

# The Latin American Efficiency Gap

Francesco Caselli\*

July 2014

## 1 Introduction

The average Latin American country produces about 1 fifth of the output per worker of the US. What are the sources of these enormous *income gaps*? This paper reports development-accounting results for Latin America. Development accounting compares differences in income per worker between developing and developed countries to counterfactual differences attributable to observable components of physical and human capital. Such calculations can serve a useful preliminary diagnostic role before engaging in deeper and more detailed explorations of the fundamental determinants of differences in income per worker. If differences in physical and human capital – or *capital gaps* – are sufficient to explain most of the difference in incomes, then researchers and policy makers need to focus on factors holding back investment (in machines and in humans). Instead, if differences in capital are insufficient to account for most of the variation in income, one must conclude that developing countries are also hampered by relatively low efficiency at using their inputs - *efficiency gaps*. The research and policy agenda would then have to focus on

---

\*London School of Economics, Centre for Macroeconomics, BREAD, CEP, CEPR, and NBER. Email: f.Caselli@lse.ac.uk. This paper is part of a research project on Latin American and Caribbean convergence financed by the Latin American and Caribbean Region of the World Bank. I am very grateful to Federico Rossi for excellent research assistance, to Ludger Woessman for patient and constructive advice, and to Jorge Araujo and the other participants in the project for helpful comments.

technology, allocative efficiency, competition, and other determinants of the efficient use of capital.<sup>1</sup>

I present development-accounting results for 2005 for three samples of Latin American countries: a “broad” sample of 22 countries, a “narrow” sample of 9, and an “intermediate” sample of 15.

The three samples differ in the data available to measure human capital. In the broad sample human capital is measured in the context of a “Mincerian” framework, where the key inputs are schooling (years of education) and health (as proxied by the adult survival rate). In the narrow and intermediate samples I augment the Mincerian framework with measures of cognitive skills, to account for additional factors such as schooling quality, parental inputs, and other influences on human capital not captured by years of schooling and health. The measures of cognitive skills are based on tests administered to school-age children. In the narrow sample, the test is a science test whose results are directly comparable between Latin America and the benchmark developed country. In the intermediate sample the tests were only administered in Latin America and can be compared to the benchmark country only on the basis of a number of ad hoc assumptions.

In all three samples I measure physical capital as an aggregate of reproducible and “natural” capital. Reproducible capital includes equipment and structures, while natural capital primarily includes subsoil resources, arable land, and timber.

Given measures of physical capital gaps, as well gaps in the components of human capital, development-accounting uses a calibration to map these gaps into counter-factual income gaps, or the income gaps that would be observed based on differences in human and capital endowments only. Because these counterfactual incomes are bundles of physical and human capital, I refer to the ratio of Latin American counterfactual incomes to the US counterfactual income as *relative capital*.

For each of the three samples I present results from two alternative calibrations, a “baseline” calibration and an “aggressive” calibration. The baseline calibration makes use of the existing body of microeconomic estimates of the Mincerian framework in the way that

---

<sup>1</sup>For a detailed exposition of development accounting see, among others, Caselli (2005).

most closely fits the theoretical framework of development accounting. As it turns out, this leads to coefficients for the components of human capital that are substantially lower than in much existing work in development accounting - leading to relatively smaller estimated capital gaps and, correspondingly, larger efficiency gaps. The aggressive calibration thus uses more conventional figures as a robustness check.

When I use my benchmark calibration, irrespective of sample/cognitive skill correction, I find that relative capital and relative efficiencies are almost identical. For example in the broad sample average relative capital and average relative efficiency are both 44% - or roughly double actual average relative incomes. Hence, both capital gaps and efficiency gaps are very large: the average Latin American country has less than half the capital (human and physical) per worker of the US, and uses it less than half as efficiently.

Using the aggressive calibration, capital gaps are naturally larger, and efficiency gaps correspondingly smaller. Nevertheless, even under this “best-case scenario” for the view that capital gaps are the key source of income gaps, average Latin American efficiency is at most 60% of the US level, still implying a vast efficiency gap.

In assessing this evidence, it is essential to bear in mind that efficiency gaps contribute to income disparity both directly – as they mean that Latin America gets less out of its capital – and indirectly – since much of the capital gap itself is likely due to diminished incentives to invest in equipment, structure, schooling, and health caused by low efficiency. The consequences of closing the efficiency gap would correspondingly be far reaching.

Explaining the Latin American efficiency gap is therefore a high priority both for scholars and for policy makers. It is likely that this task will require firm-level evidence. Firm level evidence would also be invaluable in checking the robustness of the development-accounting results, which are subject to severe data-quality limitations.

## 2 Conceptual Framework

The analytical tool at the core of development accounting is the aggregate production function. The aggregate production function maps aggregate input quantities into output. The main inputs considered are physical capital and human capital. The empirical literature so far has failed to uncover compelling evidence that aggregate input quantities deliver

large external economies, so it is usually deemed safe to assume constant returns to scale.<sup>2</sup> Given this assumption, one can express the production function in intensive form, i.e. by specifying all input and output quantities in per worker terms. In order to construct counterfactual incomes a functional form is needed. Existing evidence suggests that the share of capital in income does not vary systematically with the level of development, or with factor endowments [Gollin (2002)]. Hence, most practitioners of development accounting opt for a Cobb-Douglas specification. In sum, the production function for country  $i$  is

$$y_i = A_i k_i^\alpha h_i^{1-\alpha}, \quad (1)$$

where  $y$  is output per worker,  $k$  is physical capital per worker,  $h$  is human capital per worker (quality-adjusted labor), and  $A$  captures unmeasured/unobservable factors that contribute to differences in output per worker.

The term  $A$  is subject to much speculation and controversy. Practitioners refer to it as total factor productivity, technology, a measure of our ignorance, etc. Here I will refer to it as “efficiency”. Countries with a larger  $A$  are countries that, for whatever reasons, are more efficient users of their physical and human capital.

The goal of development accounting is to assess the relative importance of efficiency differences and physical and human capital differences in producing the differences in income per worker we observe in the data. To this end, one constructs counterfactual incomes, or capital bundles,

$$\tilde{y}_i = k_i^\alpha h_i^{1-\alpha}, \quad (2)$$

which are based exclusively on the observable inputs. Differences in these capital bundles are then compared to income differences. If counter-factual and actual income differences are similar, then observable factors are able to account for the bulk of the variation in income. If they are quite different, then differences in efficiency are important. Establishing how significant efficiency differences are has important repercussions both for research and for policy.

