	0.520	1.47	0.482	0.639	0.630	0.514	0.638	0.44	1.98	0.279	0.012	0.014	0.95	8.38
ETH	0.637	1.17	0.543	0.800	0.624	0.383	0.656	0.38	2.28	0.249	0.015	0.019	0.97	8.37
GHA	0.578	1.85	0.479	0.746	0.708	0.553	0.622	0.42	2.05	0.263	0.012	0.017	0.97	7.62
GIN	0.496	0.97	0.456	0.586	0.514	0.430	0.609	0.51	1.71	0.309	0.009	0.011	0.87	8.74
KHM	0.237	1.61	0.515	0.283	0.561	0.483	0.754	0.36	2.39	0.273	0.015	0.015	1.10	9.99
LBR	0.554	1.70	0.273	0.661	0.668	0.615	0.489	0.82	1.06	0.400	0.003	0.005	0.71	7.77
MWI	0.598	2.24	0.563	0.686	0.731	0.640	0.645	0.34	2.54	0.221	0.017	0.024	1.00	7.61
NER	0.203	2.22	0.260	0.394	0.666	0.486	0.780	0.51	1.71	0.397	0.005	0.005	1.18	11.00
NGA	0.852	1.17	0.635	0.950	0.711	0.481	0.546	0.34	2.57	0.184	0.020	0.034	0.80	6.33
PSE	0.782	1.14	0.752	0.813	0.622	0.522	0.734	0.18	4.82	0.132	0.045	0.061	1.06	7.79
RWA	0.166	3.14	0.264	0.443	0.745	0.010	1.723	0.23	3.78	0.395	0.013	0.005	2.74	20.86
SEN	0.287	2.71	0.300	0.526	0.736	0.555	0.806	0.44	1.98	0.353	0.007	0.007	1.29	9.87
SLE	0.657	1.15	0.608	0.768	0.609	0.421	0.684	0.31	2.77	0.214	0.021	0.026	1.00	8.33
TZA	0.372	2.45	0.353	0.590	0.729	0.573	0.718	0.45	1.91	0.326	0.008	0.010	1.14	8.86
UGA	0.586	2.22	0.536	0.779	0.753	0.521	0.714	0.30	2.88	0.215	0.019	0.026	1.21	7.84
ZMB	0.642	2.52	0.653	0.744	0.767	0.622	0.709	0.22	3.89	0.158	0.031	0.045	1.20	7.41
ZWE	0.698	2.58	0.714	0.696	0.760	0.764	0.667	0.23	3.77	0.153	0.033	0.047	0.95	7.42
ALB	0.759	1.33	0.680	0.791	0.650	0.581	0.666	0.25	3.40	0.170	0.028	0.040	0.97	7.57
ARG	0.812	1.48	0.683	0.859	0.700	0.592	0.623	0.26	3.39	0.159	0.028	0.044	0.93	6.90
ARM	0.978	0.95	0.638	0.989	0.746	0.629	0.494	0.37	2.34	0.183	0.018	0.035	0.71	5.80
CHL	0.834	1.79	0.765	0.869	0.742	0.625	0.665	0.16	5.34	0.108	0.050	0.083	1.02	6.61
CHN	0.391	1.87	0.771	0.431	0.635	0.482	1.000	0.13	6.63	0.131	0.065	0.062	1.55	10.38
COL	0.646	1.85	0.638	0.779	0.717	0.475	0.747	0.23	3.71	0.175	0.029	0.038	1.22	8.09
ECU	0.602	1.74	0.599	0.756	0.699	0.416	0.776	0.26	3.36	0.200	0.025	0.029	1.25	8.73
EGY	0.791	1.35	0.766	0.857	0.678	0.447	0.754	0.15	5.62	0.116	0.053	0.074	1.15	7.63
IRQ	0.406	1.56	0.477	0.572	0.627	0.363	0.792	0.36	2.44	0.282	0.014	0.014	1.21	10.16
LKA	0.623	1.32	0.444	0.759	0.638	0.518	0.565	0.53	1.63	0.301	0.009	0.012	0.83	7.73
