

# Cross-country Variation in Educational Attainment: Structural Change or Within Industry Skill Upgrading?\*

Lutz Hendricks

University of North Carolina at Chapel Hill, Department of Economics  
CESifo, Munich; CFS, Frankfurt

March 2, 2010

## Abstract

Educational attainment varies greatly across countries and within countries over time. This paper asks whether the variation in education is primarily due to industry composition or to within industry skill intensities. The main finding is that within industry variation accounts for at least two-thirds of the cross-country and the time series variation in educational attainment. The within industry education gaps are broadly consistent with a model of industry neutral cross-country differences in skilled labor productivity. These results suggest that theories of educational development should focus on skill upgrading within industries rather than structural change.

Key words: education, skill bias, industry specialization.

JEL: I2, J24

## 1 Introduction

The educational attainment of countries increases dramatically with the level of development. In cross-country data, average years of schooling range from 3.5 years in Sub-Saharan Africa to 12 years in the U.S. (Barro and Lee, 2001). A large literature has explored the causes of cross-country schooling gaps<sup>1</sup> and their contribution to differences in output per worker.<sup>2</sup> The majority of this literature is based on one sector models.

This contrasts with a development literature that emphasizes the importance of the structural transformation for income differences and growth.<sup>3</sup> Many low income countries employ more than 80% of their workers in sectors of low skill intensities, such as agriculture. High income countries, on the other hand, specialize in skill intensive industries. Gollin, Parente, and Rogerson (2002) and Restuccia, Yang, and Zhu

---

\*For helpful comments I am grateful to the editor and to two anonymous referees. I also thank seminar audiences at Concordia University and at the 2008 Midwest Macro Meetings.

<sup>1</sup>See Bills and Klenow (2000); Kaboski (2006); Schoellman (2006); Hendricks (2008); Cordoba and Ripoll (2009) on the causes of cross-country schooling gaps. Galor (2000), Galor and Weil (2000) and Galor (2005) explore the evolution of education during the process of development. Goldin (1998), Caselli and Coleman (2001), Goldin (2001), Goldin and Katz (2008), and Restuccia and Vandenbroucke (2008) study the rise in U.S. schooling during the twentieth century. A related literature studies skill-biased technical change (e.g., Acemoglu 2002, 2003).

<sup>2</sup>See Klenow and Rodriguez-Clare (1997); Hall and Jones (1999); Hendricks (2002); Erosa, Koreshkova, and Restuccia (2007); Manuelli and Seshadri (2007), among others.

<sup>3</sup>Early contributions include Lewis (1954) and Harris and Todaro (1970). More recent work includes Caselli and Coleman (2001), Gollin, Parente, and Rogerson (2002), Hansen and Prescott (2002), and Lucas (2004).

(2007) argue that variation in sectoral employment shares is important for understanding cross-country income gaps.

These findings raise the question whether sectoral employment shares are also important for understanding variation in education across countries and within countries over time. To address this question, this paper decomposes cross-country and time series variation in educational attainment into the contributions of industry specialization and within industry skill gaps.

**Implementation.** Section 2 constructs a dataset of skilled and unskilled labor inputs for 28 countries and 15 industries. The data are drawn from large, individual level datasets provided by the IPUMS International database (Minnesota Population Center 2009). Workers are divided into skilled and unskilled groups depending on their level of schooling. I explore three definitions of skilled labor: primary, secondary, or higher education completed.

Section 3 develops a decomposition of cross-country education gaps into the contributions of industry composition and within industry educational variation. Roughly speaking, the decomposition answers the question: what fraction of the measured education gaps would remain if all cross-country variation in industry composition were removed? An analogous decomposition splits the time series variation in education into the contributions of changes in industry mix and within industry skill upgrading.

**Findings.** Results are presented for cross-country and time series variation in education. In both cases, I find that within industry skill gaps account for the majority of educational variation. Section 3.1 discusses the cross-country findings. For all skill definitions, around 75% of the education gaps relative to the U.S. are due to within industry gaps. A similar fraction is obtained when the analysis is restricted to the 9 countries with the largest education gaps.

It may appear surprising that variation in industry composition contributes relatively little, given that the fraction of employment in agriculture varies twenty-fold across countries. One reason is that within industry education gaps are large. The least skill intensive U.S. industry employs more skilled labor than nearly all sectors in the least educated countries. As a result, even dramatic changes in product mix fail to eliminate cross-country education gaps.

Section 3.2 presents the time series findings for 17 countries with available data. The time span between observations ranges from 10 to 40 years. I find that around 85% of the changes in education occur within industries. For a given definition of skill, the contribution of structural change is smaller than in the cross-section.

One way of summarizing quantitative findings is as follows: Upgrading the within industry skill intensities of the least educated countries to U.S. levels would eliminate at least 71% of the education gaps relative to the U.S. Similarly, within industry skill upgrading accounts for at least 79% of the largest observed time series changes in education.

One concern with these findings is that the level of industry detail is limited. This might bias the contribution of structural change downward. Section 4 therefore considers how the findings change when more detailed industry data are used. Since detailed industry data cover only a subset of all industries, it is necessary to develop an alternative method for decomposing country education gaps. The decomposition exploits the fact that cross-country education gaps are approximately *industry neutral* in the sense of Treffer (1993). Industry neutrality means that the relative productivity of skilled labor varies by the same factor in all industries for a given country. This result may be of independent interest. I find no evidence that finer industry data reduce the role of within industry gaps in either the time series or the cross-section.

**Related literature.** The paper is related to a small literature studying the sources of cross-country variation in education (see footnote 1 for references). The objective of this paper is to document an observation which may guide the development of theories of country educational attainment.

The time series results are consistent with the literature on skill-biased technical change, which finds that roughly two-thirds of the changes in skill intensities within countries are due to within industry skill upgrading (Berman, Bound, and Griliches 1994; Autor, Katz, and Krueger 1998; Berman, Bound, and Machin 1998; Machin and Van Reenen 1998). Relative to this literature, my data cover a wider range of countries and industries, and offer a better measure of educational attainment. Most of the existing evidence is limited to skill upgrading within manufacturing in high income countries. Only Desjonqueres, Machin, and van Reenen (1999) and Berman and Machin (2000b), which is summarized in section III of Berman and Machin (2000a), present evidence for several low-income countries. Moreover, most studies use the share of non-production labor as a proxy for educational attainment. Only Desjonqueres, Machin, and van Reenen (1999) present educational attainment for non-manufacturing industries. However, their database covers only eight high-income countries.

Two of my findings resemble the literature on human capital in cities. Hendricks (2011) shows that industry neutral differences in skilled labor productivity account well for cross-city data in the U.S. Doms and Lewis (2006) and Hendricks (2011) find that the bulk of cross-city variation in education is due to within industry differences, not to variation in product mix.

## 2 Data

The data are obtained from the IPUMS International database (Minnesota Population Center, 2009). I use up to two waves for 28 countries that provide sufficient data to estimate labor inputs at the country-industry level. Table 17 in the Appendix shows summary statistics for each of the countries and years included in the database.

### 2.1 Background on the IPUMS International Database

The IPUMS International database collects and harmonizes international micro census data.<sup>4</sup> It is administered by the Minnesota Population Center in collaboration with national statistical offices. The database currently covers 130 censuses in 44 countries.

Each sample contains individual level, and mostly also household level, information about a range of topics, including education, work, income, family structure, and migration. The set of available variables and their coding varies across samples. For a subset of variables, IPUMS provides consistently recoded data. The sample design varies across samples. Most samples intend to cover the entire population of a nation. However, small sub-populations, such as transients or foreign embassy personnel, are often excluded. A small number of samples exclude larger sub-populations and even entire regions of the country.

The quality of the data is difficult to assess. IPUMS implements internal consistency checks and flags inconsistencies. However, I am not aware of any attempts to compare the IPUMS data with other sources. Hendricks (2010) compares industry and education data for a subset of samples. His findings motivate the choice of samples included in this paper.

---

<sup>4</sup>Ruggles, King, Levison, McCaa, and Sobek (2003) provide a brief description of the database. Information on some of the samples is contained in Hall, McCaa, and Thorvaldsen (2000).

## 2.2 Sample Selection

The unit of observation is an individual. To be included in the sample, a person must satisfy all of the following criteria:

1. Age (IPUMS variable AGE) between 15 and 70 years. This range trades off two objectives: (i) Labor inputs should be comprehensive. (ii) Persons with low work hours should be excluded.
2. Employment status (EMPSTAT) is "employed." Where EMPSTAT is not available, CLASSWKR is used.
3. Industry (INDGEN) is known.
4. Highest grade attained (EDATTAN) and years of schooling (YRSCHL) are known. Some countries report either intervalled years of schooling or completed school levels. These are recoded into years of schooling based on each country's typical duration of schooling by level.
5. Positive individual sampling weight (WTPER).

## 2.3 Variable Definitions

For each individual, I construct measures of labor input, education, and industry. These are then aggregated into measures of skilled ( $H_{m,i}$ ) and unskilled ( $L_{m,i}$ ) labor inputs at the country-industry level. Countries are indexed by  $m = 1, \dots, M$ . Industries are indexed by  $i = 1, \dots, I$ . My measurement choices generally follow Caselli and Coleman (2006) (hereafter CC).

**Schooling:** EDATTAN assigns each person, indexed by  $k$ , one of seven education levels: no school, some primary, primary completed, lower secondary, secondary, some tertiary, and tertiary. These levels are indexed by  $j = 1, \dots, 7$ . Following Caselli and Coleman (2006), I divide persons into "unskilled" and "skilled" groups based on their education level  $j$ . Persons with  $EDATTAN_k \geq j_0$  are skilled; others are unskilled. I consider the same three values of  $j_0$  as do CC: primary, secondary, and higher education completed.

IPUMS derives EDATTAN and YRSCHL from country specific variables that record either years of schooling or, more commonly, the highest degree attained. For several countries, these variables exhibit inconsistencies (see Hendricks 2010). In these cases, I recode the country specific variables into EDATTAN and YRSCHL.

**Individual labor input:** Within a skill group ( $H$  or  $L$ ), persons are perfect substitutes. Person  $k$  supplies  $l_k = \lambda_k \times hours_k$  efficiency units of labor, where  $hours_k$  denotes hours worked. Following CC, labor efficiency is measured by a Mincerian equation of the form  $\lambda_k = \exp(\mu(s_k) - \mu(s_0))$ , where  $s_k$  denotes years of schooling and  $\mu(s_k)$  maps years of schooling into labor efficiencies. For unskilled workers,  $s_0 = 0$ , so that labor is measured in no school equivalents. For skilled labor,  $s_0$  denotes average years of schooling completed at the highest unskilled level:  $s_0 = \sum_{j < j_0} \bar{s}_j$ , where  $\bar{s}_j$  denotes average years of schooling at level  $j$ . Thus, skilled labor is measured in education level  $j_0$  equivalent units.

I follow Hall and Jones (1999) in assuming common Mincer returns for all countries. The returns are 0.134 for the first four years of schooling, 0.1 for the next four years, and 0.068 for all additional years. The assumption that all countries share the same Mincer returns departs from CC's specification. This is motivated by recent work which suggests that cross-country variation in Mincer returns contains large

measurement errors (Banerjee and Duflo 2005). Allowing Mincer returns to vary across countries yields similar results.