---

<sup>2</sup>See, e.g. Iranzo and Peri (2009) for a recent review and some new evidence on the quantitative significance of schooling externalities.

In order to construct the counterfactual  $\tilde{y}$ s we need to construct measures of  $k_i$  and  $h_i$ , as well as to calibrate the capital-share parameter  $\alpha$ . Standard practice sets the latter to 0.33, and we stick to this practice throughout. Caselli (2005) shows that development-accounting calculations are not overly sensitive to alternative values in a reasonable range. The rest of this section focuses on the measurement of physical and human capital.<sup>3</sup>

Existing development-accounting calculations measure  $k$  exclusively on the basis of *reproducible* capital (equipment and structures). But in most developing countries, where agricultural and mining activities still represent large shares of GDP, natural capital (land, timber, ores, etc.) is also very important. Caselli and Feyrer (2007) show that omitting natural capital can lead to very significant understatements of total capital in developing countries relative to developed ones. Hence, this study will measure  $k$  as the sum of the value of all reproducible and natural capital.

Human capital per worker can vary across countries as a result of differences in knowledge, skills, health, etc. The literature has identified three variables that vary across countries which may capture significant differences in these dimensions: years of schooling [e.g., Klenow and Rodriguez-Clare (1997), Hall and Jones (1999)], health [Weil (2007)], and cognitive skills [e.g. Hanushek and Woessmann (2012a)]. In order to bring these together, we postulate the following model for human capital:

$$h_i = \exp(\beta_s s_i + \beta_r r_i + \beta_t t_i). \quad (3)$$

In this equation,  $s_i$  measures average years of schooling in the working-age population,  $r_i$  is a measure of health in the population, and  $t_i$  is a measure of cognitive skills. The coefficients  $\beta_s$ ,  $\beta_r$ , and  $\beta_t$  map differences in the corresponding variables into differences in human capital.<sup>4</sup>

The model in (3) is attractive because it offers a strategy for calibration of the para-

---

<sup>3</sup>There may well be significant heterogeneity among Latin American countries, and, more importantly, between Latin America and the benchmark rich country, in the value of  $\alpha$ . However, it is not known how to perform development-accounting with country-specific capital shares. This is because measures of the capital stock are indices, so that a requirement for the exercise to make sense is that the results should be invariants to the units in which  $k$  is measured. Now  $(k_i/k_j)^\alpha$  is unit-invariant, but  $(k_i^{\alpha_i}/k_j^{\alpha_j})$  is not.

<sup>4</sup>Some caveats as to the validity of of the functional form assumption in (3) are in order. There

eters  $\beta_s$ ,  $\beta_r$ , and  $\beta_t$ . In particular, combining (1), (3), and an assumption of perfect competition in labor markets, we obtain the “Mincerian” formulation

$$\log(w_{ij}) = \alpha_i + \beta_s s_{ij} + \beta_r r_{ij} + \beta_t t_{ij}, \quad (4)$$

where  $w_{ij}$  ( $s_{ij}$ , etc.) is the wage (years of schooling, etc.) of worker  $j$  in country  $i$ , and  $\alpha_i$  is a country-specific term. This suggests that using within-country variation in wages, schooling, health, and cognitive skills one might in principle identify the coefficients  $\beta$ . In practice, there are severe limitations in following this strategy, that we discuss after introducing the data.

### 3 Data

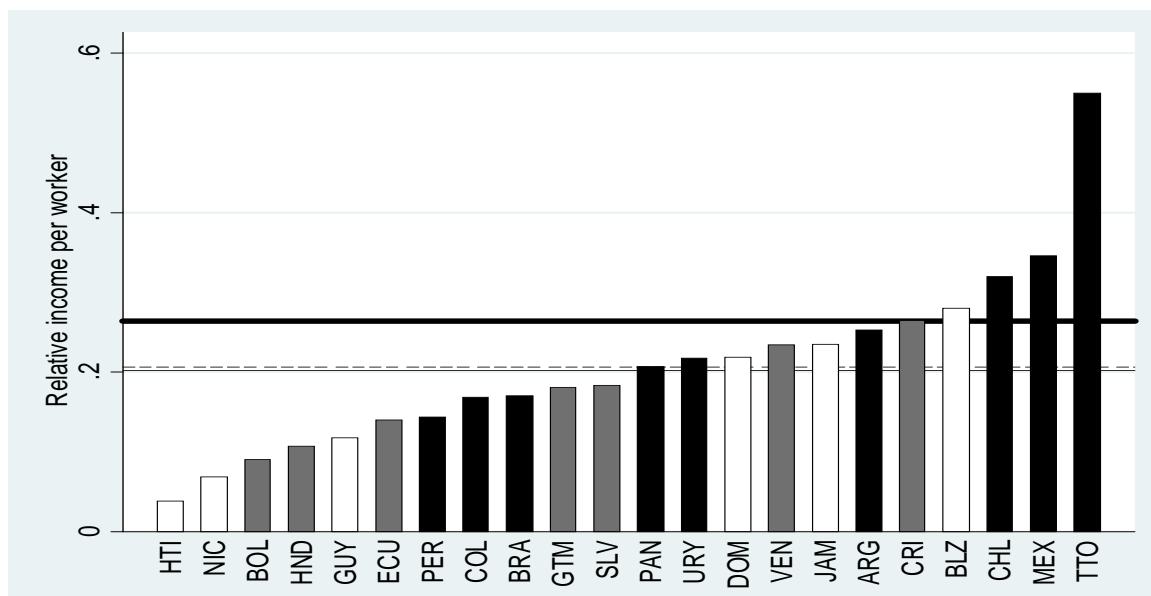
We work with three samples, broad, narrow, and intermediate. The broad data set contains all Latin American countries for which we have data for  $y$ ,  $k$ ,  $s$ , and  $r$ , all observed in 2005. There are 22 such countries (excluded are Barbados, Cuba, and Paraguay, for which we have no capital data). The other two samples add alternative measures of  $t$ . The trade-off is that one measure offers a more credible comparison with the benchmark high-income country, but is only available for 9 Latin-American economies. The more dubious but more plentiful measure is available for 15 countries. All but one of the countries in the narrow sample are also in the intermediate sample (Trinidad and Tobago is the exception). The dataset also includes data from the USA, which we use as the benchmark rich country.

