

# 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

February 4, 2009

## Abstract

Educational attainment varies greatly across countries and within countries over time. This paper asks whether the variation in education is primarily due to structural change or to within industry skill upgrading. 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 & Lee 2001).

How do countries absorb dramatically rising skilled labor endowments during the process of industrialization? The literature has developed two potential answers to this question. One answer emphasizes *structural change*. As countries develop, their product mix changes from unskilled labor intensive products, often identified with agriculture, to skilled labor intensive products (Hansen & Prescott 2002; Lucas 2004; Gollin et al. 2007). An alternative class of models highlights skill upgrading *within industries*. Countries with larger skill endowments produce a given product with more skilled labor (Galor & Weil 2000; Galor 2005).

The two views have very different implications. From a theoretical perspective, the skill upgrading interpretation suggests to study models of technology adoption, of skill biased technical change (Acemoglu 2002, 2003), or of ability biased technical change (Galor & Moav 2000). By contrast, the structural change view suggests to study multi sector models with non-balanced growth. From a policy perspective, efforts to raise a country's human capital level could either focus on removing barriers to technology adoption (Parente & Prescott 1994) or on promoting skill-intensive industries.

The purpose of this paper is to shed light on the relative importance of these two views. It decomposes cross-country and time series variation in skilled employment into the contributions of structural change and within industry skill gaps.

**Implementation.** Section 2 constructs a dataset of skilled and unskilled labor inputs for 26 countries and 15 industries. The data are drawn from large, individual level datasets provided by the IPUMS International database (Minnesota Population Center 2008). Workers are divided into skilled and unskilled groups depending on their years of schooling.

Section 3 develops a decomposition of cross-country education gaps into the contributions of product mix and within industry educational variation. The decomposition is based on two counter-factual experiments which transform the educational attainment of a country into that of the U.S. Both experiments replace a country's industry mix and within industry attainment with their U.S. counterparts. The experiments only differ in the sequencing of the changes. Experiment I changes product mix first and within industry education second. Experiment II reverses the order. I measure the contribution of within industry education gaps by the effect of changing within industry education on counter-factual country attainment.

**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 8 countries with the largest education gaps.

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

Section 3.2 presents the time series findings for 16 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 and experiment, the contribution of structural change smaller than in the cross-section.

One way of summarizing quantitative findings is as follows: Upgrading the within industry skills 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 85% 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 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 cross-section.

**Literature.** The paper is related to a small literature studying the sources of cross-country variation in human capital or education (Kaboski 2006; Schoellman 2006; Hendricks 2008). The objective of this paper is to document an observation which may guide the development of theories of country educational attainment. My findings suggest that such theories should focus on skill upgrading within industries rather than on structural change.

Two of my findings resemble the literature on human capital in cities. Hendricks (2007) shows that industry neutral differences in skilled labor productivity account well for cross-city data in

the U.S. Doms & Lewis (2006) and Hendricks (2007) 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 2008). It consists of large individual level data sets for 35 countries and several years. For this paper, I use up to two waves for 26 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 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, CLASS-WKR is used.
3. Industry (IND) 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).

### 2.2 Variable Definitions

For each individual, I construct a measure of labor input, skill level, 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  $i = 1, \dots, I$ . Table 18 in the Appendix lists detailed data for each country. My measurement choices generally follow Caselli & Coleman (2006; hereafter CC).

**Schooling:** EDATTAN assigns each person, indexed by  $h$ , one of seven education levels,  $e_h$ : no school, some primary, primary completed, lower secondary, secondary, some tertiary, and tertiary. These levels are indexed by  $j = 1, \dots, 7$ . Following Caselli & Coleman (2006), I divide persons into "unskilled" and "skilled" groups based on their education level  $j$ . Persons with  $e_h \geq j_0$  are skilled; others are unskilled. I consider the same three values of  $j_0$  as do CC: primary, secondary, and higher education completed. For Austria, no persons report less than the mandatory level of some secondary education. The country is therefore dropped when the primary skill cutoff is used.

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. 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  $h$  supplies  $l_h = \lambda_h \times \ell_h$  efficiency units of labor, where  $\ell_h$  denotes hours worked. Following CC, labor efficiency is measured by a Mincerian equation of the form  $\lambda_h = \exp(\mu(s_h) - \mu(s_0))$ , where

$s_h$  denotes years of schooling and  $\mu(s_h)$  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 & 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 error (Banerjee & 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:  $H_{m,i} = \sum_h \omega_h l_h$ , where the sum is taken over skilled persons in industry  $i$ .  $L_{m,i}$  is defined analogously.  $\omega_h$  is person  $h$ 's sample weight (PERWT).

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

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

Given that all countries share the same Mincer returns,  $\mu(s_h)$ , skill premia vary across countries only because average years of schooling by skill level do. The data force me to abstract from skill premium variation across industries within in a country. This assumption is violated in all of the four countries for which individual earnings are reported. In these countries, skill premia, measured by Mincer returns, are higher in more skill intensive industries. I argue below that this does not affect the paper's main findings.

**Average years of schooling:** For most of the paper's results, educational attainment is measured by the ratio of skilled to unskilled labor inputs

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

An alternative measure is average years of schooling, defined as  $S_{m,i} = \sum_h \omega_h \ell_h s_h / \sum_h \omega_h \ell_h$ , 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  $\eta_{m,i}$  as the measure of educational attainment.

Table 1: Dispersion of educational attainment.

	Std. dev.	U.S./Lowest 5
Primary	1.77	164.0
Secondary	1.29	69.9
College	1.17	31.5
$S_m$	2.60	13.6 / 5.0

Notes: The table shows the standard deviation of  $\eta_m = \ln(H_m/L_m)$  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.

### 3 Results

This section decomposes cross-country and time series variation in educational attainment into the contributions of structural change 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,  $\eta_m = \ln H_m/L_m$ .

The table reveals a large dispersion of  $\eta_m$  across countries. For the primary skill cutoff, the ratio between the U.S. and the least educated five countries is above 160. The dispersion 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  $\eta_{US} - \eta_m$ . The purpose of this paper is to decompose this gap into the contributions of product mix and within industry education gaps.

A natural decomposition is suggested by the identity

$$\eta_{US} - \eta_m = \ln \sum_i n_{US,i} \frac{H_{US,i}}{L_{US,i}} - \ln \sum_i n_{m,i} \frac{H_{m,i}}{L_{m,i}} \quad (3)$$

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

$$n_{m,i} = L_{m,i}/L_m \quad (4)$$

Consider two counter-factual experiments, both of which transform  $\eta_m$  into  $\eta_{US}$ .