Few countries report hours worked. For consistency, I therefore assume that workers of given age and education work the same hours in all countries and years. Each person’s work hours are set to the mean weekly hours worked in the corresponding age and education (EDATTAN) cell in the year 2000 U.S. sample.

**Industry labor input:** Industry labor inputs are weighted sums of individual labor inputs:  $H_{m,i} = \sum_k PERWT_k l_k$ , where the sum is taken over skilled persons in industry  $i$ .  $L_{m,i}$  is defined analogously. Summing  $H_{m,i}$  and  $L_{m,i}$  over industries yields country level labor inputs ( $H_m$  and  $L_m$ ). Table 18 in the Appendix lists the estimated skill intensities,  $H_m/L_m$ , for all samples.

The data contain two levels of industry detail. Most of the paper uses the variable INDGEN, which distinguishes 15 industries that are comparable across countries and years. I refer to this as the ”broad” industry definition. The data also report a finer industry classification (IND), which differs across countries and years. Section 4 examines how using this ”detailed” industry definition affects the findings.

**Skill premium:** Individual earnings are imputed from the Mincer equation  $z_k = \exp(\mu(s_k)) \times hours_k$ . Aggregate skilled earnings are then given by  $W_{H,m} = \sum_k z_k$  where the sum is taken over all skilled persons. Unskilled earnings ( $W_{L,m}$ ) are defined analogously. The skill premium is then given by

$$\frac{w_{H,m}}{w_{L,m}} = \frac{W_{H,m} L_m}{W_{L,m} H_m} \quad (1)$$

The data force me to abstract from skill premium variation across industries within a country. This assumption is violated in all of the samples for which individual earnings are reported. In these samples, skill premiums, measured by Mincer returns, are higher in more skill intensive industries. A possible concern is that abstracting from variation of skill premiums across industries could bias the estimated skill intensities.

To see this, consider an economy in which workers differ in unmeasured skills. Assume that labor mobility equates skill premiums across industries. In this case the correct measure of an industry’s skill intensity is the ratio of skilled to unskilled wage bills. By contrast, my measures of  $H_{m,i}$  and  $L_{m,i}$  abstract from unmeasured skills and are biased in this economy.

For the subset of samples that report labor earnings, I can compare the two measures of skill intensities. In all cases I find that both measures are very similar. The reason is that skill intensities vary far more across industries than do skill premiums. Whether skill premium variation is attributed to labor inputs or to skill prices is quantitatively not important.

**Educational attainment:** For most of the paper’s results, educational attainment is measured by the ratio of skilled to unskilled labor inputs

$$h_{m,i} = H_{m,i}/L_{m,i}. \quad (2)$$

An alternative measure is average years of schooling, defined as

$$S_{m,i} = \sum_k PERWT_k hours_k s_k / \sum_k PERWT_k hours_k \quad (3)$$

Table 1: Dispersion of educational attainment

	Std. dev.	U.S./Lowest 5
Primary	1.73	164.0
Secondary	1.19	69.9
Tertiary	1.18	32.2
$S_m$	2.50	13.6 / 4.7

Notes: The table shows the standard deviation of  $\ln(H_m/L_m)$  and the ratio of  $H_{US}/L_{US}$  to the average of  $H_m/L_m$  for the least educated five countries. Each row represents one definition of skilled labor. The final row shows the same statistics for average years of schooling.

where the sum is taken over persons in industry  $i$ . Where feasible, the findings for both measures are reported. Both yield similar conclusions. However, the detailed industry statistics of section 4 cannot be computed for average years of schooling. Also, theories often treat skilled and unskilled labor as imperfect substitutes. For these reasons, the main focus is placed on  $h_{m,i}$  as the measure of educational attainment.

### 3 Results

This section decomposes the cross-country and time series variation in educational attainment into the contributions of industry composition and within industry variation. The cross-sectional results use the most recent wave for each country. The time series results examine the changes within a country between two waves. The broad industry classification is used.

#### 3.1 Cross-country Education Gaps

Table 1 summarizes the cross-country variation in educational attainment. My measure of country  $m$ 's educational attainment is the ratio of skilled to unskilled labor input,  $h_m = H_m/L_m$ .

The table reveals a large dispersion of  $h_m$  across countries. For the primary skill cutoff, the ratio between the U.S. and the least educated five countries is above 160. The large ratio is due to the fact that only 4% of U.S. workers have completed at most primary schooling, compared with 95% of Rwandan workers. The dispersion of  $h_m$  declines with the skill cutoff. Even for the college cutoff, the U.S. employs 32 times more skilled labor than the five least educated countries. These findings are broadly consistent with CC's.

##### 3.1.1 Direct Decomposition

Table 1 reveals that educational attainment varies greatly across countries. My measure of the education gap between the U.S. and country  $m$  is  $h_{US}/h_m$ . The purpose of this paper is to decompose this gap into the contributions of industry composition and within industry education gaps.

A natural decomposition is suggested by the identity

$$\frac{h_{US}}{h_m} = \frac{\sum_i n_{US,i} h_{US,i}}{\sum_i n_{m,i} h_{m,i}} \quad (4)$$

Educational attainment in country  $m$  is a weighted average of the industry attainments. The weights are

the unskilled employment shares

$$n_{m,i} = L_{m,i}/L_m \quad (5)$$

Consider two counter-factual experiments, both of which transform  $h_m$  into  $h_{US}$ .

*Experiment I* answers the question: By how much would country  $m$ 's education gap decline, if the *U.S.* *downgraded* all of its industries to country  $m$ 's skill intensities? The experiment first replaces country  $m$ 's product mix ( $n_{m,i}$ ) with that of the reference country ( $n_{US,i}$ ), resulting in the counter-factual attainment level  $h_m^I = \sum_i n_{US,i} h_{m,i}$ .  $h_m^I/h_m$  measures the contribution of product mix to country education gaps. The experiment then replaces country  $m$ 's within industry education ( $h_{m,i}$ ) with their U.S. values. Thus,

$$\frac{h_{US}}{h_m^I} = \frac{\sum_i n_{US,i} h_{US,i}}{\sum_i n_{US,i} h_{m,i}} \quad (6)$$

measures the within industry education gaps.

*Experiment II* answers the complementary question: How much would country  $m$ 's education increase, if all of its industries *upgraded* their skill intensities to U.S. levels? It reverses the order of Experiment I's changes. Starting from  $h_m$ , it first replaces within industry attainment ( $h_{m,i}$ ) with  $h_{US,i}$ , resulting in  $h_m^{II} = \sum_i n_{m,i} h_{US,i}$ . Within industry education gaps are then given by

$$\frac{h_m^{II}}{h_m} = \frac{\sum_i n_{m,i} h_{US,i}}{\sum_i n_{m,i} h_{m,i}} \quad (7)$$

Then the experiment replaces the sector weights ( $n_{m,i}$ ) with their U.S. values, so that  $h_{US}/h_m^{II}$  measures the contribution of industry composition.

Both experiments offer sensible decompositions that differ only in the employment weights used to weight within industry education gaps ( $h_{US,i}$  vs.  $h_{m,i}$ ). Since both experiments yield broadly similar results, I only report a decomposition that uses the average employment shares of the U.S. and country  $i$  ( $\hat{n}_{m,i}$ ) as weights. Thus, within industry education gaps are given by

$$\frac{\hat{h}_{US,m}}{\hat{h}_m} = \frac{\sum_i \hat{n}_{m,i} h_{US,i}}{\sum_i \hat{n}_{m,i} h_{m,i}}. \quad (8)$$

The *fraction* of  $h_{US}/h_m$  attributed to within industry education gaps is given by  $\ln(\hat{h}_{US,m}/\hat{h}_m) / \ln(h_{US}/h_m)$ .

For example, if the country education gap is  $h_{US}/h_m = 16$  and the within industry gap is  $\hat{h}_{US,m}/\hat{h}_m = 4$ , one-half of the total gap is attributed to within industry gaps, while the other half is attributed to industry mix.

In almost all cases, the within industry gaps implied by Experiment II are slightly larger than those of Experiment I. Those based on average employment shares lie in between. I refer to (8) as the *direct* decomposition. In section 4, I derive an alternative decomposition based on a model of industry neutral cross-country differences in skilled labor productivity. It yields results that are quite similar to Experiment II.<sup>5</sup>

Most of the paper's findings are reported in the form of cross-country averages. Average educational attainment,  $\bar{h}$ , is defined as the geometric mean of  $h_m$  across countries. Average education gaps are

<sup>5</sup>An alternative decomposition that avoids using the U.S. as the reference country yields similar results. For each country, define a counterfactual  $h_m$  with common industry mix as  $\sum_i \bar{n}_i h_{m,i}$ , where  $\bar{n}_i$  is the employment share in industry  $i$  averaged across all countries. To measure the importance of within industry education gaps, regress the log counterfactual  $h_m$  on the log measured  $h_m$ . The fraction of cross-country education gaps that occurs within industries is defined as the slope coefficient of the regression.

Table 2: Direct decomposition. Cross-country data.

	Avg.gap	Within gap	Within fraction	$N$
Primary	31.78	12.78	0.74	27
Secondary	17.57	10.55	0.82	27
Tertiary	4.86	2.87	0.67	27
$S_m$	5.51	4.24	0.77	27

Notes: Direct decomposition of cross-country education gaps into the contributions of product mix and within industry gaps. Each row represents a skill definition. "Avg. gap" denotes the cross-country average of education gaps. The remaining columns show the average education gap due to within industry gaps, the fraction of education gaps due to within industry gaps, and the number of countries. The last row measures educational attainment by average years of schooling.

defined as the geometric means of the country level gaps ( $h_{US}/h_m$ ). However, the average *fraction* of cross-country gaps that is due to within industry gaps is defined as the arithmetic mean of the within industry gaps,  $\ln(\hat{h}_{US,m}/\hat{h}_m)$ , divided by the arithmetic mean of the country level gaps,  $\ln(h_{US}/\bar{h})$ .<sup>6</sup>

I also report results that use *average years of schooling* ( $S_m$ ) to measure a country's educational attainment. Education gaps are then defined as  $S_{US} - S_m$ . Within industry education gaps are given by a weighted average of  $S_{US,i} - S_{m,i}$ . The weights are the shares of hours worked in each industry, averaged across country  $m$  and the U.S.

### 3.1.2 Cross-country Results

Table 2 reports cross-country averages for the direct decomposition. For each skill cutoff, the table shows the average cross-country education gap and the average within industry gap. To illustrate: the first row shows results for the primary skill cutoff. The U.S. skill intensity is more than 30 times higher than the average skill intensity in the other countries. Of this 30-fold gap, a factor of 12.8 (74%) is due to within industry gaps.

The main message of Table 2 is that at least two-thirds of cross-country education gaps are due to within industry gaps. There is substantial variation across skill cutoffs in both the average gap and in the fraction due to within industry gaps. When education is measured by average years of schooling, similar findings emerge.