Per-worker income  $y_i$  is variable *rgdpwok* from version 7.1 of the Penn World Tables (PWT71). Figure 1 shows per-worker income in each country in the broad sample relative to the USA, or  $y_i/y_{US}$ . Countries that are also included in the narrow sample are in black, and countries that are in the intermediate but not the narrow sample are in grey. With the exception of Trinidad and Tobago, all Latin America countries have per-worker incomes

---

is considerable micro and macro evidence against the assumption that workers with different years of schooling are perfect substitutes [e.g. Caselli and Coleman (2006)]. In this paper I abstract from the issue of imperfect substitutability. Caselli and Ciccone (2013) argue that consideration of imperfect substitution is unlikely to reduce the estimated importance of efficiency gaps.

Figure 1: Income per worker relative to the US



White bars; only broad sample. Grey bars: only broad and intermediate samples. Black bars: all samples (except TTO not in intermediate). Dashed line: broad sample mean. Light solid line: intermediate sample mean. Heavy solid line: narrow sample mean. Source: PWT71.

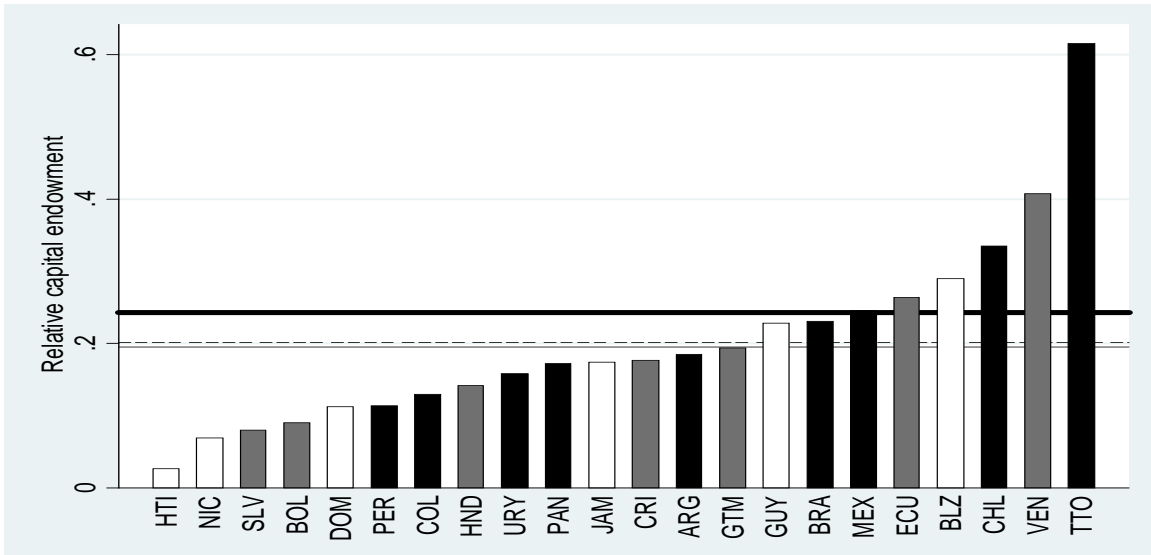
well below 40% of the US level, sometimes much below. The horizontal lines show the three (unweighted) sample averages, indicating that the average country is only one fifth as productive as the USA.<sup>5</sup>

World Bank (2012) presents cross-sectional estimates of the *total* capital stock,  $k$ , as well as its components, for various years. The total capital stock includes reproducible capital, but also land, timber, mineral deposits, and other items that are not included in standard national-account-based data sets. The basic strategy of the World Bank team that constructed these data begins with estimates of the rental flows accruing from different types of natural capital, which are then capitalized using fixed discount rates. I construct the total capital measure by adding the variables *producedplusurban* and *natcap*.

Figure 2 shows total (reproducible plus natural) capital per worker estimates for Latin American countries relative to the US,  $k_i/k_{US}$ . The average Latin American worker is endowed with approximately one fifth of the physical capital of the average US worker.

<sup>5</sup>In the narrow sample the average is higher due to the disproportionate weight of Trinidad and Tobago. Labor-force weighted averages are reported in Table 1 below.

Figure 2: Physical capital per worker relative to the US



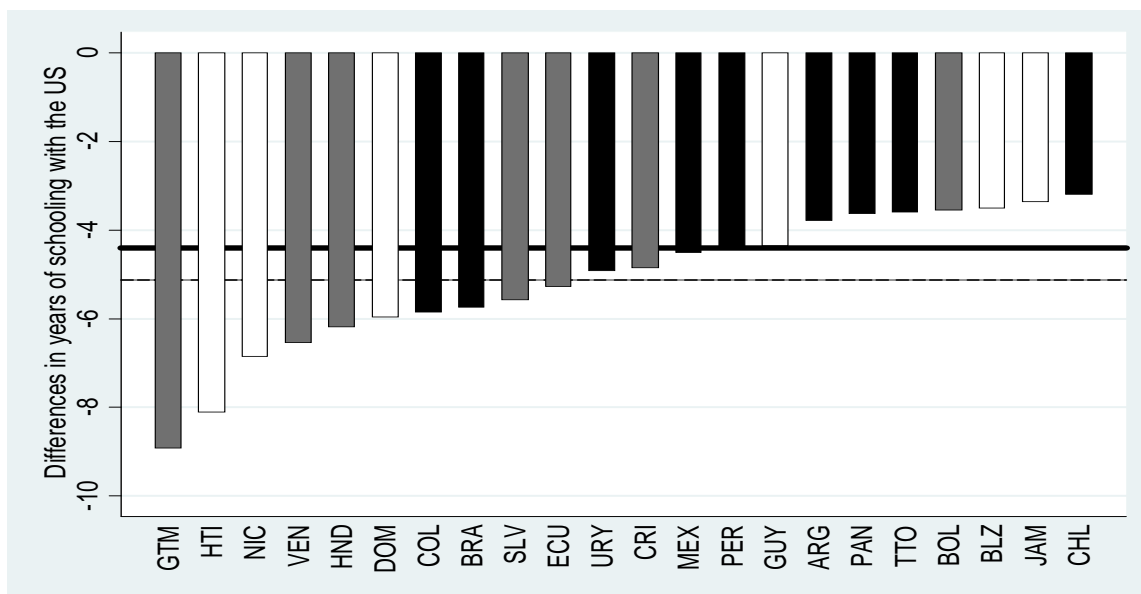
White bars; only broad sample. Grey bars: only broad and intermediate samples. Black bars: all samples (except TTO not in intermediate). Dashed line: broad sample mean. Light solid line: intermediate sample mean. Heavy solid line: narrow sample mean. Source: World Bank (2012).