MEX	0.429	1.77	0.546	0.568	0.654	0.393	0.856	0.28	3.09	0.240	0.021	0.020	1.34	10.25
MNG	0.861	1.34	0.855	0.871	0.683	0.591	0.800	0.08	10.32	0.067	0.109	0.166	1.17	7.16
NAM	0.677	3.46	0.429	0.819	0.835	0.755	0.507	0.55	1.59	0.277	0.008	0.015	0.81	6.16
PER	0.780	1.70	0.649	0.870	0.732	0.574	0.607	0.28	3.12	0.169	0.025	0.040	0.94	6.69
RUS	0.947	1.26	0.844	0.953	0.728	0.635	0.717	0.09	9.38	0.066	0.097	0.169	1.04	6.41
URY	0.622	1.59	0.721	0.698	0.662	0.471	0.809	0.18	4.86	0.144	0.043	0.052	1.26	8.58
ZAF	0.775	2.45	0.728	0.827	0.784	0.675	0.651	0.19	4.66	0.121	0.041	0.069	1.05	6.56
AUS	0.978	1.36	0.813	0.981	0.784	0.716	0.632	0.12	6.99	0.078	0.069	0.134	0.91	5.87
AUT	0.896	1.39	0.778	0.898	0.705	0.694	0.692	0.16	5.35	0.112	0.051	0.078	0.99	7.05
CHE	0.900	1.38	0.776	0.904	0.707	0.682	0.683	0.16	5.28	0.112	0.050	0.078	0.98	6.96
CYP	0.871	1.57	0.628	0.913	0.739	0.644	0.541	0.35	2.50	0.187	0.019	0.033	0.80	6.30
CZE	0.966	1.54	0.838	0.967	0.781	0.773	0.695	0.10	8.82	0.068	0.091	0.162	0.99	6.25
DEU	0.949	1.56	0.861	0.951	0.766	0.719	0.735	0.08	11.56	0.055	0.122	0.218	1.06	6.29
DNK	0.912	1.15	0.862	0.916	0.676	0.611	0.807	0.08	10.81	0.065	0.115	0.175	1.16	7.17
FIN	0.889	1.23	0.778	0.899	0.680	0.622	0.697	0.16	5.40	0.112	0.051	0.079	1.01	7.09
FRA	0.899	1.24	0.829	0.902	0.684	0.651	0.762	0.11	7.90	0.084	0.081	0.121	1.09	7.20
GBR	0.919	1.44	0.835	0.927	0.731	0.651	0.716	0.10	8.61	0.072	0.088	0.150	1.04	6.52
ISL	0.756	1.21	0.818	0.770	0.622	0.558	0.816	0.12	7.31	0.097	0.074	0.098	1.18	7.99
JPN	0.958	1.33	0.867	0.964	0.752	0.619	0.732	0.07	13.18	0.048	0.141	0.262	1.07	6.07
KOR	0.953	1.68	0.651	0.970	0.800	0.715	0.490	0.35	2.50	0.170	0.020	0.040	0.71	5.58
MLT	0.609	1.51	0.760	0.642	0.637	0.550	0.801	0.16	5.40	0.129	0.051	0.063	1.19	8.37
NLD	0.886	1.40	0.882	0.884	0.700	0.761	0.885	0.07	13.00	0.059	0.143	0.199	1.26	7.71
NOR	0.824	1.23	0.852	0.836	0.648	0.565	0.830	0.09	9.73	0.074	0.103	0.145	1.21	7.60
SVN	0.916	1.40	0.811	0.914	0.716	0.751	0.738	0.13	6.74	0.095	0.067	0.101	1.05	7.21
SWE	0.926	1.06	0.860	0.928	0.667	0.633	0.819	0.08	10.41	0.068	0.111	0.162	1.17	7.38
USA	0.959	1.79	0.883	0.965	0.805	0.632	0.748	0.04	19.67	0.033	0.214	0.433	1.17	5.70