Figure 1 shows the within industry gaps by country for the three skill definitions. The horizontal axis displays each country's education gap,  $\ln(h_{US}) - \ln(h_m)$ . The figure also shows lines with slopes 0.5 and 1. Between these lines, within industry gaps contribute between 50% and 100% of a country's education gap. This is the case for the majority of countries, especially for those with low education. Hence, the averages reported in table 2 capture a feature of the data that is common to many countries.

**Country groups.** Table 3 studies how the findings vary between countries of different education levels. For this purpose, the countries are divided into three equally sized groups according to their  $h_m$ . Group 1 consists of the least educated countries.

Table 3 reports group averages for country education gaps and for the contributions of within industry gaps. The figures are constructed as in Table 2. The main finding is that within industry gaps account

<sup>6</sup>A natural alternative measure, the arithmetic mean of country level fractions due to within industry gaps, is sensitive to outliers.



Table 3: Direct decomposition. Cross-country data.

Primary skill cutoff	Avg.gap	Within industry gap	Fraction
Group 1 RWA, UGA, KHM, BRA, GHA, IND, PRT, VNM, COL	136.81	33.30	0.71
Group 2 ARG, BOL, VEN, MYS, ECU, CHN, MEX, CRI, PHL	47.38	21.87	0.80
Group 3 ZAF, PAN, CHL, JOR, ESP, GRC, ROM, ITA, CAN	4.95	2.86	0.66
Secondary skill cutoff			
Group 1 RWA, UGA, GHA, CHN, VNM, ARG, KHM, COL, IND	54.11	20.80	0.76
Group 2 VEN, MEX, CRI, PRT, ECU, BRA, BOL, ZAF, PAN	16.33	12.17	0.89
Group 3 ESP, CHL, MYS, ITA, PHL, JOR, GRC, ROM, CAN	6.14	4.63	0.84
Tertiary skill cutoff			
Group 1 RWA, KHM, CHN, COL, UGA, GHA, VNM, VEN, ZAF	19.68	8.77	0.73
Group 2 BOL, IND, CHL, ARG, ECU, ESP, BRA, ITA, MYS	3.34	2.30	0.69
Group 3 ROM, CRI, PAN, MEX, PRT, PHL, JOR, GRC, CAN	1.74	1.18	0.30

Notes: Direct decomposition of cross-country education gaps. Countries are grouped by  $h_m$ . The table shows the average education gap, the average gap due to within industry gaps, and the fraction of cross-country education gaps due to within industry gaps.

for at least 66% of the education gaps for all country groups. The only exception is the most educated group with the tertiary skill definition.

One statistic summarizes the findings of this section: If the least educated countries upgraded their within industry attainments to U.S. levels, holding product mix constant, at least 71% of the observed country education gaps would vanish.<sup>7</sup>

### 3.1.3 Intuition

The results presented so far indicate that product mix accounts for less than one-third of cross-country education gaps. This may appear surprising, given that countries differ greatly in their industry composition. Notably, the fraction of labor employed in agriculture ranges from 2% in the U.S. to 88% in Rwanda. In this section, I investigate why large differences in product mix do not translate into large contributions to cross-country education gaps. I document two facts: (i) More educated countries employ more skilled labor in all industries. (ii) In nearly all industries, the least educated countries employ less skilled labor than the U.S. median sector. Thus, even dramatic sectoral reallocation is not sufficient to eliminate large parts of the observed education gaps.

<sup>7</sup>This is the thought experiment of Experiment II, for which the contribution of within industry education gaps is slightly larger than what is shown in Table 3.

Table 4: within industry and aggregate education gaps.

Skill cutoff	Avg. $\beta$	Avg. $R^2$	Avg. N	N industries
Primary	0.88 (0.10)	0.82 (0.11)	27.5	15
Secondary	0.78 (0.14)	0.74 (0.12)	27.6	15
Tertiary	0.74 (0.22)	0.70 (0.17)	27.5	15

Notes: The table shows the results of regressing, by industry,  $\ln(h_{US,i}) - \ln(h_{m,i})$  on  $\ln(h_{US}) - \ln(h_m)$ . The statistics shown are the average slope coefficient (with its standard deviation in parentheses), the average regression  $R^2$ , the average number of countries included in the regression, and the number of industry regressions estimated.

**Within industry education gaps.** Figure 2 shows that more educated countries employ more skilled labor in all industries. For the largest eight industries, as measured by U.S. employment, the figure plots the industry education gap,  $\ln(h_{US,i}) - \ln(h_{m,i})$ , against the country gap,  $\ln(h_{US}) - \ln(h_m)$ . Results are shown for the "secondary" skill definition. In all cases, a strong positive relationship is apparent. The slope of an OLS regression is above 0.5 for each industry.

Table 4 shows that the same holds for the remaining industries. It displays the results of regressing, by industry, within industry gaps on country level gaps. The average slope coefficient is near 0.75, as is the average  $R^2$  statistic. Consistent with the previous findings, even if all countries employed the same product mix, around 75% of the cross-country education gaps would remain. The findings are roughly consistent across skill cutoffs.

**Distributions of industry attainment.** Figure 3 establishes the second fact: The distributions of skill intensities do not overlap much between low skill and high skill countries. That is, the least skill intensive U.S. industry is more skill intensive than most countries. Conversely, few industries in low education countries employ more skilled labor than even the least skill intensive U.S. sector.

Figure 3 shows the distribution of  $h_{m,i}$  by country. The horizontal axis displays country skill intensities,  $\ln(h_m)$ . The vertical axis shows  $\ln(h_{m,i})$  for all industries in country  $m$  together with the population weighted median of  $\ln(h_{m,i})$ . For the primary and secondary cutoffs, among the ten least educated countries at most one industry exceeds the skill intensity of median U.S. industry. Conversely, the least educated U.S. industry (agriculture or personal services, depending on the definition of skill) employs more skilled labor than the lowest two-thirds of countries ( $h_{US,Ag} > h_m$ ). The distributions of  $h_{m,i}$  overlap more for the college cutoff. Even in that case, in the four least educated countries no industry is more skilled than the U.S. median.

### 3.2 Time Series Decomposition

This section applies the decomposition developed in section 3.1 to the changes in educational attainment within countries over time. Denote by  $h_{t,m,i} = H_{t,m,i}/L_{t,m,i}$  the educational attainment of industry  $i$  in country  $m$  at date  $t$ . The goal is to decompose country level changes in attainment,  $\ln(h_{2,m}) - \ln(h_{1,m})$ , into the contributions of product mix ( $n_{t,m,i}$ ) changes and within industry ( $h_{t,m,i}$ ) changes.

The decomposition proceeds in the same way as for the cross-sectional results (see equation 8). Country  $m$ 's within industry education gap is defined as the weighted average of  $\ln(h_{2,m,i}) - \ln(h_{1,m,i})$ . The weights are the unskilled labor input shares, averaged across the two dates.

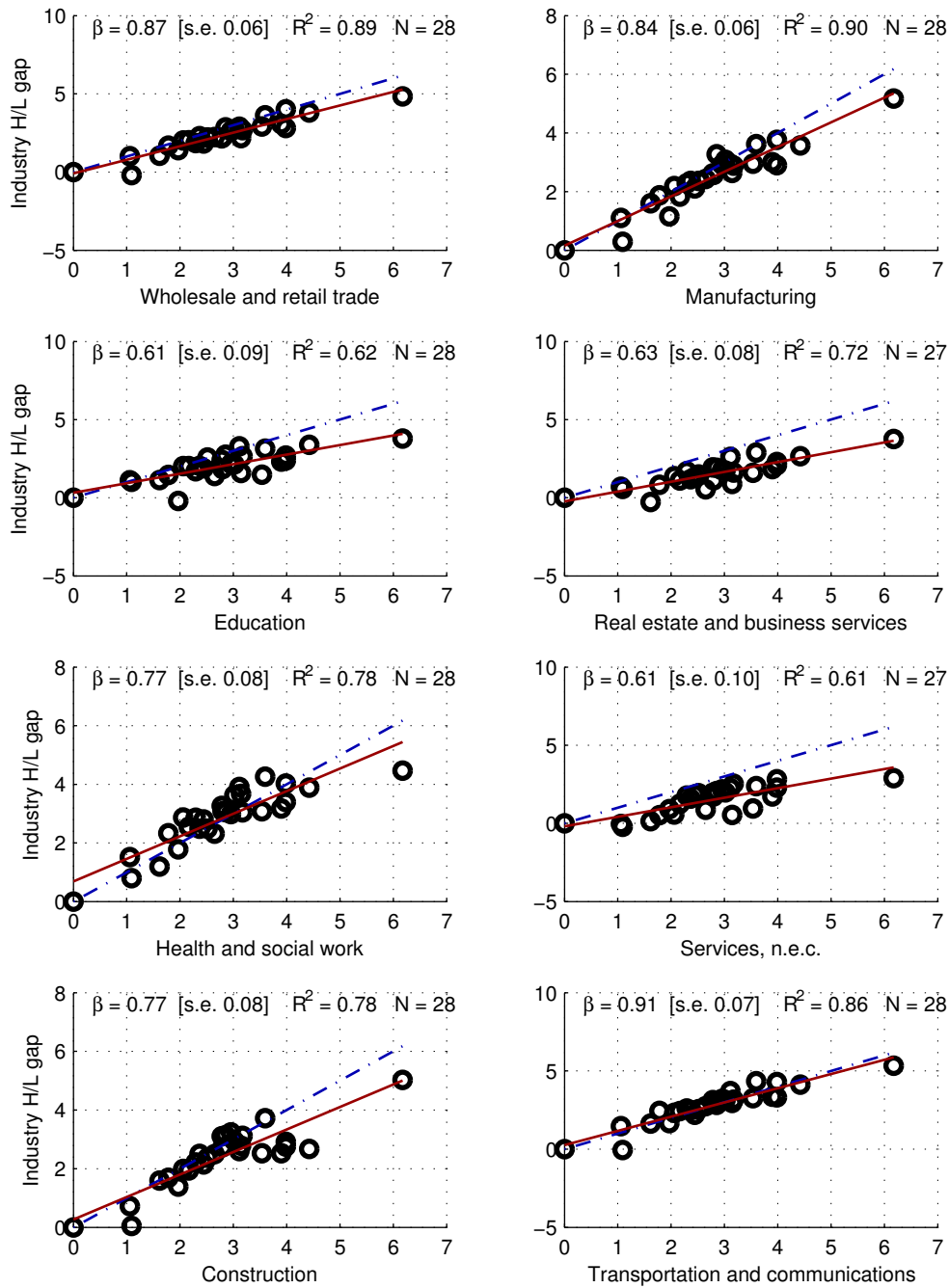


Figure 2: within industry and aggregate education gaps. Cross-country data.

The figure shows within industry education gaps,  $\ln(h_{US,i}/h_{m,i})$ , and country education gaps,  $\ln(h_{US}/h_m)$ , for the eight industries with the largest employment shares. Also shown are 45-degree lines (dashed) and OLS regression lines (solid).