*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  $\eta_m^I = \ln \sum_i n_{US,i} \exp(\eta_{m,i})$ .  $\eta_m^I - \eta_m$  measures the contribution of product mix to country education gaps. The experiment then replaces country  $m$ 's within industry education ( $\eta_{m,i}$ ) with their U.S. values. Thus,

$$\eta_{US} - \eta_m^I = \ln \sum_i n_{US,i} \exp(\eta_{US,i}) - \ln \sum_i n_{US,i} \exp(\eta_{m,i}) \quad (5)$$

measures the contribution of within industry attainment 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  $\eta_m$ , it first replaces within industry attainment ( $\eta_{m,i}$ ) with  $\eta_{US,i}$ , resulting in  $\eta_m^{II} = \ln \sum_i n_{m,i} \exp(\eta_{US,i})$ . The contribution of within industry education gaps is then given by

$$\eta_m^{II} - \eta_m = \ln \sum_i n_{m,i} \exp(\eta_{US,i}) - \ln \sum_i n_{m,i} \exp(\eta_{m,i}) \quad (6)$$

Then the experiment replaces the sector weights ( $n_{m,i}$ ) with their U.S. values, so that  $\eta_{US} - \eta_m^{II}$  measures the contribution of product mix.

Both experiments offer sensible decompositions that differ only in the employment weights used to weight within industry education gaps ( $\eta_{US,i}$  vs.  $\eta_{m,i}$ ). 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

$$\hat{\eta}_{US,m} - \hat{\eta}_m = \ln \sum_i \hat{n}_{m,i} \exp(\eta_{US,i}) - \ln \sum_i \hat{n}_{m,i} \exp(\eta_{m,i}). \quad (7)$$

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 (7) 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.

Most of the paper's findings are reported in the form of cross-country averages. I denote average educational attainment by

$$\bar{\eta} = M^{-1} \sum_{m=1}^M \eta_m \quad (8)$$

Average education gaps are defined as simple averages of the country level gaps (7). Similarly, the average fraction of cross-country gaps that is due to within industry gaps is defined as the mean of  $\hat{\eta}_{US,m} - \hat{\eta}_m$  divided by  $\eta_{US} - \bar{\eta}$ . To facilitate the interpretation, all of the paper's tables display average gaps as exponentials. For example, the average cross-country gap is displayed as  $\exp(\eta_{US} - \bar{\eta})$ , so that a gap of 10 means that U.S. attainment ( $H_{US}/L_{US}$ ) is 10 times larger than the average attainment in other countries.

I also report results that use *average years of schooling* to measure a country's educational attainment. Education gaps are then defined as  $S_{US} - S_m$ . Within industry education gaps are given by  $\sum_i (S_{US} - S_{m,i}) f_{m,i}$ , where  $f_{m,i}$  denotes the share of hours worked in industry  $i$ , averaged across country  $m$  and the U.S. Denote by  $\bar{S}$  the cross-country average of  $S_m$ .

### 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 education gap and the average gap due to within industry  $\eta_{m,i}$  gaps. To illustrate: the first row shows results for the primary skill cutoff. The average gap in  $\hat{\eta}_{US,m} - \hat{\eta}_m$  implies a more than 30-fold difference in skilled to unskilled labor inputs. Of this 30-fold gap, a factor of 12 (73%) is due to within industry gaps.

The main message of table 2 is that at least 2/3 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.

Table 2: Direct decomposition. Cross-sectional data.

	Avg.gap	Within gap	Within fraction	$N$
Primary	30.17	12.02	0.73	24
Secondary	15.69	9.49	0.82	25
College	4.77	2.88	0.68	25
$S_m$	5.22	4.01	0.77	25

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.

Figure 1 shows the within industry gaps by country. The horizontal axis displays each country's education gap,  $\hat{\eta}_{US,m} - \hat{\eta}_m$ . Each panel represents a skill definition. The figure also shows lines with slopes 0.5 and 1. These represent points where within industry gaps contribute 50% or 100% to a country's education gap. The majority of countries lie between these lines, indicating that within industry gaps account for more than half of most country's education gaps. For nearly all countries with low education, this is the case. 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  $\eta_m$ . 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 for at least 65% 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>1</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.

**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,  $\eta_{US,i} - \eta_{m,i}$ , against the country gap,  $\eta_{US} - \eta_m$ . 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.

<sup>1</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.

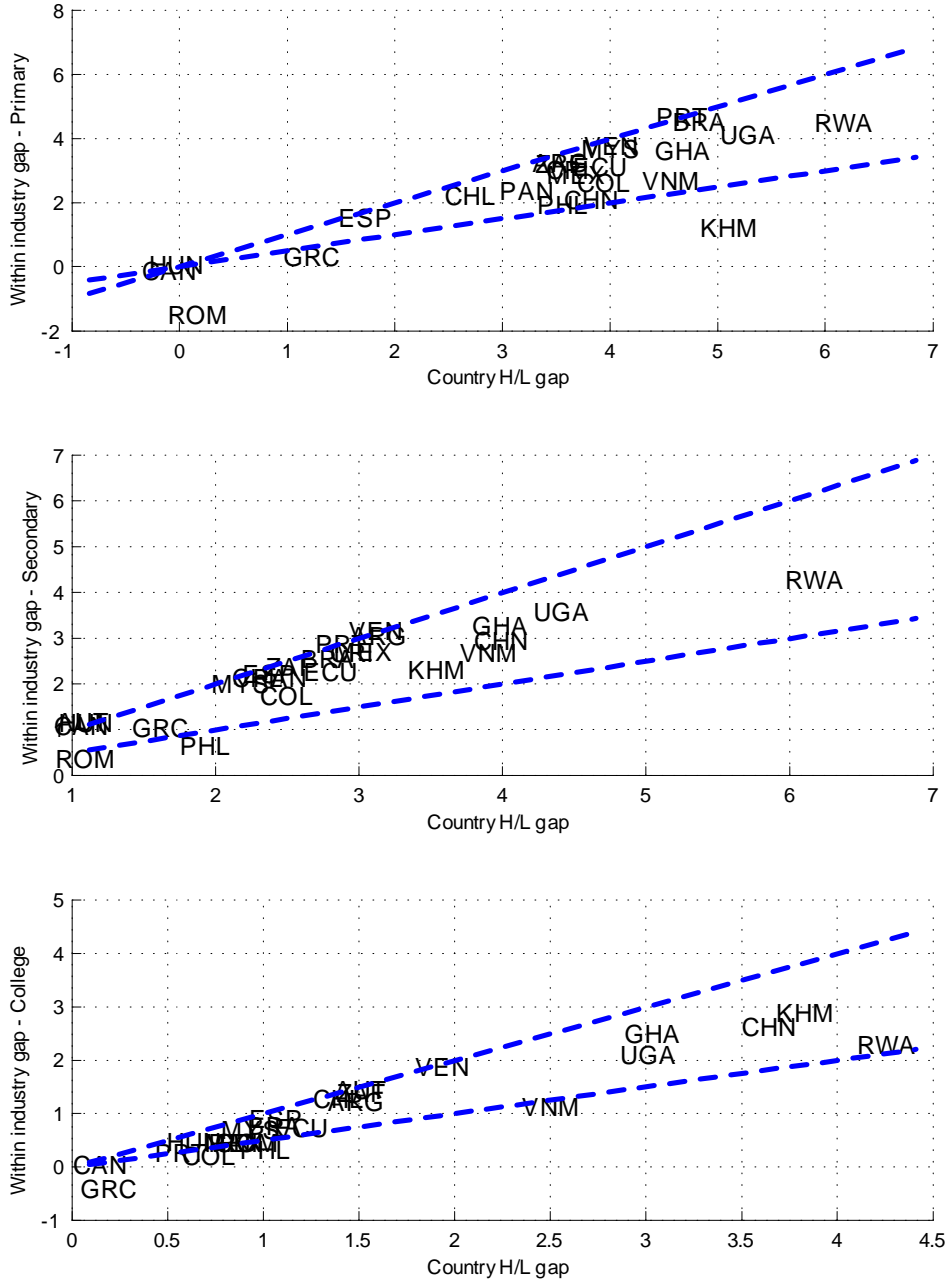


Figure 1: Direct decomposition. Cross-country data.