For average years of schooling in the working-age population (which is defined as between 15 and 99 years of age) I rely on Barro and Lee (2013). Note from equation (3) that for the purposes of constructing *relative* human capital  $h_i/h_{US}$  what is relevant is the *difference* in years of schooling  $s_i - s_{US}$ . The same will be true for  $r$  and  $t$ . Accordingly, in Figure 3 I plot schooling-year differences with the USA in 2005. Latin American workers have always at least three year less schooling than American ones, and five on average.

As a proxy for the health status of the population,  $r$ , Weil (2007) proposes using the *adult survival rate*. The adult survival rate is a statistic computed from age-specific mortality rates at a point in time. It can be interpreted as the probability of reaching the age of 60, conditional on having reached the age of 15, *at current rates of age-specific mortality*. Since most mortality before age 60 is due to illness, the adult survival rate is a reasonably good proxy for the overall health status of the population at a given point in time. Relative to more direct measures of health, the advantage of the adult survival rate is that it is available for a large cross-section of countries. I construct the adult survival rate from the World Bank’s World Development Indicators. Specifically, this is the weighted average of male and female survival rates, weighted by the male and female share in the population.



Figure 3: Differences in years of schooling with the US



White bars; only broad sample. Grey bars: only broad and intermediate samples. Black bars: all samples (except TTO not in intermediate). Dashed line: broad sample mean. Light solid line: intermediate sample mean. Heavy solid line: narrow sample mean. Source: Barro and Lee (2013).

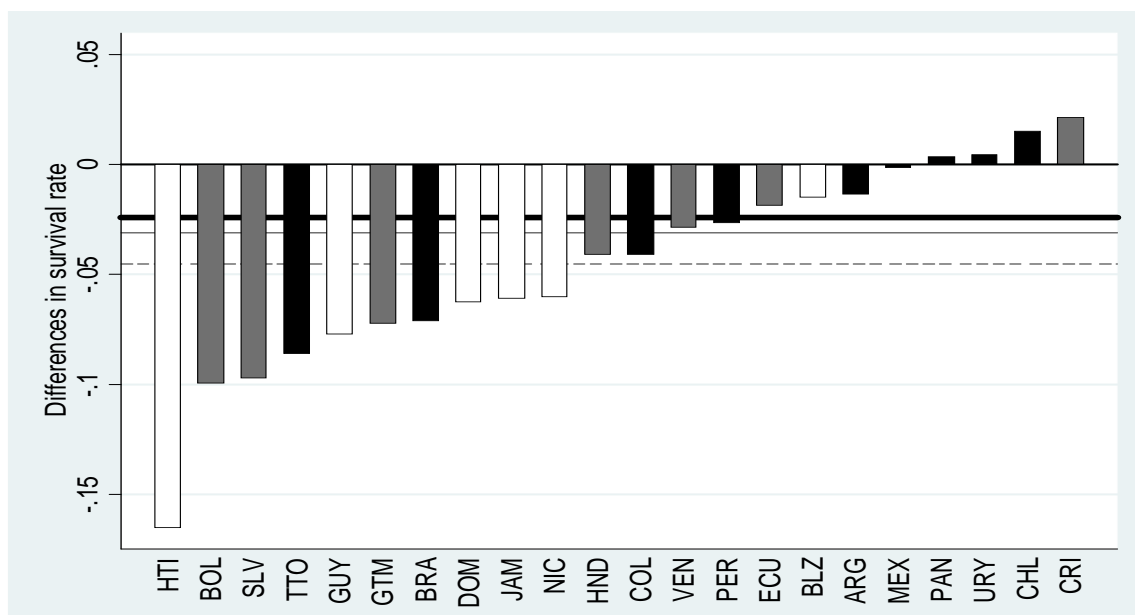
In Figure 4 I plot adult survival rate differences with the USA. Survival rate probabilities are lower in Latin America than in the US, but perhaps not vastly so. On average, Latin American 15-year olds are only 4 percentage points less likely to reach the age of 60 than US 15-year olds.<sup>6</sup>

Following work by Gundlach, Rudman, and Woessman (2002), Woessman (2003), Jones and Schneider (2010) and Hanushek and Woessmann (particularly 2012a), we also wish to account for differences in cognitive skills not already accounted for by years of schooling and health. The ideal measure would be a test of average cognitive ability in the working population. Hanushek and Zhang (2009) report estimates of one such test for a dozen countries, the International Adult Literacy Survey (IALS), but only one of these is in Latin America (Chile).

As a fallback, I rely on internationally comparable test scores taken by school-age children. In the narrow sample, I will use scores from a science test administered in 2009

<sup>6</sup>The population-weighted mean survival rate in the broad sample is 0.85.

Figure 4: Differences in survival rate with the US



White bars; only broad sample. Grey bars: only broad and intermediate samples. Black bars: all samples (except TTO not in intermediate). Dashed line: broad sample mean. Light solid line: intermediate sample mean. Heavy solid line: narrow sample mean. Source: WDI.

to 15 year olds by PISA (Program for International Student Assessment). There are in principle several other internationally-comparable tests (by subject matter, year of testing, and organization testing) that could be used in alternative to or in combination with the 2009 PISA science test. However there would be virtually no gain in country coverage by using or combining with other years (the PISA tests of 2009 are the ones with the greatest participation, and virtually no Latin American country participated in other worldwide tests and not in the 2009 PISA tests).<sup>7</sup> Focusing only on one test bypasses potentially thorny issues of aggregation across years, subjects, and methods of administration. Cross-country correlations in test results are very high anyway, and very stable over time.<sup>8</sup> Data on PISA test score results are from the World Bank’s Education Statistics.

Aside from the world-wide tests of cognitive skills used in the narrow sample, there are

<sup>7</sup>The only exception is Belize, which participated in some of the reading tests administered by PIRLS (Progress in International Reading Literacy Study).

<sup>8</sup>Repeating all my calculations using the PISA math scores yielded results that were virtually indistinguishable from those using the science test.

also two “regional” tests of cognitive skills that have been administered to a group of Latin American countries: the first in 1997 by the Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE), covering reading and math in the third and fourth grade; the second in 2006 by the Latin American bureau of the UNESCO, covering the same subjects in third and sixth grade. These tests are described in greater detail in, e.g., Hanushek and Woessman (2012a), who also argue that these tests may better reflect *within Latin-American* differences in cognitive skills.