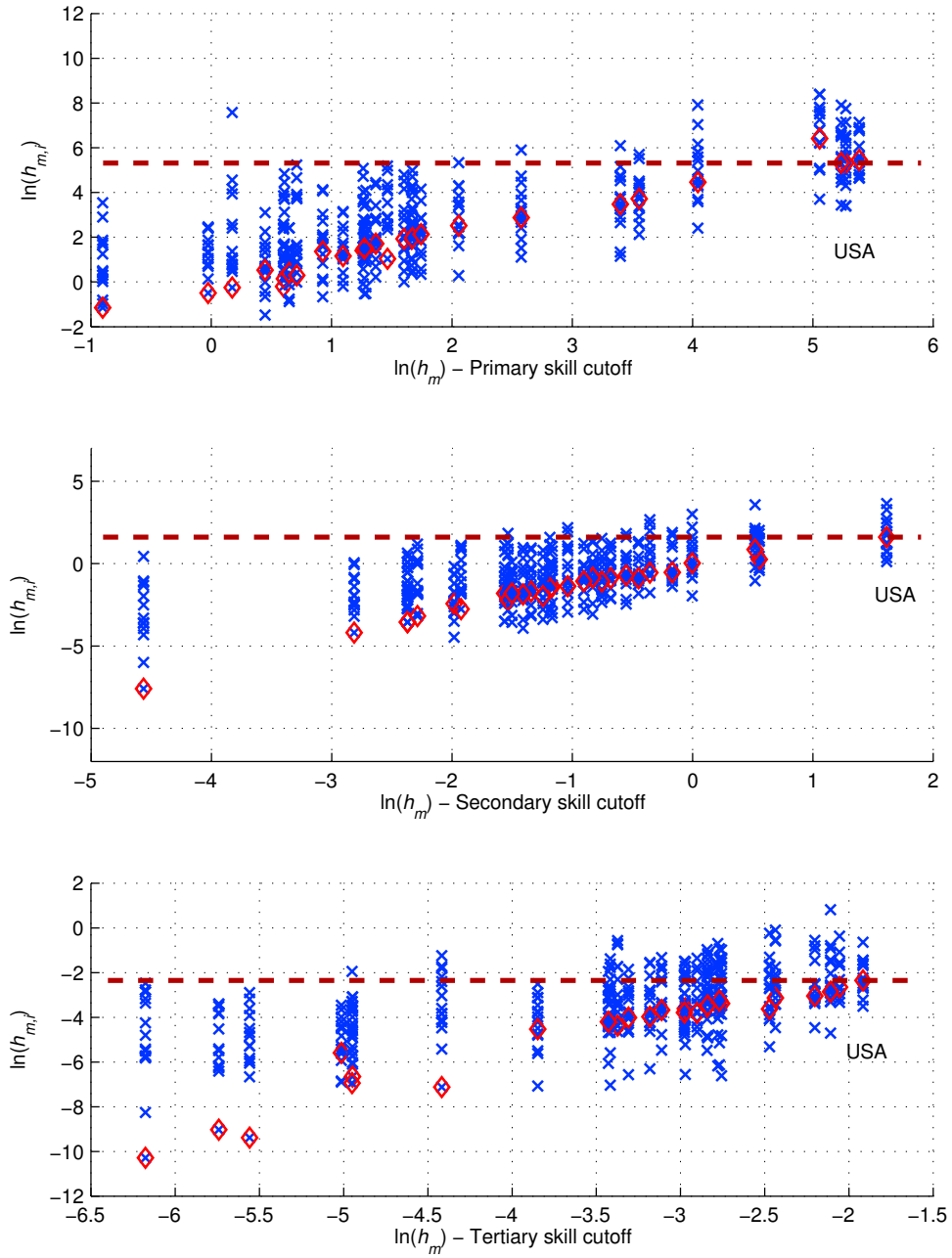


Figure 3: Range of  $h_{m,i}$  by country.

The figure shows the skill intensities of all broad industries in every country in the sample. The red diamonds indicate population weighted median values of  $\ln(h_{m,i})$  for each country. The dashed line shows the same value for the U.S. The solid line . Each plot corresponds to one skill definition.

Table 5: Time series decomposition of  $H/L$  changes.

	Avg.gap	Within gap	Within fraction	$N$
Primary	4.66	4.09	0.91	17
Secondary	3.93	3.74	0.97	17
Tertiary	2.39	1.96	0.77	17
$S_m$	3.02	2.36	0.78	17

Notes: The table shows the fraction of time series variation in  $h_{t,m}$  due to within industry changes in  $h_{t,m,i}$ . The last row shows time series changes in average years of schooling. For notes see Table 2.

**Results.** Table 5 shows the average contributions of within industry changes in  $h_{t,m,i}$ . It is the time series analog to Table 2. Consistent with the cross-sectional findings, the majority of education changes occur within industries. The average fractions range from 0.77 to 0.97, depending on the skill definition. Relative to the cross-sectional results, the within industry contributions are somewhat larger. Similar findings are obtained when educational attainment is measured by average years of schooling, as shown in the last row of Table 5.

Figure 4 shows country level results. It plots within industry changes against country-level changes in skill intensities,  $\ln(h_{2,m}) - \ln(h_{1,m})$ . For most countries, within industry changes account for more than 50%, or at times even for more than 100%, of the country level changes in skill intensities.

A related literature on skill-biased technical change also finds that the bulk of skill upgrading occurs within industries.<sup>8</sup> My findings are not directly comparable with this literature. Almost all of the existing evidence is limited to manufacturing industries. Lacking data on education, the literature generally uses the shares of non-production employment or wage bills as proxies for skill intensities.

Perhaps closest to my findings is Desjonquieres, Machin, and van Reenen (1999) who present data for eight high-income countries, mostly during the 1980s. Their data cover industries outside of manufacturing. Desjonquieres, Machin, and van Reenen (1999) find that the majority of skill upgrading occurs within industries. Autor, Katz, and Krueger (1998) examine U.S. data over the same time period covered by my data. They find that 87% of the increase in the fraction of college employment occurs within industries. The corresponding fraction for my decomposition is 70%.

For non-OECD countries, the literature covers only manufacturing industries. Berman and Machin (2000a) present data for non-production wage bill shares in 37 countries during the 1970s and 1980s. On average, 82% of skill upgrading occurs within industries.

**Group results.** It is of interest to focus on countries which experienced the largest increases in education. I therefore divide countries into three groups, based on their changes in  $h_{t,m}$  over time. Group 1 countries experience the largest increases in  $h_{t,m}$ .

Table 6 shows the average contributions of within industry changes for each group. In all cases but one (group 3 with the college skill cutoff), at least three quarters of the attainment increases occur within industries. One useful statistic summarizing the time series results is: Of the largest increases in education (in group 1), at least 79% were achieved through within industry skill upgrading.

**Within industry education changes.** One reason why within industry changes in attainment are important is that countries with rising education tend to upgrade the skill intensities of most industries.

<sup>8</sup>The seminal study is Berman, Bound, and Griliches (1994).

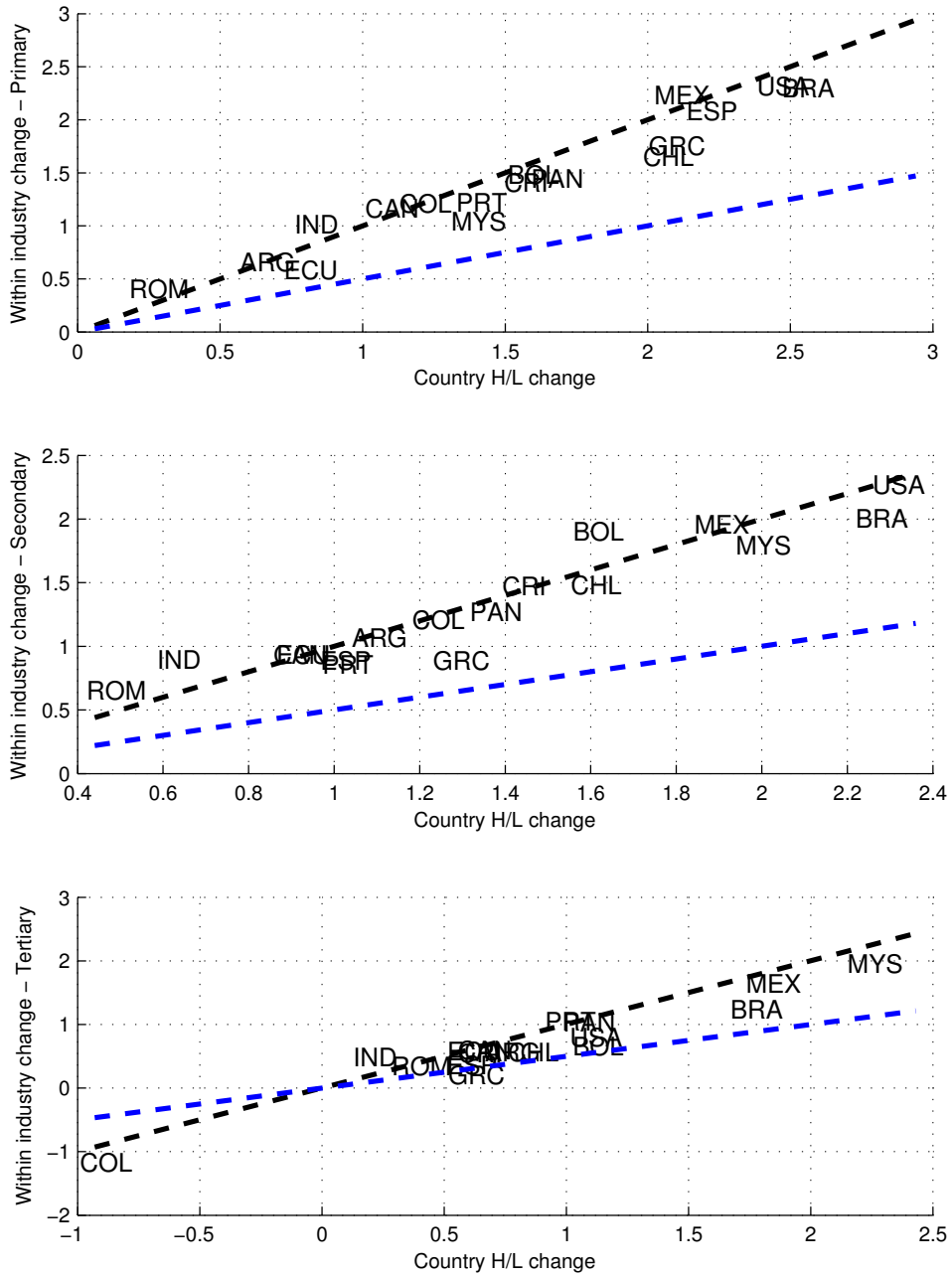


Figure 4: within industry and country level changes in  $H/L$ .

The figure shows within industry education changes and country-level education changes,  $\ln(h_{2,m}/h_{2,m})$ . Each panel represents one skill definition. Also shown are lines with slopes 0.5 and 1. Between these lines, within industry changes account for between 50% and 100% of country-level education changes.

Table 6: Direct decomposition of time series education changes.