Table 3: Direct decomposition. Cross-country data.

Primary skill cutoff	Avg.gap	Within industry gap	Fraction
Group 1 RWA, UGA, KHM, BRA, GHA, PRT, VNM, VEN	135.86	37.70	0.74
Group 2 MYS, COL, ECU, CHN, MEX, CRI, PHL, ARG	42.73	16.83	0.75
Group 3 ZAF, PAN, CHL, ESP, GRC, ROM, HUN, CAN	4.73	2.74	0.65
Secondary skill cutoff			
Group 1 RWA, UGA, GHA, CHN, VNM, KHM, ARG, VEN	56.36	23.41	0.78
Group 2 MEX, CRI, PRT, ECU, BRA, ZAF, COL, PAN, ESP	14.79	10.86	0.89
Group 3 CHL, MYS, PHL, GRC, HUN, ROM, AUT, CAN	4.67	3.31	0.78
College skill cutoff			
Group 1 RWA, KHM, CHN, UGA, GHA, VNM, VEN, ZAF	19.39	8.13	0.71
Group 2 AUT, ARG, CHL, ECU, ESP, BRA, PHL, MYS, ROM	3.23	2.37	0.73
Group 3 CRI, PAN, MEX, COL, HUN, PRT, GRC, CAN	1.83	1.27	0.40

Notes: Direct decomposition of cross-country education gaps. Countries are grouped by  $\eta_m$ . 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.

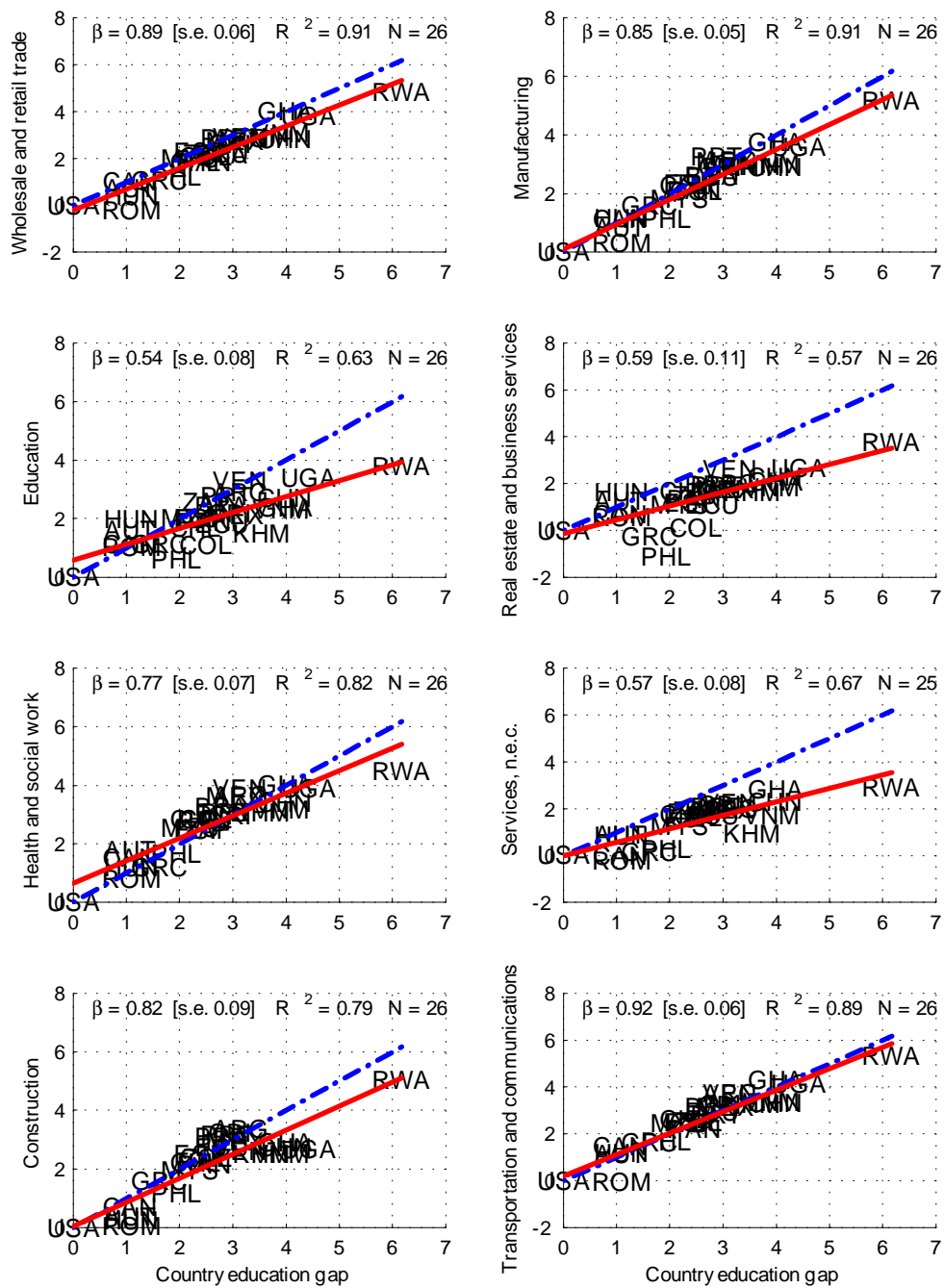


Figure 2: Within industry and aggregate education gaps. Cross-country data.

Table 4: Within industry and aggregate education gaps.

Skill cutoff	Avg. $\beta$	Avg. $R^2$	Avg. N	No.of.ind.
Primary	0.88 (0.11)	0.82 (0.12)	24.6	15
Secondary	0.78 (0.16)	0.76 (0.12)	25.7	15
College	0.71 (0.25)	0.69 (0.18)	25.6	15

Notes: The table shows the results of regressing, by industry,  $\eta_{US,i} - \eta_{m,i}$  on  $\eta_{US} - \eta_m$ . 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.

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, nearly 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  $\eta_{m,i}$  by country. The horizontal axis displays country attainment,  $\eta_m$ . The vertical axis shows  $\eta_{m,i}$  for all industries in country  $m$  together with the population weighted median of  $\eta_{m,i}$ . 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 2/3 of countries ( $\eta_{US,Ag} > \eta_m$ ). The distributions of  $\eta_{m,i}$  overlap more for the college cutoff. Even in that case, in the three least educated countries no industry is more skilled than the U.S. median.