Notes. The table reports the calibration targets and the internally calibrated parameters for each country group, with the United States shown separately for reference. Countries are grouped by GDP per capita (PPP) as in Table I. Quantities reported to the right of the calibrated parameters are model implications and are not targeted.

F.1. Lower value of the elasticity of substitution: $\rho = 1.5$

TABLE F.II
Calibration to the US economy– $\rho = 1.5$

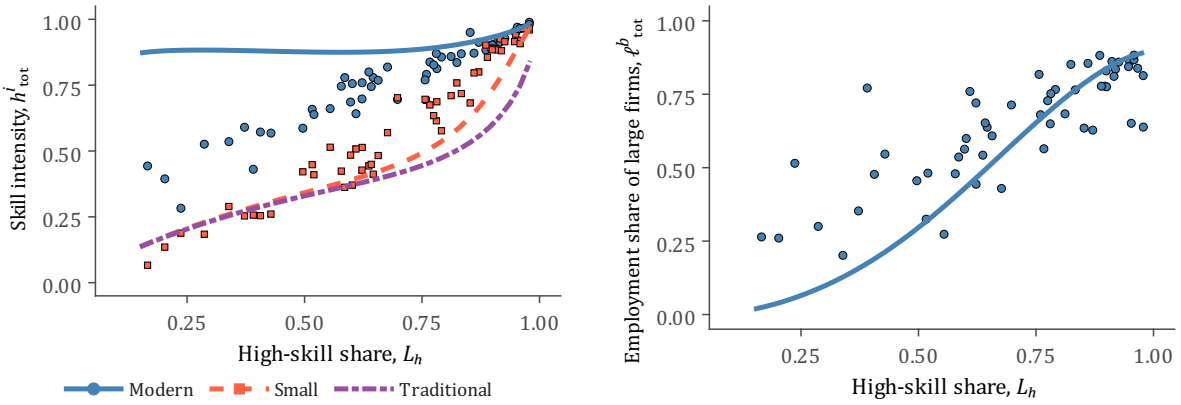
Parameter / moment	Value	Source / target
Pre-set parameters		
ρ	1.5	Hendricks and Schoellman (2023); Bils et al. (2024)
γ^m	0.85	Atkeson and Kehoe (2007)
$z_{\min}$	1	normalization
Calibrated parameters		
α	7.533	Tail index of employment = 1.13 (BDS)
μ^m	0.943	Skill intensity of large firms, $h^b_{\text{tot}} = 0.97$
μ^t	0.510	Employment share of large firms, $\ell^b_{\text{tot}} = 0.88$
γ^t	0.241	Skill premium, $\omega = 1.79$
c^f	0.743	Count share of large firms, $m^b = 0.214$
c^e	0.100	normalization
Supply-side input		
L_h	0.959	Skilled share of the US labor force, L_h
Implied equilibrium objects		
m^m	0.446	Share of modern firms
ℓ^m	0.990	Employment share in modern firms
$L^t(z^x)$	1.08	Employment of smallest active firm (incl. overhead)
$L^m(z^m)$	5.58	Employment of smallest modern firm (incl. overhead)
$c^f M$	0.033	Fixed overhead share of employment
$1/M$	19.68	Average firm size (incl. overhead)

Notes. As Table IV, but with the elasticity of substitution between high- and low-skill workers set to $\rho = 1.5$ instead of the baseline value. The table lists each model parameter together with its value and the target or source used to set it. The skill weights μ^m and μ^t, returns to scale in the traditional technology γ^t, and the fixed cost c^f are calibrated internally to match the same four US moments exactly: the skill intensity of large firms, the share of employment in large firms, the skill premium and the share of firms that are large. All other parameters are set externally or are normalizations, as in Table IV.

TABLE F.III
Calibration to each country income group separately – $\rho = 1.5$

	Data				Calibrated parameters			Firms			Firm shares		Optimal employment	
	L_h	ω	ℓ^b_{tot}	h^b_{tot}	μ^m	μ^t	c^f	nr firms	avg size	$c^f M$	large	modern	$L^t(z^x)$	$L^m(z^m)$
LIC	0.513	1.84	0.478	0.646	0.734	0.353	0.813	0.38	2.28	0.309	0.014	0.014	1.14	9.84
MIC	0.702	1.69	0.661	0.780	0.798	0.395	0.765	0.25	3.44	0.193	0.028	0.036	1.08	8.26
HIC	0.893	1.39	0.810	0.901	0.859	0.712	0.709	0.14	6.37	0.097	0.063	0.105	0.95	6.66
US	0.959	1.79	0.883	0.965	0.943	0.510	0.743	0.04	19.68	0.033	0.214	0.446	1.08	5.58

Notes. As Table V, but with the elasticity of substitution between high- and low-skill workers set to $\rho = 1.5$. The table reports the calibration targets and the internally calibrated parameters for each country income group, with the United States shown separately. Returns to scale in the traditional technology, γ^t, are held at the US value of Table F.II; the skill weights μ^m and μ^t and the fixed cost c^f are set to match the share of employment in large firms, the skill intensity of large firms and the skill premium. Income groups are defined by GDP per capita (PPP) as in Table I.



(A) Skill endowment and the firm's skill mix (B) Skill endowment and large-firm employment

FIGURE F.II
The skill endowment, firms' input and technology choices – $\rho = 1.5$.

Notes. As Panels (a) and (c) of Figure VI, but under the calibration with elasticity of substitution $\rho = 1.5$ (Table F.II). The horizontal axis in both panels is the economy-wide skill endowment L_h, which is varied while every other model input is held at its US value and the equilibrium is recomputed. Lines show model outcomes and markers the data, one marker per country. Panel (a) shows the skill intensity of modern (equivalently, large) firms, of traditional firms and of small firms; Panel (b) the share of employment in large firms, ℓ^b. The US data point at the top right is matched exactly by construction; nothing to its left is targeted.

TABLE F.IV
Determinants of the variation in employment in large firms across countries – $\rho = 1.5$

	$\log \ell^b_{\text{tot}}$			$\log \omega$		
Parameter	alone	drop	Shapley	alone	drop	Shapley
L_h	322.1%	51.6%	161.2%	557.0%	-385.6%	94.2%
μ^m	4.2%	-86.0%	-48.1%	341.0%	-337.2%	-1.1%
μ^t	6.3%	30.1%	8.2%	55.2%	-97.3%	-6.4%
c^f	53.3%	-77.4%	-21.3%	31.2%	3.7%	13.3%

Notes. As Table VI, but under the calibration with elasticity of substitution $\rho = 1.5$. The table decomposes the cross-country variance of $\ln \ell^b$ implied by the country-specific calibrations into the contributions of the three model inputs that vary across countries — the skill endowment L_h, the skill weights μ^i and the fixed operating cost c^f — under the three schemes described in Table VI: *alone*, *drop* and *Shapley*.