Primary skill cutoff	Avg.change	Within industry gap	Fraction
Group 1 BRA, USA, ESP, MEX, GRC, CHL	9.58	7.81	0.91
Group 2 PAN, BOL, CRI, PRT, MYS, COL	4.41	3.67	0.88
Group 3 CAN, IND, ECU, ARG, ROM	2.10	2.14	1.03
Secondary skill cutoff			
Group 1 USA, BRA, MYS, MEX, BOL, CHL	7.08	6.68	0.97
Group 2 CRI, PAN, GRC, COL, ARG, PRT	3.49	3.08	0.90
Group 3 ESP, ECU, CAN, IND, ROM	2.23	2.36	1.07
Tertiary skill cutoff			
Group 1 MYS, MEX, BRA, USA, BOL, PAN	4.66	3.38	0.79
Group 2 PRT, CHL, ARG, CAN, CRI, GRC	2.16	1.79	0.76
Group 3 ECU, ESP, ROM, IND, COL	1.22	1.12	0.60

Notes: The table shows the fraction of time series variation in  $h_{t,m}$  due to within industry changes in  $h_{t,m,i}$ . Countries are divided into groups according to their changes in  $h_{t,m}$ . The table shows average education changes, and the average fraction due to within industry changes.

Table 7: within industry and aggregate changes in  $H/L$ .

Skill cutoff	Avg. $\beta$	Avg. $R^2$	Avg. N	N industries
Primary	0.96 (0.20)	0.65 (0.23)	16.7	15
Secondary	0.93 (0.13)	0.54 (0.18)	16.8	15
Tertiary	0.97 (0.24)	0.64 (0.21)	16.4	15

Notes: The table shows the results of regressing, by industry,  $\ln(h_{2m,i}) - \ln(h_{1m,i})$  on  $\ln(h_{2,m}) - \ln(h_{1,m})$ . See Table 4 for notes.

To demonstrate this point, I regress, by industry, within industry changes ( $\ln(h_{2m,i}) - \ln(h_{1m,i})$ ) on country level changes ( $\ln(h_{2,m}) - \ln(h_{1,m})$ ) in attainment. Table 7 shows the average slope coefficients and  $R^2$  statistics of these regressions.

The results are consistent with the cross-sectional analysis (Table 4). The average slope coefficients are close to unity (albeit with large standard deviations). The  $R^2$  statistics are on the order of 0.6.

**Summary.** The evidence presented so far suggests that the majority of cross-country education gaps is due to within industry education gaps. Similarly, the majority of education changes within countries over time is due to within industry skill upgrading. Product mix makes a smaller contribution. Exactly how large the contribution of within industry gaps is depends on the definition of skilled labor. Focusing on the countries with the largest education gaps: If these countries upgraded their within industry skill levels to U.S. levels, at least 71% of education gaps would vanish (Table 3). Similarly, of the largest increases in education, at least 79% were achieved through within industry upgrading (Table 6).

## 4 Detailed Industries

Models of development commonly focus on broad sectoral shifts from unskilled ("agriculture") to skilled sectors ("industry").<sup>9</sup> This focus is motivated by the large decline in the employment share of agriculture observed during the process of development. Even though the data presented so far distinguish only 15 industries, they are sufficient to speak to this class of models.

However, one concern is that aggregating multiple products into one industry may hide some industry specialization. It is possible that countries differ in the mix of products produced *within* one of the 15 broad sectors. It is therefore of interest to ask how the previous results change when more detailed industry data are used.

**Detailed industry data.** The IPUMS variable IND contains more detailed industry information. The level of detail varies across countries and year. For example, in the year 2000, the U.S. sample distinguishes 265 industries.

An important drawback of using this data is that each country delineates industries differently. However, for most samples it is possible to map a large subset of industries into a common classification of 62 categories. The common classification is taken from the CHL 2002 sample. The reason is that several countries use classifications that are nearly identical to Chile's. The mapping of industries is often not one-to-one. At times, one of country  $m$ 's industries comprises several of Chile's industries. These cases are dropped. In other cases, one or more of country  $m$ 's industries comprise only part of a Chilean industry.

<sup>9</sup>Examples include Hansen and Prescott (2002) or Gollin, Parente, and Rogerson (2007).

These cases are retained. While it is not possible to map all industries into the common classification, the mapped industries typically account for around 90% of employment.<sup>10</sup> Clearly, mapping industries based on descriptive labels introduces classification errors. Below, I provide indirect evidence that the mapping is not too noisy.

A second drawback of using detailed industry data is that the decomposition developed in section 3.1.1 cannot be calculated. The reason is that sector weights ( $n_{m,i}$ ) cannot be computed for all industries. To address this problem, I develop an alternative decomposition of education gaps which can be implemented with incomplete industry coverage.

#### 4.1 Industry Neutral Model

This section develops the alternative decomposition and compares it with the direct decomposition of section 3. The idea is to infer within industry education gaps for industries that cannot be mapped into the common classification from those that can be. The decomposition exploits the finding of section 3.1.3 that within industry gaps are correlated across industries. I show that educational attainment is consistent with a model of industry neutral differences in skilled labor productivity.

The decomposition is based on a simple labor demand model which extends CC's model to multiple sectors. Each industry is endowed with a production function of the form

$$y_{m,i} = A_{m,i} \left[ \gamma_{m,i} H_{m,i}^\rho + (1 - \gamma_{m,i}) L_{m,i}^\rho \right]^{1/\rho} \quad (9)$$

where  $A_{m,i}$  is a neutral productivity term and  $\gamma_{m,i}$  determines the relative productivity of skilled labor. If labor is paid its marginal product, the labor demand equation is given by

$$\ln \left( \frac{w_{H,m}}{w_{L,m}} \right) = \phi_{m,i} + (\rho - 1) \ln(h_{m,i}) \quad (10)$$

where

$$\phi_{m,i} = \ln \left( \frac{\gamma_{m,i}}{1 - \gamma_{m,i}} \right) \quad (11)$$

measures the relative productivity of skilled labor in industry ( $m, i$ ).

If estimates of  $\phi_{m,i}$  were available for all industries, they could be used to predict  $h_{m,i}$  and thus to decompose cross-country education gaps. When using detailed industry data,  $\phi_{m,i}$  can be estimated from the factor demand equation (10). However, this is possible only for industries that are mapped into the common classification. To infer  $\phi_{m,i}$  in the remaining industries, I assume that cross-country differences in  $\phi_{m,i}$  are *industry neutral* in the sense of Treffer (1993). Industry neutrality means that

$$\phi_{m,i} = \alpha_m + \beta_i. \quad (12)$$

The relative productivity of skilled labor ( $\phi_{m,i} - \phi_{\hat{m},i}$ ) is the same in all industries for a given country pair. I call  $\alpha_m$  country  $m$ 's *skill bias*.

Treffer (1993) shows that industry neutral differences in factor augmenting productivity help account for observed international trade flows. I show below that a model with industry neutral skill bias differences accounts for a large share of the cross-country and cross-industry variation in  $\phi_{m,i}$ .

---

<sup>10</sup>Table 17 in the Appendix lists the number of detailed industries that can be mapped into the common classification for all samples. A table listing the mapping from country specific to CHL 2002 industries is available at [www.lhendricks.org](http://www.lhendricks.org).

If skill bias differences are industry neutral, a natural decomposition of cross-country education gaps is based on the identity

$$\frac{H_m}{L_m} = \sum_i n_{m,i} \frac{H_{m,i}}{L_{m,i}} \quad (13)$$

$$= \sum_i n_{m,i} \exp \left[ \frac{\alpha_m + \beta_i - \ln \left( \frac{w_{H,m}}{w_{L,m}} \right)}{1 - \rho} \right] \quad (14)$$

so that

$$\ln(h_m) = \frac{\alpha_m - \ln(w_{H,m}/w_{L,m})}{1 - \rho} + \ln \sum_i n_{m,i} \exp \left( \frac{\beta_i}{1 - \rho} \right) \quad (15)$$

This suggests to decompose cross-country gaps into three components:

$$\ln(h_{US}) - \ln(h_m) = \Delta W_m + \Delta \alpha_m + \Delta L_m + \varepsilon_m \quad (16)$$

where

$$\Delta W_m = \frac{\ln \left( \frac{w_{H,m}}{w_{L,m}} \right) - \ln \left( \frac{w_{H,US}}{w_{L,US}} \right)}{1 - \rho} \quad (17)$$

measures the contribution of skill premiums,

$$\Delta \alpha_m = \frac{\alpha_{US} - \alpha_m}{1 - \rho} \quad (18)$$

measures the role of skill bias, and

$$\Delta L_m = \ln \left( \sum_i n_{US,i} \exp \left( \frac{\beta_i}{1 - \rho} \right) \right) - \ln \left( \sum_i n_{m,i} \exp \left( \frac{\beta_i}{1 - \rho} \right) \right) \quad (19)$$

measures the contribution of industry composition.  $\varepsilon_m$  is a residual which captures model misspecification and measurement error.  $\Delta L_m$  can only be computed if industry coverage is complete. However,  $\alpha_m$  can be estimated even with incomplete industry data.

## 4.2 Estimation

Given measures of  $H_{m,i}$ ,  $L_{m,i}$  and skill premiums, I estimate the parameters of the industry neutral model,  $\alpha_m$  and  $\beta_i$ , as follows. Based on the evidence discussed in CC, I set the elasticity of substitution to  $(1 - \rho)^{-1} = 1.4$ . For each skill definition, I calculate  $\phi_{m,i}$  from (10) for the subset of industries with available data. I then regress  $\phi_{m,i}$  on country and industry dummies using weighted least squares. Thus,  $\alpha_m$  and  $\beta_i$  minimize

$$\sum_m \sum_i n_{m,i} [\phi_{m,i} - \alpha_m - \beta_i]^2 \quad (20)$$

If detailed industries are used, the industry coverage of countries varies. Only industries present in at least 10 countries are retained.

If skill premiums vary across industries, the estimates of  $\phi_{m,i}$  are biased. Note, however, that country specific variation in skill premiums does not bias  $\phi_{m,i}$ ; it was removed in the estimation of  $\phi_{m,i}$  using (10). Therefore, (20) yields biased estimates of  $\beta_i$ , which play no role for the paper's results, whereas there is no reason to expect the estimated  $\alpha_m$  to be biased.

Table 8: Fraction of variation in  $\phi_{m,i}$  accounted for by the industry neutral model.

Skill cutoff	Broad industries			Detailed industries		
	$R_i^2$	$R_m^2$	D.f.	$R_i^2$	$R_m^2$	D.f.
Primary	0.90	0.84	369	0.85	0.75	1107
Secondary	0.79	0.86	370	0.73	0.80	1171
Tertiary	0.63	0.76	369	0.40	0.72	1123

Notes: The table shows the means of  $R_i^2$  across industries and of  $R_m^2$  across countries. D.f. denotes the degrees of freedom in the regression of  $\phi_{m,i}$  on industry and country dummies. Detailed industries and cross-country data are used.

### 4.3 Cross-country Results

This section presents the decomposition of cross-country education gaps based on the industry neutral model using detailed industry data. Before presenting the results, it is useful to verify that the industry neutral model accounts well for the data and yields similar results to the decomposition of section 3 when broad industries are used.