One implication of figure 3 is that large cross-country education gaps would remain, even if all U.S. employment were reallocated to agriculture. Figure 4 displays the results of this thought experiment. If all U.S. labor were moved to agriculture, the remaining education gaps would be  $\eta_{US,Ag} - \eta_m$ . Figure 4 plots these gaps against the observed ones,  $\eta_{US} - \eta_m$ . Also shown are lines representing 50% and 100% contribution of within industry gaps. For all skill cutoffs, more than half of the observed education gaps remain for countries in group 1.

This experiment provides an upper bound for the role of product mix and thus a lower bound for the role of within industry gaps. Even if nothing is known about the industry composition of countries other than the U.S., the role of product mix cannot exceed what is shown in figure 4. This result establishes that the paper’s main finding is robust against incorrectly assigning industries to workers in low income countries.

### 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  $\eta_{t,m,i} = \ln H_{t,m,i}/L_{t,m,i}$  the educational attainment of industry  $i$  in country  $m$  at date  $t$ . The goal is to decompose country level changes in attainment,  $\eta_{2,m} - \eta_{1,m}$ , into the contributions of product mix ( $n_{t,m,i}$ ) changes and within industry ( $\eta_{t,m,i}$ ) changes.

The decomposition proceeds in the same way as for the cross-sectional results. Country  $m$ ’s

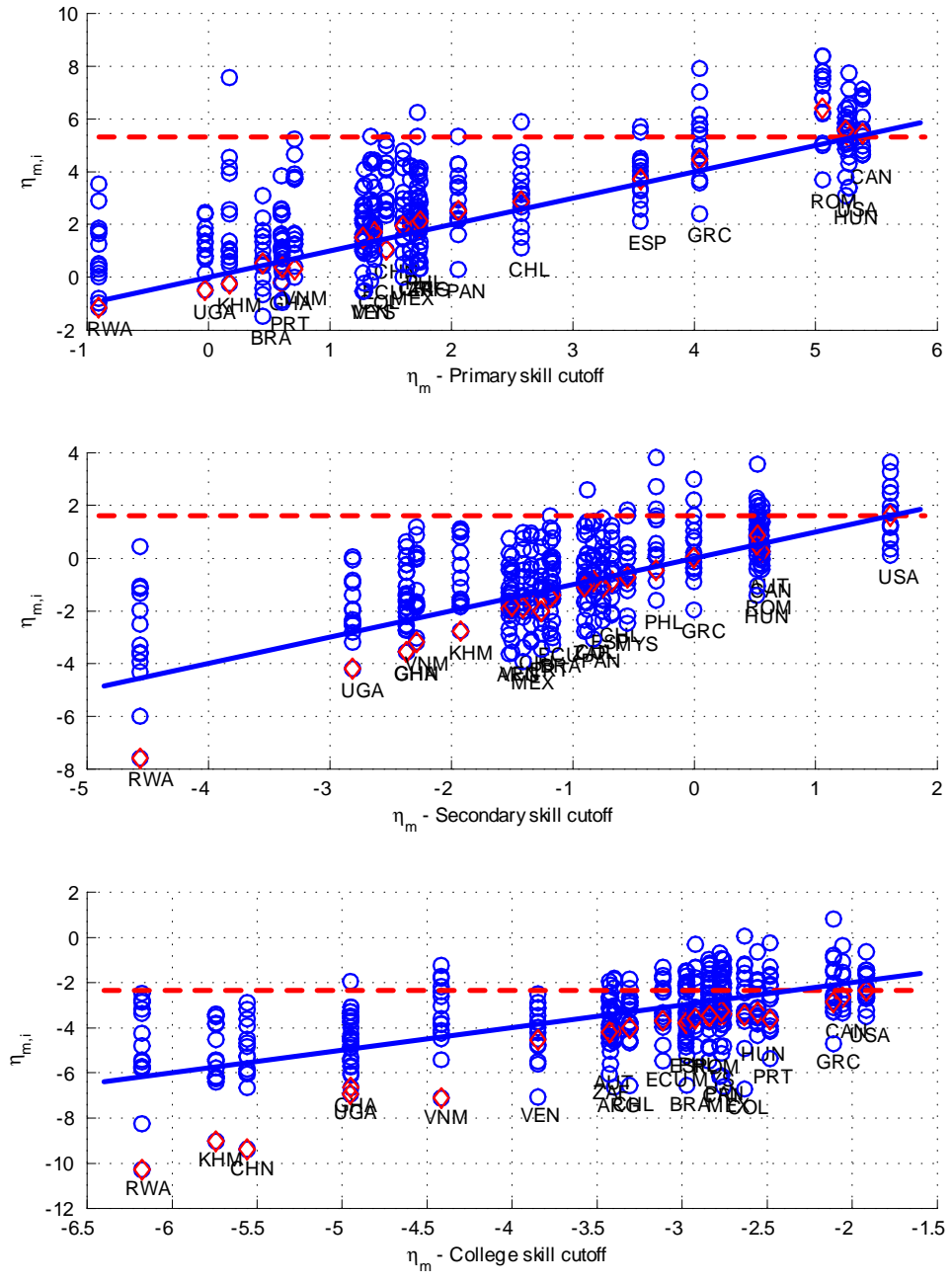


Figure 3: Range of  $\eta_{m,i}$  by country.