F.2. Alternative definition of skilled as ≥ 13 years of schooling

TABLE F.V
Calibration to the US economy – define skilled as ≥ 13 years of schooling

Parameter / moment	Value	Source / target
Pre-set parameters		
ρ	4	Hendricks and Schoellman (2023); Bils et al. (2024)
γ^m	0.85	Atkeson and Kehoe (2007)
$z_{\min}$	1	normalization
Calibrated parameters		
α	7.533	Tail index of employment $= 1.13$ (BDS)
μ^m	0.661	Skill intensity of large firms, $h^b_{\text{tot}} = 0.69$
μ^t	0.488	Employment share of large firms, $\ell^b_{\text{tot}} = 0.88$
γ^t	0.451	Skill premium, $\omega = 1.60$
c^f	0.818	Count share of large firms, $m^b = 0.214$
c^e	0.100	normalization
Supply-side input		
L_h	0.672	Skilled share of the US labor force, L_h
Implied equilibrium objects		
m^m	0.351	Share of modern firms
ℓ^m	0.966	Employment share in modern firms
$L^t(z^x)$	1.70	Employment of smallest active firm (incl. overhead)
$L^m(z^m)$	6.74	Employment of smallest modern firm (incl. overhead)
$c^f M$	0.036	Fixed overhead share of employment
$1/M$	19.54	Average firm size (incl. overhead)

Notes. As Table IV, but high-skill workers are defined as those with at least thirteen years of schooling when constructing the empirical targets. The table lists each model parameter with its value and the target or source used to set it; the skill weights μ^m and μ^t, returns to scale in the traditional technology γ^t, and the fixed cost c^f are calibrated internally to match the skill intensity of large firms, the share of employment in large firms, the skill premium and the share of firms that are large, all measured under the alternative skill definition. All other parameters are set externally or are normalizations, as in Table IV.

TABLE F.VI
Calibration to each country income group separately – define skilled as $\geq$ 13 years of schooling

	Data				Calibrated parameters			Firms			Firm shares		Optimal employment	
	L_h	ω	ℓ^b_{tot}	h^b_{tot}	μ^m	μ^t	c^f	nr firms	avg size	$c^f M$	large	modern	$L^t(z^x)$	$L^m(z^m)$
LIC	0.266	2.20	0.478	0.375	0.660	0.543	0.567	0.35	2.48	0.198	0.014	0.018	1.13	8.28
MIC	0.347	1.91	0.661	0.426	0.639	0.484	0.690	0.19	4.66	0.128	0.038	0.046	1.40	8.49
HIC	0.599	1.30	0.810	0.620	0.596	0.520	0.632	0.10	8.40	0.065	0.083	0.146	1.19	6.30
US	0.672	1.60	0.883	0.692	0.661	0.488	0.818	0.04	19.54	0.036	0.214	0.351	1.70	6.74

Notes. As Table V, but with high-skill workers defined as those with at least thirteen years of schooling when constructing the targets. The table reports the calibration targets and the internally calibrated parameters for each country income group, with the United States shown separately. Returns to scale in the traditional technology, γ^t, are held at the US value of Table F.V; the skill weights μ^m and μ^t and the fixed cost c^f are set to match the share of employment in large firms, the skill intensity of large firms and the skill premium. Income groups are defined by GDP per capita (PPP) as in Table I.

TABLE F.VII
Determinants of the variation in employment in large firms across countries – define skilled as $\geq$ 13 years of schooling

	$\log \ell^b_{\text{tot}}$			$\log \omega$		
Parameter	alone	drop	Shapley	alone	drop	Shapley
L_h	31.0%	63.2%	74.7%	38.0%	41.1%	37.9%
μ^m	18.8%	-20.8%	19.4%	34.9%	59.4%	45.4%
μ^t	20.9%	-37.3%	3.0%	5.0%	9.1%	6.3%
c^f	52.9%	-88.4%	2.8%	3.6%	19.4%	10.4%

Notes. As Table VI, but with high-skill workers defined as those with at least thirteen years of schooling when constructing the calibration targets. The table decomposes the cross-country variance of $\text{ł}\ln \ell^b$ implied by the country-specific calibrations into the contributions of the skill endowment L_h, the skill weights μ^i and the fixed operating cost c^f, under the three schemes described in Table VI: *alone*, *drop* and *Shapley*. See Table F.IV for the analogous decomposition under a lower elasticity of substitution.