#### 4.3.1 Evaluating the industry neutral model

To evaluate whether the industry neutral restriction offers a good approximation to the data, I compare the predicted values of  $\phi_{m,i}$  with their observed counterparts. The fraction of  $\phi_{m,i}$  variation in industry  $i$  which is accounted for by the model is measured by the pseudo  $R^2$

$$R_i^2 = 1 - \frac{\sum_m (\phi_{m,i} - \alpha_m - \beta_i)^2}{\sum_m \left( \phi_{m,i} - \frac{1}{M_i} \sum_m \phi_{m,i} \right)^2} \quad (21)$$

The numerator is the model's residual sum of squares. The denominator measures the total sum of squares.  $M_i$  denotes the number of countries with data for industry  $i$ . The values of  $\phi_{m,i}$  are calculated from the labor demand equation (10). Predicted values of  $\phi_{m,i}$  are given by  $\alpha_m + \beta_i$ . The pseudo  $R^2$  for cross industry variation within country  $m$ ,  $R_m^2$ , is defined analogously.

Table 8 shows that the industry neutral model accounts for 63% to 90% of the variation in  $\phi_{m,i}$  across countries and industries. The sole exception is  $R_i^2$  for detailed industries with the tertiary skill cutoff. This is a recurring finding. Possibly due to larger measurement error in labor inputs, the industry neutral model generally accounts less well for the data when the tertiary skill cutoff is used.

An alternative way of evaluating the industry neutral model compares the predicted and observed values of educational attainment,  $h_{m,i}$ . The predicted values,  $h_{m,i}^{pred}$ , are computed from the firm's labor demand equation (10), imposing the industry neutral assumption  $\phi_{m,i} = \alpha_m + \beta_i$ .

Table 9 shows pseudo  $R^2$  values for industries and countries. Industry  $i$ 's pseudo  $R^2$  is defined as

$$R_i^2 = 1 - \frac{\sum_m \left( \ln(h_{m,i}) - \ln(h_{m,i}^{pred}) \right)^2}{\sum_m \left( \ln(h_{m,i}) - \frac{1}{M_i} \sum_m \ln(h_{m,i}) \right)^2} \quad (22)$$

The country pseudo  $R_m^2$  is defined analogously.

Table 9: Fraction of variation in  $h_{m,i}$  accounted for by the industry neutral model

Skill cutoff	Broad industries		Detailed industries	
	$R_i^2$	$R_m^2$	$R_i^2$	$R_m^2$
Primary	0.91	0.84	0.86	0.75
Secondary	0.79	0.86	0.74	0.80
Tertiary	0.63	0.76	0.44	0.72

Notes: The table shows the fraction of variation in  $h_{m,i}$  across countries and industries accounted for by the industry neutral model.

The industry neutral model generally accounts for at least 72% of the variation in  $\ln(h_{m,i})$  across industries and countries. The exception is again the tertiary skill cutoff, where the model accounts for less than one-half of the variation. I conclude that the industry neutral model offers a useful approximation to the data. The good fit of the model also suggests that errors in the mapping of industries are not too large.

### 4.3.2 Comparison with the direct decomposition

A useful consistency check is to ask whether the decomposition of cross-country education implied by the industry neutral model is similar to the decomposition of section 3. Since the direct decomposition requires complete industry coverage, this comparison can only be implemented for broad industries.

Table 10 compares the two decompositions. For the industry neutral model, the table reports the average contributions of skill bias ( $\Delta\alpha_m$ ), skill premiums ( $\Delta W_m$ ), and product mix ( $\Delta L_m$ ). The within industry component is given by  $\Delta\alpha_m + \Delta W_m$ . As before, average contributions are defined as simple averages of the country level contributions. For example, the average contribution of skill bias is defined as  $\exp(M^{-1} \sum \Delta\alpha_m)$ . The fraction of cross-country education gaps due to skill bias is given by  $\sum \Delta\alpha_m / \sum (\ln(h_{US}) - \ln(h_m))$ .

Within industry gaps account for 80% to 87% of cross-country gaps. The bulk of these are due to skill bias differences, not to variation in skill premiums. This result does not change when country specific Mincer returns are used to estimate skill premiums and labor inputs. For comparison, the direct decomposition attributes between 67% and 82% of cross-country education gaps to within industry gaps (see Table 2).<sup>11</sup>

Figure 5 compares the direct and the industry neutral decompositions. For each country, it shows the within industry education gaps obtained from both decompositions. The two decompositions are similar for the majority of countries. I conclude that the industry neutral model offers a useful method for decomposing cross-country education gaps using detailed industry data.

### 4.3.3 Results

Table 11 shows the industry neutral decomposition for detailed industry data. The findings are quite similar to those for broad industries in Table 10. The fractions of cross-country education gaps due to within industry gaps are very similar. This similarity continues to hold when countries are divided into groups (as in Table 3). The details are not shown to conserve space.

<sup>11</sup>The industry neutral model's decomposition is close to that of Experiment II developed in section 3.1.1.

Table 10: Industry neutral decomposition. Broad industries.

Skill cutoff	Avg.gap	$\Delta\alpha_m$	$\Delta W_m$	$\Delta L_m$	$\Delta\alpha_m + \Delta W_m$
Primary	31.90	15.14 ( 0.78)	1.04 ( 0.01)	2.32 ( 0.24)	15.79 ( 0.80)
Secondary	17.62	13.50 ( 0.91)	0.90 (-0.03)	1.81 ( 0.21)	12.21 ( 0.87)
Tertiary	4.89	3.72 ( 0.83)	0.99 (-0.00)	1.74 ( 0.35)	3.70 ( 0.82)

Notes: The table shows the industry neutral decomposition of cross-country education gaps for broad industries. Each column shows the average of a cross-country education gap. Fractions of average education gaps are shown in parentheses.

Table 11: Industry neutral decomposition of cross-country education gaps. Detailed industries.

Skill cutoff	Avg.gap	$\Delta\alpha_m$	$\Delta W_m$	$\Delta L_m$	$\Delta\alpha_m + \Delta W_m$
Primary	38.41	17.04 ( 0.78)	1.04 ( 0.01)	1.00 ( 0.00)	17.75 ( 0.79)
Secondary	18.21	13.48 ( 0.90)	0.90 (-0.04)	1.00 ( 0.00)	12.09 ( 0.86)
Tertiary	4.70	3.22 ( 0.76)	1.00 (-0.00)	1.00 ( 0.00)	3.22 ( 0.76)

Notes: The table shows the industry neutral decomposition of cross-country education gaps for detailed industries. See Table 10 for notes.

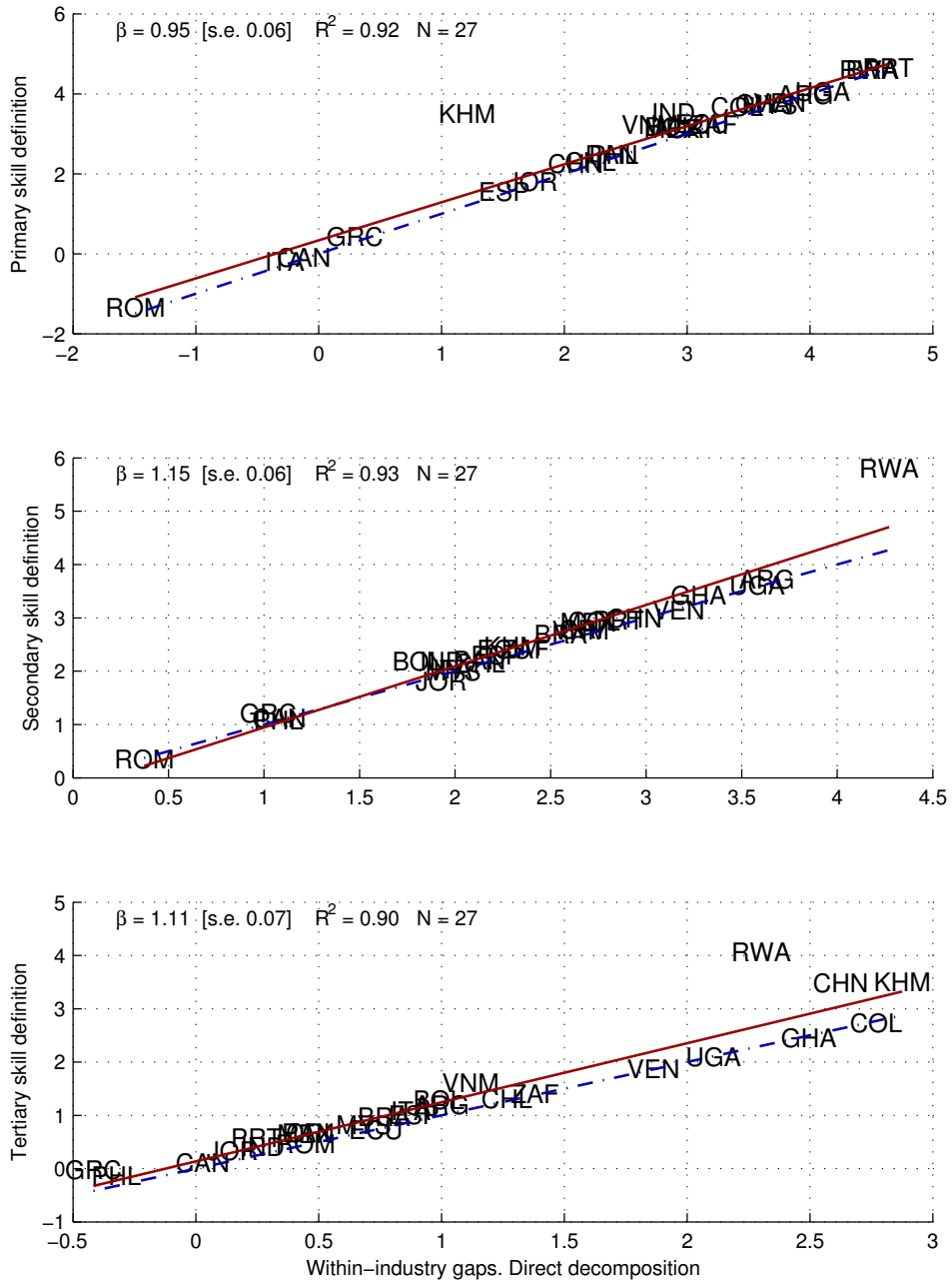


Figure 5: Direct and industry neutral decomposition.

The figure shows the within industry education gaps obtained from the direct decomposition and from the industry neutral decomposition. Also shown are a 45-degree line (dashed) and an OLS regression line (solid). Each panel represents one skill definition.

Table 12: Industry detail and within industry education gaps.

Skill cutoff	Broad industries			Detailed industries		
	Avg.gap	Within gap	Fraction	Avg.gap	Within gap	Fraction
Primary	28.2	14.3	0.80	33.0	15.7	0.79
Secondary	15.9	11.2	0.87	16.1	10.9	0.86
Tertiary	4.6	3.5	0.82	4.4	3.1	0.76

Notes: The table compares the fraction of country education gaps that are due to within industry gaps for broad and detailed industries. Avg.gap is the cross-country average of  $\ln(h_{US}) - \ln(h_m)$ . "Within gap" denotes the average within industry gap. "Fraction" is the average fraction due to within industry gaps.