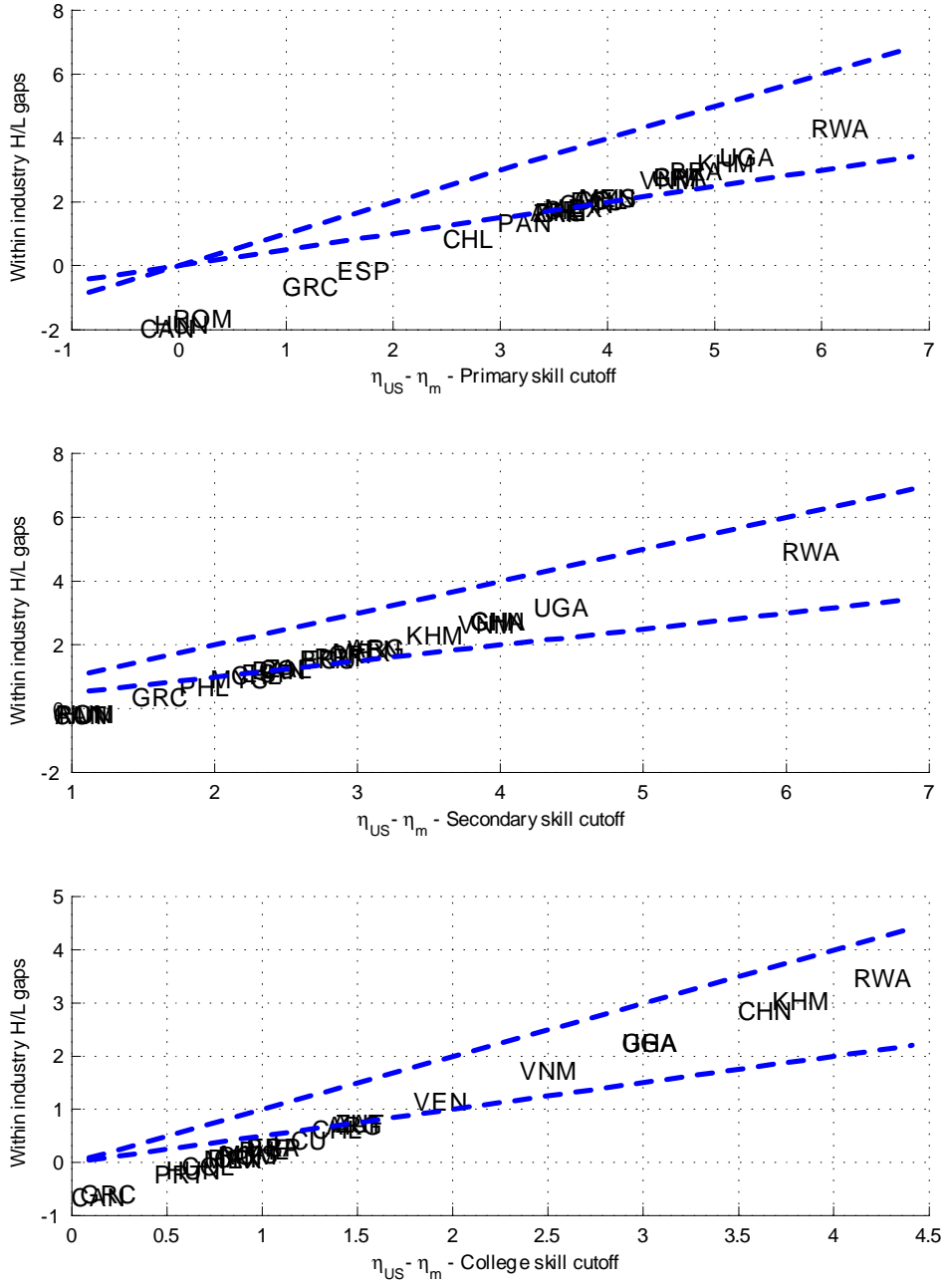


Figure 4: Within industry skill gaps with all U.S. labor in agriculture.

Table 5: Time series decomposition of H/L changes

	Avg.gap	Within gap	Within fraction	$N$
Primary	4.41	3.77	0.90	15
Secondary	3.68	3.39	0.94	16
College	2.61	2.10	0.77	16
$S_m$	2.88	2.19	0.76	16

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

2.

within industry education gap is defined as the weighted average of  $\eta_{2,m,i} - \eta_{1,m,i}$ . The weights are the unskilled labor input shares, averaged across the two dates.

**Results.** Table 5 shows the average contributions of within industry changes in  $\eta_{t,m,i}$ . 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 0.94, depending on the skill definition. Relative to the cross-sectional result, 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 5 shows country level results. It plots counter-factual within industry changes for both experiments against the total changes in attainment. For all but one country, within industry changes account for more than 50%, or at times even for more than 100%, of the country level changes in attainment.

**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  $\eta_{t,m}$  over time. Group 1 countries experience the largest increases in  $\eta_{t,m}$ .

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 3/4 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 85% were achieved through within industry skill upgrading.

**Within industry education changes.** One reason why within industry changes in attainment are important is that they are highly correlated with country level attainment changes. To demonstrate this point, I regress, by industry, within industry changes ( $\eta_{2,m,i} - \eta_{1,m,i}$ ) on country level changes ( $\eta_{2,m} - \eta_{1,m}$ ) 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 high; in fact, they are above unity (albeit with large standard deviations). The  $R^2$  statistics are on the order of 0.7, suggesting that countries with growing education upgrade the skill intensities of most sectors.

**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 85% were achieved through within

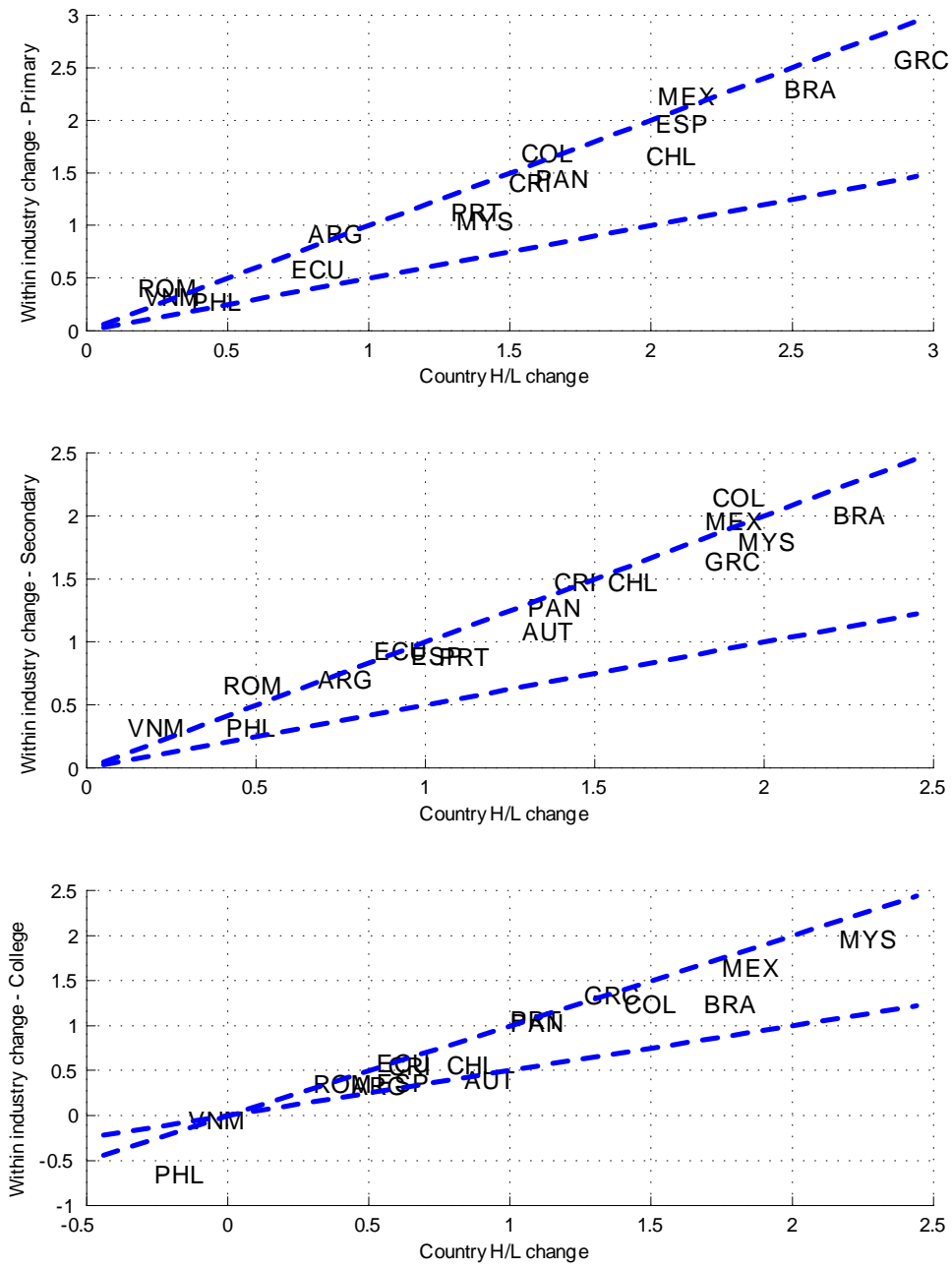


Figure 5: Within industry and country level changes in  $H/L$ .