From the perspective of this study, the main attraction of these alternative measures of cognitive skills is that they cover a significantly larger sample. The biggest problem, of course, is that they exclude the United States (or any high-income country) and so, on the face of it, they are unusable for constructing counterfactual relative incomes. However, Hanushek and Woessman (2012a) propose a methodology to “splice” the regional scores into their worldwide sample. While this splicing involves a large number of assumptions that are difficult to evaluate, it is worthwhile to assess the robustness of my results to these data.<sup>9</sup>

Needless to say measuring  $t$  by the above-described test scores is clearly very unsatisfactory, as in most cases the tests reflect the cognitive skills of individuals who have not joined the labor force as of 2005, much less those of the average worker. The average Latin American worker in 2005 was 36 years old, so to capture their cognitive skills we would need test scores from 1984.<sup>10</sup> Implicitly, then, we are interpreting test-score gaps in current children as proxies for test score gaps in current workers. If Latin America and the US have experienced different trends in cognitive skills of children since 1984 this assumption is problematic.

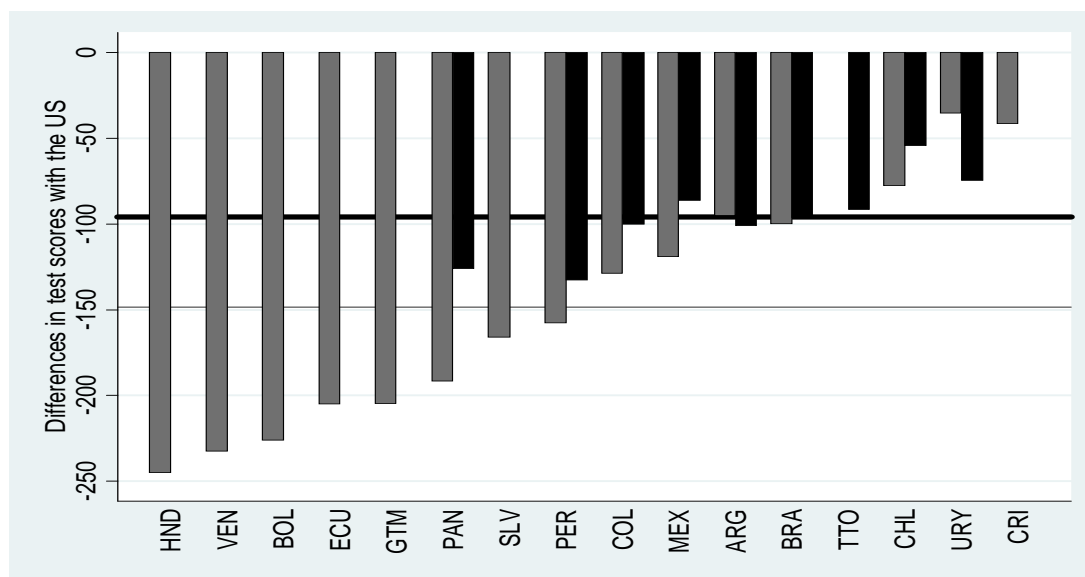
The 2009 PISA science tests are reported on a scale from 0 to 1000, and they are normalized so that the average score *among OECD countries* (i.e. among all pupils taking the test in this set of countries) is (approximately) 500 and the standard deviation is (approx-

---

<sup>9</sup>Hanushek and Woessman (2012a) splice the regional scores into world-wide scores that are themselves aggregates of multiple waves and multiple subject areas - obtained with a methodology described in Hanushek and Woessman (2012b).

<sup>10</sup>The method for estimating the average age of workers is described in footnote 25 of Caselli (2005).

Figure 5: Differences in test scores with the US



Grey bars: regional test/intermediate sample. Black bars: PISA test/narrow sample. Light solid line: regional-test mean. Heavy solid line: PISA-test mean. Source: World Bank's Education Statistics and Hanushek and Woessman (2012a, 2012b).

mately) 100.<sup>11</sup> The regional scores are put on the PISA scale by Hanushek and Woessman's splicing, so they can be directly compared. Figure 5 shows test score differences  $t_i - t_{US}$  for the narrow and intermediate samples. Differences in PISA scores are very significant: the average Latin American student in 2009 shows cognitive skills that are below those of his US counterpart by about one standard deviation of the OECD distribution of cognitive skills. Only Chile is a partial stand-out, with a cognitive gap closer to one half of one standard deviation. Differences in Hanushek and Woessman's spliced regional tests are even more significant, with the average gap exceeding 1.5 standard deviations. Recall that the PISA scores are directly comparable between Latin American and USA, while

<sup>11</sup>I say approximately in parenthesis because the normalization was applied to the 2006 wave of the test. The 2009 test was graded to be comparable to the 2006 one. Hence, it is likely that the 2009 mean (standard deviation) will have drifted somewhat away from 500 (100) - though probably not by much. The PISA math and reading tests were normalized in 2000 and 2003, respectively, so their mean and standard deviation are more likely to have drifted away from the initial benchmark. This is one reason why I use the science test for my baseline calculations.

the spliced regional tests – while arguably giving a more accurate sense of within Latin America differences – are less suitable for poor country-rich country comparisons. Hence, the discrepancy in cognitive-skill gaps between the PISA and the regional scores implies that the latter should be treated with caution.

## 4 Calibration

The last, and most difficult, step in producing counter-factual income gaps between US and Latin America is to calibrate the coefficients  $\beta_s$ ,  $\beta_r$ , and  $\beta_t$ . As discussed, equation (4) indicates that, using *within country* data on  $w$ ,  $s$ ,  $r$ , and  $t$ , one could in principle identify these coefficients by running an extended Mincerian regression for log-wages. In implementing this plan, we are confronted with (at least) two important problems.