In order to show clearly how more industry detail affects the contribution of within industry gaps, Table 12 compares the industry neutral decompositions for broad and detailed industries. Except for the "tertiary" skill cutoff, the results are quite close. This suggests that industry specialization within the 15 broad sectors is not an important concern. It is, of course, possible that specialization occurs at an even finer level of industry detail. However, an earlier version of this paper arrived at similar conclusions using the 265 industry classification of the year 2000 U.S. sample.

#### 4.4 Time Series Decomposition

This section applies the industry neutral decomposition to the changes in education within countries over time. The decomposition is analogous to the cross-sectional one. The only difference is that cross-country differences, such as  $\alpha_{US} - \alpha_m$ , are replaced by time series changes, such as  $\alpha_{2,m} - \alpha_{1,m}$ .

The model parameters are estimated as follows. For date 2, the cross-sectional estimates are used ( $\alpha_{2,m}$ ,  $w_{2,H,m}/w_{2,L,m}$ , and  $\beta_i$ ). I assume that industry skill intensities ( $\beta_i$ ) and skill premiums are time invariant. The latter assumption reflects the lack of good time series data for Mincer returns. The values of  $\alpha_{1,m}$  are estimated by minimizing the date 1 deviation between actual and predicted skill bias parameters,  $\sum_i (\phi_{1,m,i} - \alpha_{1,m} - \beta_i)^2$ , where  $\phi_{1,m,i}$  solves the factor demand equation (10). This is analogous to the cross-sectional estimation, except that the values of  $\beta_i$  are predetermined.

##### 4.4.1 Evaluating the industry neutral model

Before presenting the findings, I examine how well the industry neutral model accounts for the time series data. Table 13 compares the predicted values of  $h_{1,m,i}$  with the observed ones. For each industry and country, a pseudo  $R^2$  is calculated, analogously to (22) for the cross-section. Table 13 reports the means of these pseudo  $R^2$  values. Consistent with the cross-sectional results, the model fits the data well, except for the "tertiary" skill cutoff.

Table 14 examines how well the industry neutral model accounts for the changes of  $h_{t,m,i}$  over time. For each industry, I regress predicted against observed changes in  $h_{t,m,i}$ . The predicted changes are the same for all industries:  $\frac{\alpha_{2,m} - \alpha_{1,m}}{1 - \rho}$ . The table reports the average regression slope and the average pseudo  $R^2$  for each skill cutoff.

The industry neutral model accounts for more than half of the changes, except for the "tertiary" skill cutoff. The average slope coefficients are around 0.65, compared with a model prediction of 1. Attenuation bias may reduce the slope coefficients.

Table 13: Fraction of date 1 variation in  $h_{m,i}$  accounted for by the industry neutral model.

Skill cutoff	Broad industries		Detailed industries	
	$R_i^2$	$R_m^2$	$R_i^2$	$R_m^2$
Primary	0.92	0.82	0.90	0.77
Secondary	0.69	0.80	0.70	0.75
Tertiary	0.27	0.77	0.18	0.68

Notes: The table shows the fraction of variation in  $h_{1,m,i}$  across countries and industries accounted for by the industry neutral model.  $R_i^2$  denotes the average pseudo  $R^2$  comparing predicted and observed  $h_{1,m,i}$  within industries, across countries.  $R_m^2$  is the corresponding  $R^2$  comparing  $h_{1,m,i}$  within countries, across industries.

Table 14: Predicted / actual changes in  $H/L$ . Industry neutral.

Skill cutoff	Pseudo $R^2$	Avg.Slope (s.e.)
Primary	0.54 (0.39)	0.71 (0.21)
Secondary	0.33 (0.43)	0.61 (0.21)
Tertiary	0.35 (0.54)	0.65 (0.20)

Notes: The table compares predicted and actual changes in  $h_{t,m,i}$ . The predictions are those of the industry neutral model. The columns report the average pseudo  $R^2$  and the average regression slope. Standard deviations are shown in parentheses.

#### 4.4.2 Comparison with the direct decomposition

The final consistency check asks whether the industry neutral model decomposition is similar to the direct decomposition of section 3.2. This comparison can only be implemented for broad industries.

Table 15 compares the average fractions due to within industry changes in education implied by the two decompositions. For the industry neutral model, within industry changes are given by  $\frac{\alpha_{2,m} - \alpha_{1,m}}{1 - \rho}$ . This contains changes in skill bias and skill premiums.

As in the cross-sectional case, the industry neutral model generally overstates the fraction of education changes due to within industry changes. The gap is particularly large for the "tertiary" skill definition, where the industry neutral model generally does not perform well. However, in contrast to the cross-sectional case, the industry neutral model also overstates the within industry contributions relative to Experiment II. This should be kept in mind when interpreting the results for detailed industries.

Table 15: Comparison of industry neutral and direct decomposition. Time series data.

Skill cutoff	Ind.neutral	Direct	$N$
Primary	0.97	0.91	17
Secondary	1.12	0.97	17
Tertiary	1.06	0.77	17

Notes: The table shows the fraction of  $h_{t,m}$  changes due to within industry changes. It compares the industry neutral model with the direct decomposition.  $N$  is the number of countries with data.

Table 16: Industry neutral decomposition of time series education changes.

Skill cutoff	Broad industries			Detailed industries		
	Avg.gap	Within gap	Fraction	Avg.gap	Within gap	Fraction
Primary	4.7	4.5	0.97	5.2	5.3	1.01
Secondary	3.9	4.6	1.12	4.2	5.4	1.17
Tertiary	2.4	2.5	1.08	2.8	3.1	1.11

Notes: The table compares the industry neutral decomposition of changes in  $h_{t,m}$  for broad and detailed industries. For notes see Table 12.

### 4.4.3 Results

Even though the industry neutral model overstates the role of within industry changes, it may be used to investigate how more detailed industry data affect the results. Table 16 shows the average contributions of within industry changes for the broad and the detailed industry classifications. The figures are calculated similarly to the cross-sectional results. The average contribution of within industry changes is given by  $M^{-1} \sum \frac{\alpha_{m,2} - \alpha_{m,1}}{1 - \rho}$ . The average total change in education is  $M^{-1} \sum (\ln(h_{2,m}) - \ln(h_{1,m}))$ .

The finer industry detail increases the role of within industry changes in the time series, but only slightly. There is no indication that the large within industry changes found throughout this paper are due to lack of industry detail in the data.

## 5 Conclusion

This paper shows that at least two-thirds of the variation in education across countries and over time occurs within industries. At most one-third is due to variation in industry composition. The data are broadly consistent with a model of industry neutral differences in skilled labor productivity. The role of within industry variation in education does not diminish when a finer industry classification is used.

These findings are consistent with theories that emphasize skill upgrading within industries, such as theories of skill-biased technical change (e.g., Acemoglu 2002, 2003). Future research should investigate why variation in education occurs in an industry neutral manner.

## References

- ACEMOGLU, D. (2002): “Directed technical change,” *Review of Economic Studies*, 69, 781–810.
- (2003): “Patterns of skill premia,” *Review of Economic Studies*, 70, 199–230.
- AUTOR, D. H., L. F. KATZ, AND A. B. KRUEGER (1998): “Computing Inequality: Have Computers Changed the Labor Market?,” *The Quarterly Journal of Economics*, 113(4), 1169–1213.
- BANERJEE, A., AND E. DUFLO (2005): “Growth theory through the lens of development,” in *Handbook of Economic Growth*, ed. by P. Aghion, and S. N. Durlauf, vol. 1A, pp. 473–554. Elsevier.
- BARRO, R. J., AND J.-W. LEE (2001): “International Data on Educational Attainment: Updates and Implications,” *Oxford Economic Papers*, 53(3), 541–563.

- BERMAN, E., J. BOUND, AND Z. GRILICHES (1994): “Changes in the Demand for Skilled Labor within U.S. Manufacturing: Evidence from the Annual Survey of Manufacturers,” *The Quarterly Journal of Economics*, 109(2), 367–397.
- BERMAN, E., J. BOUND, AND S. MACHIN (1998): “Implications of Skill-Biased Technological Change: International Evidence,” *The Quarterly Journal of Economics*, 113(4), 1245–1279.
- BERMAN, E., AND S. MACHIN (2000a): “Skill-Biased Technology Transfer Around the World,” *Oxford Review of Economic Policy*, 16(3), 12–22.
- (2000b): “Skill-Biased Technology Transfer: Evidence of Factor Biased Technological Change in Developing Countries,” Mimeo. Boston University.
- BILS, M., AND P. J. KLENOW (2000): “Does Schooling Cause Growth?,” *The American Economic Review*, 90(5), 1160–1183.
- CASELLI, F., AND W. J. COLEMAN (2001): “The U.S. Structural Transformation and Regional Convergence: A Reinterpretation,” *The Journal of Political Economy*, 109(3), 584–616.
- (2006): “The World Technology Frontier,” *American Economic Review*, 96(3), 499–522.
- CORDOBA, J. C., AND M. RIPOLL (2009): “Schooling and Development: The Role of Credit Limits, Public Education, Fertility and Mortality,” Mimeo. Iowa State University.
- DESJONQUERES, T., S. MACHIN, AND J. VAN REENEN (1999): “Another Nail in the Coffin? Or Can the Trade Based Explanation of Changing Skill Structures Be Resurrected?,” *The Scandinavian Journal of Economics*, 101(4), 533–554.
- DOMS, M., AND E. LEWIS (2006): “Labor Supply and Computer Adoption,” Mimeo. Federal Reserve Bank of San Francisco.
- EROSA, A., T. KORESHKOVA, AND D. RESTUCCIA (2007): “How Important is Human Capital? A Quantitative Theory of World Income Distribution,” Mimeo. University of Toronto.
- GALOR, O. (2000): “Ability Biased Technological Transition, Wage Inequality, and Economic Growth,” *Quarterly Journal of Economics*, 115, 469–498.
- (2005): “From Stagnation to Growth: Unified Growth Theory,” in *Handbook of Economic Growth*, ed. by P. Aghion, and S. N. Durlauf, vol. 1A, pp. 171–293. Elsevier.
- GALOR, O., AND D. N. WEIL (2000): “Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and beyond,” *The American Economic Review*, 90(4), 806–828.
- GOLDIN, C. (1998): “America’s Graduation from High School: The Evolution and Spread of Secondary Schooling in the Twentieth Century,” *The Journal of Economic History*, 58(2), 345–374.
- (2001): “The Human-Capital Century and American Leadership: Virtues of the Past,” *The Journal of Economic History*, 61(2), 263–292.
- GOLDIN, C., AND L. F. KATZ (2008): *The Race between Education and Technology*. Harvard University Press.