Table 6: Direct decomposition of time series education changes.

Primary skill cutoff	Avg.change	Within industry gap	Fraction
Group 1 GRC, BRA, MEX, ESP, CHL	10.64	8.53	0.91
Group 2 PAN, COL, CRI, MYS, PRT	4.64	3.81	0.87
Group 3 ARG, ECU, PHL, VNM, ROM	1.73	1.65	0.91
Secondary skill cutoff			
Group 1 BRA, MYS, COL, MEX, GRC	7.43	6.73	0.95
Group 2 CHL, CRI, PAN, AUT, PRT, ESP	3.76	3.25	0.89
Group 3 ECU, ARG, ROM, PHL, VNM	1.77	1.79	1.02
College skill cutoff			
Group 1 MYS, MEX, BRA, COL, GRC	5.76	4.39	0.85
Group 2 PRT, PAN, AUT, CHL, CRI, ECU	2.40	2.00	0.79
Group 3 ESP, ARG, ROM, VNM, PHL	1.31	1.07	0.24

Notes: The table shows the fraction of time series variation in  $\eta_{t,m}$  due to within industry changes in  $\eta_{t,m,i}$ . Countries are divided into groups according to their changes in  $\eta_{t,m}$ . 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	No.of.ind.
Primary	1.04 (0.20)	0.74 (0.19)	15.5	15
Secondary	1.05 (0.14)	0.63 (0.16)	16.5	15
College	1.00 (0.36)	0.56 (0.26)	16.1	15

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

industry upgrading (table 6).

## 4 Detailed Industries

Models of development commonly focus on broad sectoral shifts from unskilled ("agriculture") to skilled sectors ("industry").<sup>2</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. It is based on each country's separate industry classification. 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. These cases are retained. While it is not possible to map all industries into the common classification, the mapped industries typically employ around 90% of the work force.<sup>3</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 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

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

where  $A_{m,i}$  is a productivity term. If labor is paid its marginal product, the labor demand equation is given by

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

<sup>2</sup>Examples include Hansen & Prescott (2002) or Gollin et al. (2007).

<sup>3</sup>A table listing the mapping from country specific to CHL 2002 industries is available at [www.lhendricks.org](http://www.lhendricks.org).

where

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

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  $\eta_{m,i}$  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 Trefler (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*.

Trefler (1993) shows that industry neutral differences in factor augmenting productivities 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  $\phi_{m,i}$ .

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

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

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

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

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 premia,

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

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 product mix.  $\varepsilon_m$  is a residual which captures model mis-specification 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 premia, 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 value of  $j_0$ , 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 premia vary across industries, the estimates of  $\phi_{m,i}$  are biased. Note, however, that country specific variation in skill premia 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.

## 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  $R^2$

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

The numerator is the model's residual sum of squares. The denominator measures the total sum of squares. 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 72% 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,  $\eta_{m,i}$ . The predicted values,  $\eta_{m,i}^{pred}$ , 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

$$R_i^2 = 1 - \frac{\sum_m (\eta_{m,i} - \eta_{m,i}^{pred})^2}{\sum_m (\eta_{m,i} - \frac{1}{n} \sum_m \eta_{m,i})^2} \quad (22)$$

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

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

Skill cutoff	Broad ind.			Detailed ind.		
	$R_i^2$	$R_m^2$	D.f.	$R_i^2$	$R_m^2$	D.f.
Primary	0.90	0.82	328	0.84	0.73	1087
Secondary	0.78	0.78	343	0.73	0.74	1198
College	0.56	0.75	342	0.40	0.72	1147

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.

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

Skill cutoff	Broad ind.		Detailed ind.	
	$R_i^2$	$R_m^2$	$R_i^2$	$R_m^2$
Primary	0.91	0.82	0.85	0.73
Secondary	0.78	0.78	0.74	0.74
College	0.58	0.75	0.43	0.72

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

The industry neutral model generally accounts for at least 74% of the variation in  $\eta_{m,i}$  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 implies 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 premia ( $\Delta W_m$ ), 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  $\sum \Delta\alpha_m / \sum (\eta_{US} - \eta_m)$ .

Within industry gaps account for 79% to 84% of cross-country gaps. The bulk of these are due to skill bias differences, not to variation in skill premia. This result does not change when country specific Mincer returns are used to estimate skill premia and labor inputs. For comparison, the direct decomposition attributes between 68% and 82% of cross-country education gaps to within industry gaps (see table 2).<sup>4</sup>

Figure 6 compares the direct and the industry neutral decompositions by country. 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.

<sup>4</sup>The industry neutral model's decomposition is close to that of Experiment II developed in section 3.1.1.

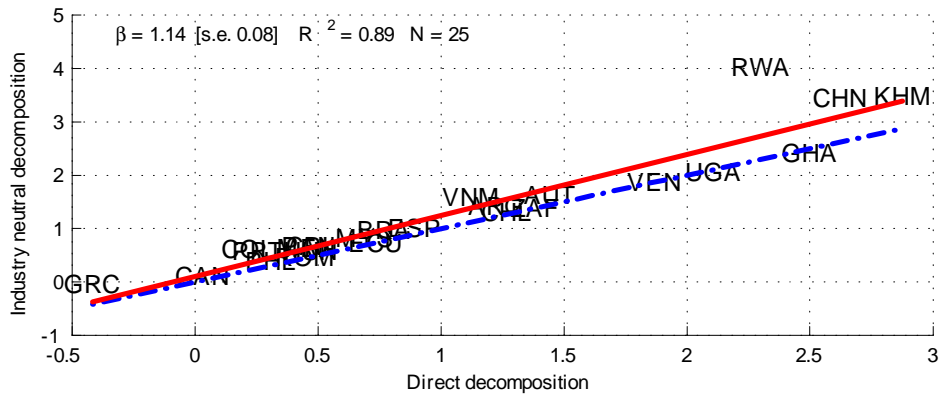
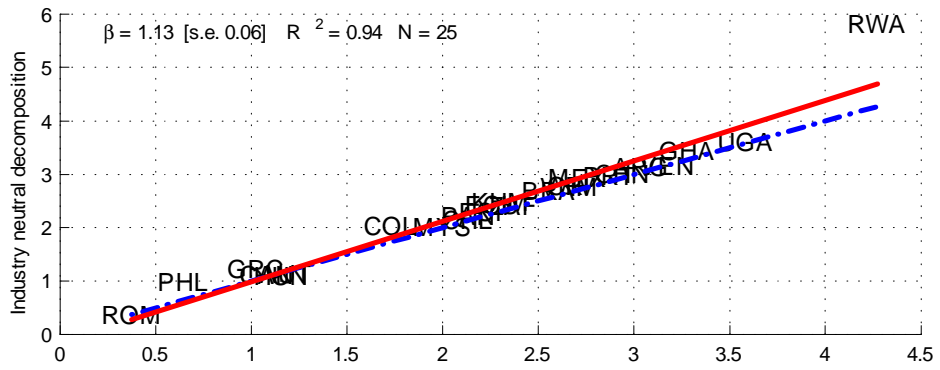
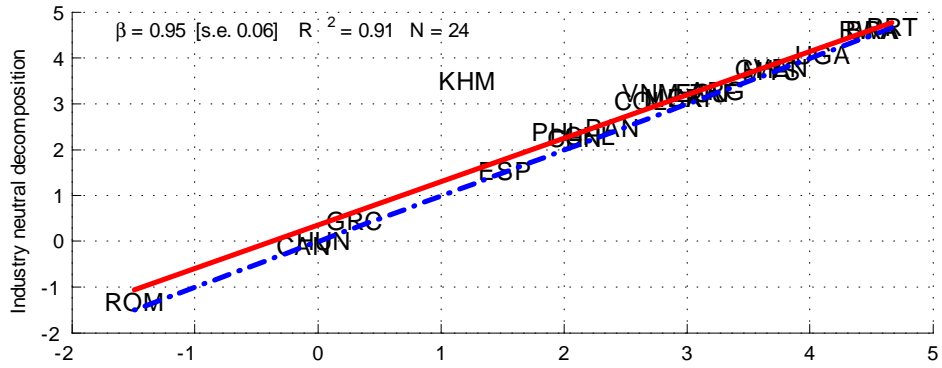