The first problem is that one of the explanatory variables, the adult survival rate  $r$ , by definition does not vary within countries. Estimating  $\beta_r$  directly is therefore a logical impossibility. To solve this problem Weil (2007) notices that, in the time series (for a sample of ten countries for which the necessary data is available), there is a fairly tight relationship between the adult survival rate and average height. In other words, he postulates  $c_i = \alpha_c + \gamma_c r_i$ , where  $c_i$  is average height and the coefficient  $\gamma_c$  is estimated from the above-mentioned time series relation (he obtains a coefficient of 19.2 in his preferred specification). Since height does vary within countries as well as between countries, this opens the way to identifying  $\beta_r$  by means of the Mincerian regression

$$\log(w_{ij}) = \alpha_i + \beta_s s_{ij} + \beta_c c_{ij} + \beta_t t_{ij},$$

where  $\beta_r = \beta_c \gamma_c$ .<sup>12</sup>

The second problem is that measures of  $t$  are not consistent at the macro and at the micro level. In particular, while we do have micro data sets reporting both results from tests of cognitive skills and wages, the test in question is simply *a different test* from the tests we have available at the level of the cross-section of countries. Call the alternative test

---

<sup>12</sup>Needless to say if we had cross-country data on average height there would be no need to use the survival rate at all.

available at the micro level  $d$ . Once again the solution is to *assume* a linear relationship  $d_i = \gamma_d t_i$ . The difference with the case of height-survival rate is that, as far as I know, there is no way to check the empirical plausibility of this assumption. Given the assumed linear relationship, one can back out  $\gamma_d$  as the ratio of the *within country* standard deviation of  $d_{ij}$  and  $t_{ij}$ . With  $\gamma_d$  at hand, one can back out  $\beta_t$  from the modified Mincerian regression

$$\log(w_{ij}) = \alpha_i + \beta_s s_{ij} + \beta_c c_{ij} + \beta_d d_{ij}, \quad (5)$$

using  $\beta_t = \beta_d \gamma_d$ .

In choosing values for  $\beta_s$ ,  $\beta_c$ , and  $\beta_d$  from the literature it is highly desirable to focus on microeconomic estimates of equation (5) that include all three right-hand variables. This is because  $s$ ,  $c$ , and  $d$  are well-known to be highly positively correlated.<sup>13</sup> Hence, any OLS estimate of one of the coefficients from a regression that omits one or two of the other two variables will be biased upward.<sup>14</sup>

A search of the literature yielded one and only one study reporting all three coefficients from equation (5). Vogl (2014) uses the two waves (2002 and 2005) of the nationally-representative Mexican Family Life Survey to estimate (5) on a subsample of men aged 25-65. In his study,  $w$  is measured as hourly earnings,  $s$  as years of schooling,  $c$  is in centimeters, and  $d$  is the respondent's score on a cognitive-skill test administered at the time of the survey.<sup>15</sup> The cognitive skill measure is scaled so its standard deviation in the Mexican population is 1.<sup>16</sup>

The coefficients reported by Vogl are as follows (see his Table 4, column 7). The return to schooling  $\beta_s$  is 0.072, which can be plugged directly in equation (3). The “return to height”  $\beta_c$  is 0.013. Hence, the coefficient associated with the adult survival rate in (3)

---

<sup>13</sup>See, e.g., the literature review in Vogl (2014).

<sup>14</sup>An alternative would be to use IV estimates of the  $\beta$ s, but instruments for the variables on the right hand side of equation 5 are often somewhat controversial - especially for height and cognitive skills.

<sup>15</sup>The test is the short-form Raven's Progressive Matrices Test.

<sup>16</sup>Needless to say there are aspects of Vogl's treatment that imply the regressions he runs are not a perfect fit for the conceptual framework of the paper. It may have been preferable for our purposes to include both men and women. He also controls for ethnicity, age, and age squared, which do not feature in my framework. Finally, he notes that the Raven's core is a coarse measure of cognitive skills, giving raise to concerns with attenuation bias (more on this below).

is  $0.013 \times 19.2 = 0.25$ , where I have used Weil’s mapping between height and the adult survival rate. Finally, the reported return to cognitive skills  $\beta_d$  is 0.011. Since the standard deviation of  $d$  is one by construction, and the standard deviation of the 2009 Science PISA test in Mexico is 77, the implied coefficient on the PISA test for the purposes of constructing  $h$  is  $0.011/77=0.00014$ .<sup>17</sup>

The coefficients in my baseline calibration are considerably lower than those used in other development-accounting exercises. For schooling, applications usually gravitate towards the “modal” Mincerian coefficient of 0.10. For the adult survival rate, Weil (2007) uses 0.65, on the basis of considerably higher estimates of the returns to height than those reported by Vogl. For the return to cognitive skills, Hanushek and Woessmann (2012a) advocate 0.002, which is more than one order of magnitude larger than the value I derive from the Vogl’s estimates.<sup>18</sup>

The fact that the parameters calibrated on Vogl’s estimates are smaller than those commonly used is consistent with the discussion above. In particular, the alternative estimates are often based on regressions that omit one or two of the variables in (5), and are therefore upward biased. Another consideration is that there is considerable cross-country heterogeneity in the estimates, and that researchers often focus on estimates from the USA, which are often larger.<sup>19,20</sup>

---

<sup>17</sup>Hanushek and Woessman’s splicing procedure implies that the same coefficient can be used for the regional tests used in the intermediate sample. In particular, the relevant standard error is the average of the standard deviations of Pisa science and math tests in Mexico, which is 80. Then we have  $0.011/80 = 0.00014$ .

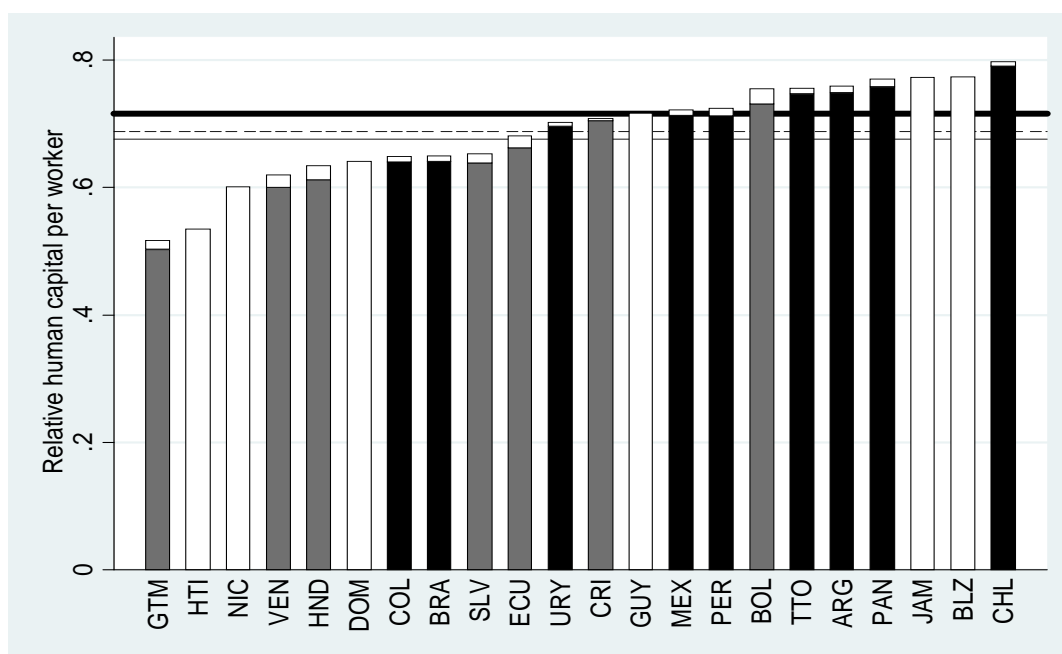
<sup>18</sup>This is based on Hanushek and Zhang (2009), who use the International Adult Literacy Survey (IALS) to estimate the return to cognitive skills in a set of 13 countries. The value of 0.002 is the one for the USA.