- GOLLIN, D., S. PARENTE, AND R. ROGERSON (2002): “The Role of Agriculture in Development,” *The American Economic Review*, 92(2), 160–164.
- GOLLIN, D., S. L. PARENTE, AND R. ROGERSON (2007): “The Food Problem and the Evolution of International Income Levels,” *Journal of Monetary Economics*, 54, 1230–1255.
- HALL, P. K., R. MCCAA, AND G. THORVALDSEN (eds.) (2000): *Handbook of International Historical Microdata for Population Research*. The Minnesota Population Center.
- HALL, R. E., AND C. I. JONES (1999): “Why do some countries produce so much more output per worker than others?,” *Quarterly Journal of Economics*, 114, 83–116.
- HANSEN, G. D., AND E. C. PRESCOTT (2002): “Malthus to Solow,” *American Economic Review*, 92(4), 1205–1217.
- HARRIS, J. R., AND M. P. TODARO (1970): “Migration, Unemployment and Development: A Two-Sector Analysis,” *The American Economic Review*, 60(1), 126–142.
- HENDRICKS, L. (2002): “How Important Is Human Capital for Development? Evidence from Immigrant Earnings,” *The American Economic Review*, 92(1), 198–219.
- (2008): “Why Does Education Differ Across Countries?,” Mimeo. Iowa State University.
- (2010): “Validation of IPUMS International Industry and Education Data,” Mimeo. University of North Carolina.
- (2011): “The Skill Composition of U.S. Cities,” *International Economic Review*, (Forthcoming).
- KABOSKI, J. P. (2006): “Endogenous Mincerian Returns: Explaining Cross-country Variation in the Returns to Schooling,” Mimeo: Ohio State University.
- KLENOW, P. J., AND A. RODRIGUEZ-CLARE (1997): “The Neoclassical Revival in Growth Economics: Has It Gone Too Far?,” *NBER Macroeconomics Annual*, 12, 73–103.
- LEWIS, W. A. (1954): “Economic Development With Unlimited Supplies of Labor,” *Manchester School of Economics and Social Studies*, 22, 139–151.
- LUCAS, R. E. (2004): “Life Earnings and Rural-urban Migration,” *Journal of Political Economy*, 112(1, part 2), S29–59.
- MACHIN, S., AND J. VAN REENEN (1998): “Technology and Changes in Skill Structure: Evidence from Seven OECD Countries,” *The Quarterly Journal of Economics*, 113(4), 1215–1244.
- MANUELLI, R., AND A. SESHADRI (2007): “Human Capital and the Wealth of Nations,” Mimeo. University of Wisconsin.
- MINNESOTA POPULATION CENTER (2009): “Integrated Public Use Microdata Series - International: Version 5.0,” Machine-readable database.
- RESTUCCIA, D., AND G. VANDENBROUCKE (2008): “The Evolution of Education: A Macroeconomic Analysis,” Mimeo. University of Toronto.

- RESTUCCIA, D., D. T. YANG, AND X. ZHU (2007): “Agriculture and Aggregate Productivity: A Quantitative Cross-Country Analysis,” *Journal of Monetary Economics*, 55(2), 234–250.
- RUGGLES, S., M. L. KING, D. LEVISON, R. MCCAA, AND M. SOBEK (2003): “IPUMS-International,” *Historical Methods*, 36(2), 60–65.
- SCHOELLMAN, T. (2006): “The Causes and Consequences of Cross-country Differences in School Attainment,” Mimeo. Clemson University.
- TREFLER, D. (1993): “International Factor Price Differences: Leontief Was Right!,” *Journal of Political Economy*, 101(6), 961–987.

## A Summary Statistics

Table 17 shows summary statistics by country.  $N$  is the number of individuals in the sample (in thousands). School denotes average years of schooling. The employment shares by education are measured by the fraction of persons having partially or fully completed a given level of schooling (primary, secondary, or tertiary). The detailed industry block shows the number of detailed industries available for each sample (“N”), the number of industries that can be matched to the common classification (“Matched”), and the fraction of employment covered by the matched industries (“Empl.frac.”).

Table 18 shows summary statistics for skilled and unskilled labor. The table shows the estimates of the model variables and parameters for all countries. Each panel represents one skill cutoff,  $j_0$ .  $\alpha_m$  is estimated only for years used in the cross-country analysis. For variable definitions, see the text of the paper.

Table 17: Country statistics.

Country	Year	N	School	Employment shares			Detailed industries		
				Primary	Secondary	Tertiary	N	Matched	Empl.frac.
USA	2000	5458.3	13.6	0.04	0.38	0.58	265	54	87.1
	1960	562.8	10.5	0.31	0.50	0.20	151	50	92.5
ARG	1991	750.3	8.6	0.54	0.31	0.15	66	59	98.6
	1970	148.1	6.6	0.75	0.20	0.04	223	53	81.4
BOL	2001	228.2	8.0	0.55	0.27	0.18	150	55	99.3
	1976	124.7	4.0	0.84	0.12	0.03	147	55	95.4
BRA	2000	3092.9	6.8	0.65	0.24	0.10	221	58	92.4
	1960	872.2	2.1	0.97	0.02	0.01	137	43	84.2
CAN	2001	322.6	13.4	0.04	0.36	0.59	n/a	n/a	n/a
	1981	174.8	12.0	0.14	0.44	0.41	n/a	n/a	n/a
CHL	2002	485.0	10.8	0.30	0.43	0.28	62	62	100.0
	1970	220.3	6.2	0.73	0.23	0.05	165	54	93.0
CHN	1990	1329.3	6.5	0.58	0.40	0.02	328	48	94.3
COL	1993	737.2	6.1	0.64	0.24	0.11	n/a	n/a	n/a
	1973	397.6	3.8	0.86	0.11	0.03	98	50	81.1
CRI	2000	109.8	7.8	0.63	0.21	0.16	173	53	93.7
	1973	48.0	5.2	0.85	0.08	0.06	161	54	95.4
ECU	2001	329.0	7.9	0.58	0.24	0.18	159	58	92.9
	1982	151.7	5.7	0.79	0.13	0.08	n/a	n/a	n/a
ESP	2001	746.3	9.6	0.24	0.58	0.19	60	60	100.0
	1981	581.0	5.6	0.64	0.26	0.10	65	50	90.0
GHA	2000	651.9	5.2	0.63	0.32	0.05	64	56	99.7
GRC	2001	366.2	10.8	0.29	0.40	0.31	224	58	92.8
	1981	315.1	7.7	0.64	0.25	0.11	240	55	96.4
IND	1999	76.3	4.9	0.61	0.30	0.09	159	61	98.2
	1983	197.0	3.0	0.80	0.17	0.03	419	58	45.4
ITA	2001	991.1	10.0	0.12	0.75	0.13	n/a	n/a	n/a
JOR	2004	103.0	10.8	0.16	0.50	0.34	158	51	98.8
KHM	1998	308.9	5.4	0.48	0.52	0.01	159	58	98.1
MEX	2000	2881.9	8.2	0.48	0.38	0.14	117	48	85.9
	1960	79.2	3.7	0.90	0.07	0.02	132	47	91.5
MYS	2000	143.6	8.1	0.30	0.59	0.12	156	61	99.6
	1970	49.2	4.1	0.82	0.17	0.01	223	53	96.9
PAN	2000	86.6	9.2	0.47	0.32	0.21	162	60	99.9
	1970	44.4	5.6	0.79	0.16	0.06	71	47	93.3
PHL	1995	2021.6	9.6	0.40	0.32	0.28	39	30	96.4
PRT	2001	215.9	7.6	0.72	0.14	0.13	228	57	93.9
	1981	160.8	5.3	0.81	0.11	0.07	81	48	89.9
ROM	2002	755.0	10.8	0.08	0.75	0.17	61	61	100.0
	1992	926.9	9.8	0.14	0.75	0.11	100	56	97.3
RWA	2002	300.5	3.1	0.95	0.04	0.01	84	43	99.6
UGA	2002	672.0	5.0	0.79	0.20	0.01	n/a	n/a	n/a
VEN	1990	399.9	7.1	0.51	0.42	0.07	79	51	94.8
VNM	1999	1068.5	6.9	0.72	30	0.25	159	58	97.4
ZAF	2001	698.7	9.0	0.37	0.57	0.06	197	57	91.6

Table 18: Summary statistics by skill cutoff.

Country	Year	Primary skill definition		Secondary skill definition		Tertiary skill definition	
		$H_m/L_m$	$\alpha_m$	$H_m/L_m$	$\alpha_m$	$H_m/L_m$	$\alpha_m$
USA	2000	100.0	0.000	100.0	0.000	100.0	0.000
	1960	8.4	–	9.8	–	32.6	–
ARG	1991	1.5	-2.696	2.7	-2.662	28.2	-0.852
	1970	0.7	–	0.9	–	12.9	–
BOL	2001	1.8	-2.263	7.1	-1.553	22.3	-0.845
	1976	0.4	–	1.4	–	7.5	–
BRA	2000	0.8	-3.288	6.2	-1.986	34.8	-0.765
	1960	0.1	–	0.6	–	6.3	–
CAN	2001	111.7	0.079	34.7	-0.725	87.1	-0.081
	1981	37.3	–	13.9	–	44.7	–
CHL	2002	6.7	-1.637	10.1	-1.462	24.8	-0.930
	1970	0.8	–	2.0	–	10.5	–
CHN	1990	2.2	-1.612	1.9	-2.202	2.6	-2.482
COL	1993	1.3	-2.616	4.2	-2.166	4.5	-2.082
	1973	0.4	–	1.2	–	10.9	–
CRI	2000	2.7	-2.237	5.2	-2.055	42.2	-0.546
	1973	0.6	–	1.2	–	22.1	–
ECU	2001	2.0	-2.311	6.1	-1.786	30.2	-0.456
	1982	0.9	–	2.4	–	16.2	–
ESP	2001	18.0	-1.101	9.4	-1.772	34.7	-0.723
	1981	1.9	–	3.4	–	18.7	–
GHA	2000	0.9	-2.672	1.9	-2.446	4.8	-1.540
GRC	2001	29.2	-0.307	19.9	-0.935	82.4	0.020
	1981	3.6	–	5.4	–	43.8	–
IND	1999	0.9	-2.510	4.3	-1.700	23.3	-0.428
	1983	0.4	–	2.3	–	18.8	–
ITA	2001	96.2	0.140	12.7	-1.595	37.3	-0.768
JOR	2004	15.3	-1.286	16.8	-1.353	75.2	-0.236
KHM	1998	0.6	-2.499	2.9	-2.094	2.2	-2.434
MEX	2000	2.5	-2.231	4.9	-2.151	43.2	-0.459
	1960	0.3	–	0.7	–	6.8	–
MYS	2000	1.9	-2.657	11.5	-1.493	39.6	-0.584
	1970	0.5	–	1.5	–	4.4	–
PAN	2000	4.0	-1.766	8.7	-1.579	42.7	-0.481
	1970	0.7	–	2.2	–	15.2	–
PHL	1995	2.7	-1.639	14.0	-0.841	59.5	0.096
PRT	2001	1.0	-3.181	5.8	-2.237	57.5	-0.479
	1981	0.2	–	2.0	–	20.8	–
ROM	2002	80.4	0.961	33.5	-0.488	39.8	-0.334
	1992	60.9	–	20.5	–	26.5	–
RWA	2002	0.2	-3.289	0.2	-4.209	1.4	-2.906
UGA	2002	0.5	-2.758	1.2	-2.572	4.8	-1.496
VEN	1990	1.8	-2.692	4.4	-2.384	14.5	-1.337
VNM	1999	1.0	-2.168	312.0	-2.036	8.2	-1.154
ZAF	2001	2.9	-2.174	8.1	-1.713	22.1	-1.065