Figure 6: Direct and industry neutral decomposition.

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	30.29	14.44 ( 0.78)	1.04 ( 0.01)	2.32 ( 0.25)	15.02 ( 0.79)
Secondary	15.69	12.20 ( 0.91)	0.91 (-0.03)	1.71 ( 0.20)	11.12 ( 0.87)
College	4.78	3.72 ( 0.84)	1.00 (-0.00)	1.68 ( 0.33)	3.72 ( 0.84)

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	32.63	14.87 ( 0.77)	1.03 ( 0.01)	1.00 ( 0.00)	15.39 ( 0.78)
Secondary	15.70	12.15 ( 0.91)	0.90 (-0.04)	1.00 ( 0.00)	10.93 ( 0.87)
College	4.77	3.48 ( 0.80)	1.00 (-0.00)	1.00 ( 0.00)	3.48 ( 0.80)

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

### 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.

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 premia 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

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	26.4	13.5	79.4	28.0	13.7	78.4
Secondary	14.1	10.1	87.5	14.0	9.9	86.9
College	4.5	3.5	84.0	4.5	3.3	79.9

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  $\eta_{US} - \eta_m$ . "Within gap" denotes the average within industry gap. "Fraction" is the average fraction due to within industry gaps.

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

Skill cutoff	Broad ind.		Detailed ind.	
	$R_i^2$	$R_m^2$	$R_i^2$	$R_m^2$
Primary	0.90	0.82	0.91	0.76
Secondary	0.74	0.81	0.75	0.75
College	0.44	0.77	0.30	0.67

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

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  $\eta_{1,m,i}$  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  $\eta_{t,m,i}$  over time. For each industry, I regress predicted against observed changes in  $\eta_{t,m,i}$ . 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 0.7, compared with a model prediction of 1. Attenuation bias may reduce the slope coefficients.

#### 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 premia.

As in the cross-sectional case, the industry neutral model generally overstates the fraction of

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

Skill cutoff	Pseudo $R^2$	Avg.Slope (s.e.)
Primary	0.65 (0.29)	0.77 (0.19)
Secondary	0.50 (0.32)	0.69 (0.19)
College	0.40 (0.24)	0.59 (0.26)

Notes: The table compares predicted and actual changes in  $\eta_{t,m,i}$ . 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.

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

Skill cutoff	Ind.neutral	Direct	$N$
Primary	0.99	0.90	15
Secondary	1.11	0.94	16
College	1.07	0.77	16

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

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.

#### 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  $M^{-1} \sum (\eta_{2,m} - \eta_{1,m})$ .

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.

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	4.6	97.8	5.1	5.3	102.2
Secondary	3.9	4.5	109.7	4.2	5.2	115.1
College	2.6	2.8	106.7	2.8	3.1	111.3

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

## 5 Conclusion

Educational attainment rises dramatically during the process of development. This paper investigates how countries absorb large or rising endowments of skilled labor. The cross-country and time series variation in educational attainment is decomposed into the contributions of structural change and within industry skill upgrading.

The main finding is that at least  $2/3$  of the variation in educational attainment across countries and over time occurs within industries. These findings suggest that theories of country educational attainment should focus on skill bias or factor quality differences. Structural change from unskilled labor intensive products to skill intensive products is less important, though by no means negligible.

Future research should investigate the sources of within industry differences in educational attainment.

## **6 Appendix**

Table 17 shows summary statistics by country. Table 18 shows summary statistics for skilled and unskilled labor.

Table 17: Summary statistics

Country	Year	N	School	No.ind.	Matched ind.	Empl.frac.
USA	2000	5458.3	13.6	265	54	87.1
	1960	562.8	10.5	151	50	92.5
ARG	1991	637.6	8.5	66	59	98.3
	1970	139.2	7.0	223	53	80.8
AUT	2001	352.4	11.7	79	42	93.1
	1971	307.4	10.5	79	42	93.5
BRA	2000	3092.9	6.8	221	58	92.4
	1960	872.2	2.1	137	43	84.2
CAN	2001	322.6	13.4	n/a	n/a	n/a
CHL	2002	485.0	10.8	62	62	100.0
	1970	220.3	6.2	165	54	93.0
CHN	1990	1329.3	6.5	328	48	94.3
COL	2005	499.8	7.5	62	61	100.0
	1973	397.6	3.8	98	50	81.1
CRI	2000	109.8	7.8	173	53	93.7
	1973	48.0	5.2	161	54	95.4
ECU	2001	329.0	7.9	159	58	92.9
	1982	151.7	5.7	99	50	94.4
ESP	2001	746.3	9.6	60	60	100.0
	1981	581.0	5.6	65	50	90.0
GHA	2000	651.9	5.2	64	56	99.7
GRC	2001	366.2	10.8	224	58	92.8
	1971	270.0	6.6	158	44	98.0
HUN	2001	184.5	12.1	60	60	100.0
KHM	1998	308.9	5.4	159	58	98.1
MEX	2000	2881.9	8.2	117	48	85.9
	1960	79.2	3.7	132	47	91.5
MYS	2000	143.6	8.1	156	61	99.6
	1970	49.2	4.1	223	53	96.9
PAN	2000	86.6	9.2	162	60	99.9
	1970	44.4	5.6	71	47	93.3
PHL	2000	1292.7	9.7	63	57	68.2
	1990	1464.1	8.7	140	51	96.6
PRT	2001	220.0	7.5	228	57	93.9
	1981	181.2	4.9	81	48	90.7
ROM	2002	755.0	10.8	61	61	100.0
	1992	926.9	9.8	100	56	97.3
RWA	2002	300.5	3.1	84	43	99.6
UGA	2002	672.0	5.0	n/a	n/a	n/a
VEN	1990	399.9	7.1	79	51	94.8
VNM	1999	1068.5	6.9	159	58	97.4
	1989	1066.8	6.4	n/a	n/a	n/a
ZAF	2001	698.7	9.0	197	57	91.6

Notes:  $N$  is the number of individuals in the sample (in thousands). School denotes average years of schooling. "No. ind" denotes the number of detailed industries and "Matched ind" is the number of these industries that are mapped into the common classification. "Empl.frac." denotes the fraction of persons employed in mapped industries.