<sup>19</sup>For example, in Hanushek and Zhang (2009), the estimated market return to cognitive skills varies (from minimum to maximum) by a factor of 10! The estimate from the USA, which is used in Hanushek and Woessman (2012a) is the *maximum* of this distribution.

<sup>20</sup>This is actually an issue with the capital share  $\alpha$  as well. However, the issue there is less severe as observed capital shares do not vary systematically with  $y$ , so it should be possible to ascribe the observed variation to measurement error. In other words the patterns of variation in  $\alpha$  do not necessarily rise the issue of model misspecification.

On the other hand, Vogl’s regressions are admittedly estimated via OLS, and there is a real concern with attenuation bias from measurement error. In order to gauge the sensitivity of my results to possibly excessively low values of the calibration parameters due to attenuation bias, I will also present results based on an “aggressive” calibration, which uses a Mincerian return of 0.10, Weil’s 0.65 value for the mapping of the adult survival rate to human capital, and Hanushek and Woessman’s 0.002 coefficient on the PISA test.<sup>21</sup>

Figure 6: Human capital per worker relative to the US - baseline calibration



Overall height: relative human capital per worker without cognitive-skill correction. Grey (Black) bars: relative human capital per worker with cognitive-skill correction based on regional (PISA) tests. Dashed line: average with no cognitive-skill correction. Light (heavy) solid line: average with regional-(PISA-)test correction.

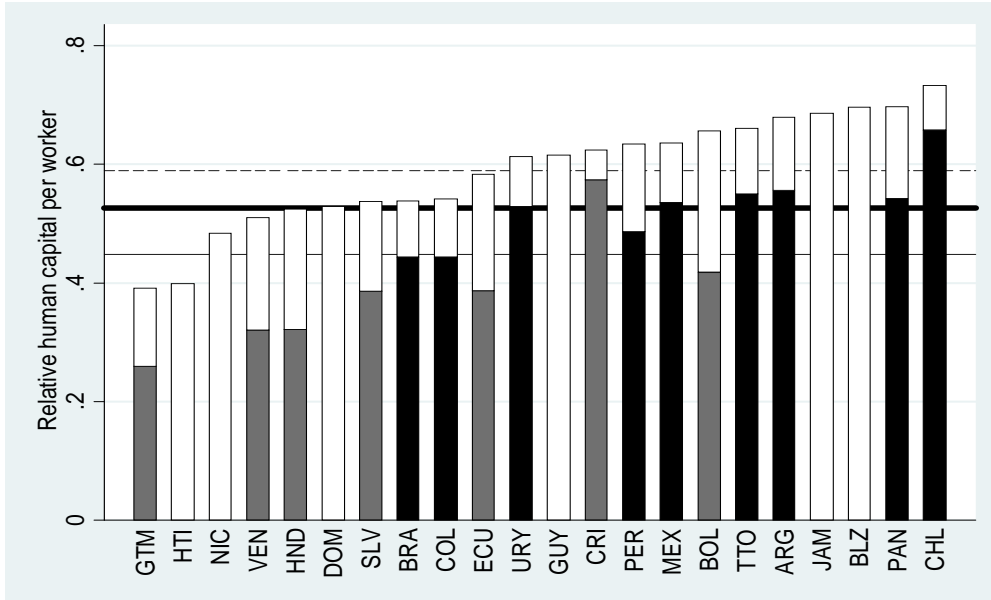
Figure 6 shows human-capital per worker estimates for Latin American countries relative to the US,  $h_i/h_{US}$ , under my baseline calibration. The full height of the bar shows

<sup>21</sup>As described above the Hanushek and Zhang estimate for the US comes from a test  $d$  different from  $t$ . In order to go from their coefficient  $\beta_d$  to the coefficient of interest  $\beta_t$  we need to multiply the former by the ratio of the standard deviation of  $d_{US,i}$  to the standard deviation of  $t_{US,i}$ . Since Hanushek and Zhang standardize the variable  $d$ , we just have to multiply by the inverse of the standard deviation of  $t_{US,i}$ . But in the test we are using this is just 0.98, so the correction would be immaterial. I use the same value both in the narrow and in the intermediate sample.



the value of  $h_i/h_{US}$  when excluding cognitive skills, and is thus fully comparable across all countries in the figure. The solid bars are the values when including cognitive skills. Irrespective of sample and cognitive-skill correction the average Latin American worker is endowed with approximately 70% of the human capital of the average US worker.

Figure 7: Human capital per worker relative to the US - aggressive calibration



Overall height: relative human capital per worker without cognitive-skill correction. Grey (Black) bars: relative human capital per worker with cognitive-skill correction based on regional (PISA) tests. Dashed line: average with no cognitive-skill correction. Light (heavy) solid line: average with regional-(PISA-)test correction.