Table 18: Summary statistics by skill cutoff

Country	Year	$H_m/L_m$	Primary $w_{H,m}/w_{L,m}$	$\alpha_m$	$H_m/L_m$	Secondary $w_{H,m}/w_{L,m}$	$\alpha_m$	$H_m/L_m$	College $w_{H,m}/w_{L,m}$	$\alpha_m$
USA	2000	100.0	100.0	0.000	100.0	100.0	0.000	100.0	100.0	0.000
ARG	1991	2.9	100.0	-2.340	4.3	100.0	-2.238	22.7	100.0	-1.010
AUT	2001	n/a	n/a	n/a	34.5	93.4	-0.873	22.2	100.0	-1.159
BRA	2000	0.8	100.0	-3.293	6.2	93.4	-1.993	34.8	93.4	-0.758
CAN	2001	111.7	100.0	0.076	34.7	107.0	-0.726	87.1	100.0	-0.081
CHL	2002	6.7	100.0	-1.642	10.1	107.0	-1.468	24.8	100.0	-0.925
CHN	1990	2.2	100.0	-1.618	1.9	93.4	-2.208	2.6	100.0	-2.469
COL	2005	2.0	100.0	-2.184	8.3	93.4	-1.523	48.9	93.4	-0.509
CRI	2000	2.7	100.0	-2.241	5.2	93.4	-2.063	42.2	93.4	-0.540
ECU	2001	2.0	100.0	-2.318	6.1	93.4	-1.796	30.2	107.0	-0.449
ESP	2001	18.0	100.0	-1.103	9.4	87.3	-1.778	34.7	100.0	-0.721
GHA	2000	0.9	100.0	-2.682	1.9	100.0	-2.456	4.8	122.6	-1.531
GRC	2001	29.2	100.0	-0.311	19.9	93.4	-0.940	82.4	100.0	0.023
HUN	2001	97.5	100.0	-0.003	33.3	100.0	-0.789	52.8	100.0	-0.412
KHM	1998	0.6	100.0	-2.505	2.9	71.5	-2.101	2.2	107.0	-2.423
MEX	2000	2.5	100.0	-2.238	4.9	93.4	-2.160	43.2	100.0	-0.456
MYS	2000	1.9	100.0	-2.659	11.5	93.4	-1.501	39.6	100.0	-0.583
PAN	2000	4.0	100.0	-1.771	8.7	100.0	-1.587	42.7	100.0	-0.474
PHL	2000	2.9	114.3	-1.579	14.6	93.4	-0.774	36.6	100.0	-0.277
PRT	2001	0.9	114.3	-3.208	5.7	87.3	-2.255	56.8	93.4	-0.487
ROM	2002	80.4	100.0	0.956	33.5	79.0	-0.495	39.8	100.0	-0.330
RWA	2002	0.2	100.0	-3.294	0.2	93.4	-4.213	1.4	100.0	-2.879
UGA	2002	0.5	114.3	-2.762	1.2	100.0	-2.577	4.8	100.0	-1.481
VEN	1990	1.8	100.0	-2.698	4.4	87.3	-2.391	14.5	100.0	-1.333
VNM	1999	1.0	114.3	-2.177	2.0	93.4	-2.044	8.2	100.0	-1.142
ZAF	2001	2.9	114.3	-2.179	8.1	100.0	-1.720	22.1	93.4	-1.057

Notes: The table shows the estimates of the model variables and parameters for all countries. Each panel represents one skill cutoff,  $j_0$ . For variable definitions, see the text of the paper.

## References

- [1] Acemoglu, Daron (2002). "Directed technical change." *Review of Economic Studies* 69: 781-810.
- [2] Acemoglu, Daron (2003). "Patterns of skill premia." *Review of Economic Studies* 70: 199-230.
- [3] Banerjee, Abhijit; Esther 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 Science, North-Holland Publishers.
- [4] Barro, Robert; Jong-Wha Lee (2001). "International Data on Educational Attainment. Updates and Implications." *Oxford Economic Papers* 53(3): 541-63.
- [5] Caselli, Francesco; Wilbur John Coleman II (2006). "The World Technology Frontier." *American Economic Review* 96(3): 499-522.
- [6] Doms, Mark; Ethan Lewis (2006). "Labor supply and computer adoption." Mimeo. Federal Reserve Bank of San Francisco.
- [7] Galor, Oded (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 Science, North-Holland Publishers.
- [8] Galor Oded; Omer Moav (2000). "Ability Biased Technological Transition, Wage Inequality, and Economic Growth." *Quarterly Journal of Economics* 115: 469-498.
- [9] Galor, Oded; David N. Weil (2000). "Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond." *American Economic Review* 90(4): 806-28.
- [10] Gollin, Douglas; Stephen L. Parente; Richard Rogerson (2007). "The food problem and the evolution of international income levels." *Journal of Monetary Economics* 54: 1230-55.
- [11] Hall, Robert E.; Charles I. Jones (1999). "Why do some countries produce so much more output per worker than others?" *Quarterly Journal of Economics*, 114: 83-116.
- [12] Hansen, Gary D.; Edward C. Prescott (2002). "Malthus to Solow." *American Economic Review* 92(4): 1205-1217.
- [13] Hendricks, Lutz (2008). "Why Does Education Differ Across Countries?" Mimeo. Iowa State University.
- [14] Hendricks, Lutz (2007). "Educational Attainment in U.S. Cities." Mimeo. Iowa State University.
- [15] Minnesota Population Center (2008). *Integrated Public Use Microdata Series — International: Version 4.0*. Minneapolis: University of Minnesota. [www.ipums.org](http://www.ipums.org).
- [16] Kaboski, Joseph P. (2006). "Endogenous Mincerian Returns: Explaining Cross-Country Variation in the Returns to Schooling." Mimeo. Ohio State University.
- [17] Lucas, Robert E. (2004). "Life earnings and rural-urban migration." *Journal of Political Economy* 112 (1, part 2): S29-59.
- [18] Parente, Stephen L.; Edward C. Prescott (1994). "Barriers to Technology Adoption and Development." *Journal of Political Economy* 102(2): 298-321.
- [19] Psacharopoulos, George (1994). "Returns to Investment in Education: A Global Update." *World Development* 22(9): 1325-43.
- [20] Schoellman, Todd (2006). "The causes and consequences of cross-country differences in school attainment." Mimeo. Clemson University.
- [21] Treffer, Daniel (1993). "International factor price differences: Leontief was right!" *Journal of Political Economy* 101(6): 961-87.