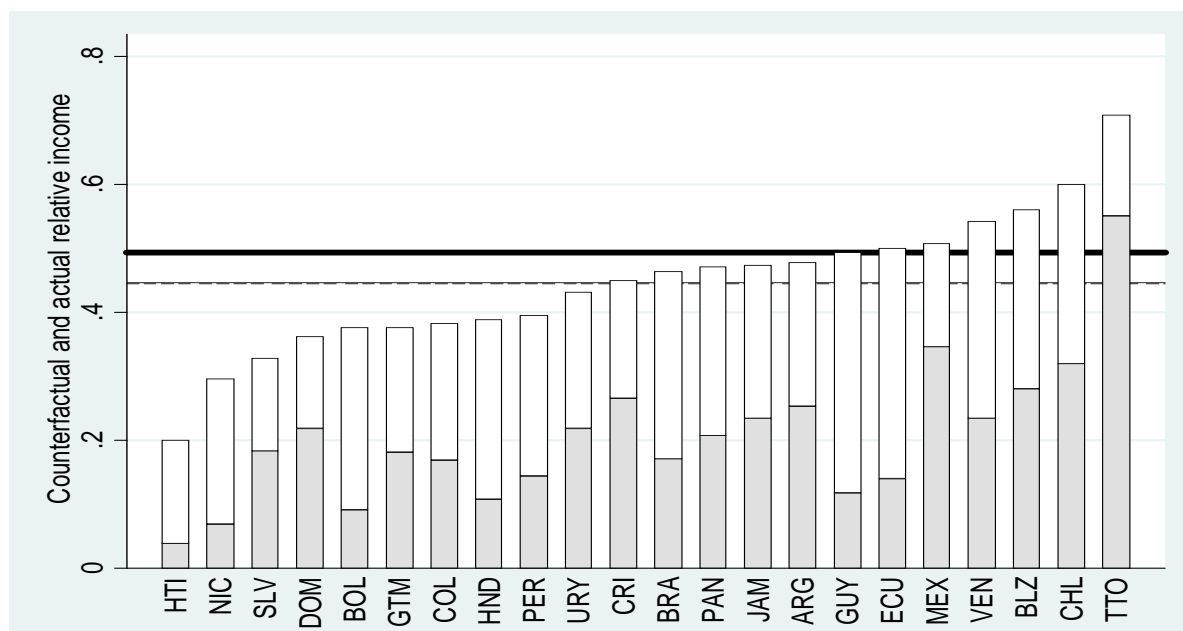
Figure 7 is analogous to Figure 6 but shows the aggressive calibration instead. Not Surprisingly, using the aggressive calibration results in significantly lower relative human capital for Latin America, since the impact of differentials in schooling, health, and cognitive skills is magnified. Human capital gaps become particularly large when including the cognitive-skill corrections.

## 5 Results

### 5.1 Baseline Calibration

In the large sample we lack cognitive skill information for more than half of the countries, so we set  $\beta_t = 0$ . Figure 8 shows each country's counterfactual income relative to the US (relative capital) in 2005,  $\tilde{y}_i/\tilde{y}_{US}$ , as well as the relative incomes  $y_i/y_{US}$  already shown in

Figure 8: Relative capital, baseline calibration, no cognitive-skill correction



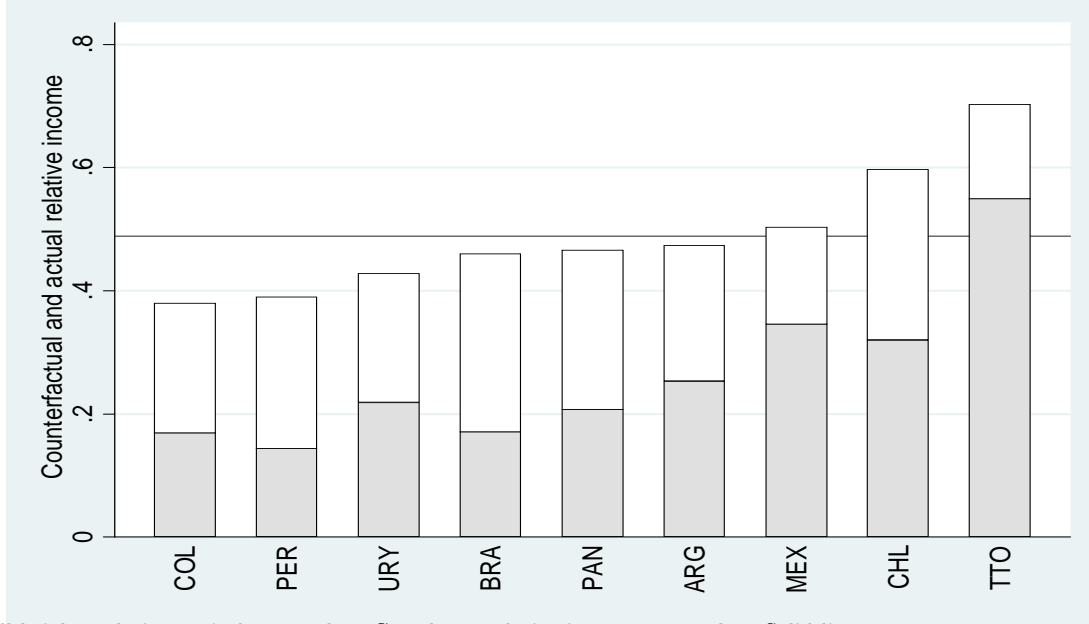
Overall height: relative capital per worker. Grey bars: relative income per worker. Dashed line: broad sample mean. Light solid line: intermediate sample mean. Heavy solid line: narrow sample mean.

Figure 1. In particular, for each country the overall height of the bar is relative capital, while the height of the shaded bar is relative income.

As is apparent, there is a lot of variation in relative capital, ranging from 20% to almost 70%. This reflects considerable heterogeneity in rates of physical and human-capital accumulation among Latin American countries, as seen above. Sample means are between 44% (broad and intermediate sample) and 49% (narrow sample). This means that observed distributions of physical and human capital are consistent with Latin American workers being between 44 and 49% as productive as USA ones. We can interpret this measure as a measure of the *capital gap* between Latin America and the US.

In Figure 9 we extend our calculations to include information on cognitive skills based on worldwide PISA test scores. The sample size correspondingly drops to 9 countries. The effect of including cognitive skills under my baseline calibration is virtually nil: the mean remains unchanged at 0.49. This result is expected given the very small calibrated “loading” on cognitive skills implied by Vogl’s estimates. Very similar patterns emerge when using the regional scores/intermediate sample, as seen in Figure 10.

Figure 9: Relative capital, baseline calibration, PISA cognitive skills



Overall height: relative capital per worker. Grey bars: relative income per worker. Solid line: mean.

## 5.2 “Aggressive” Calibration

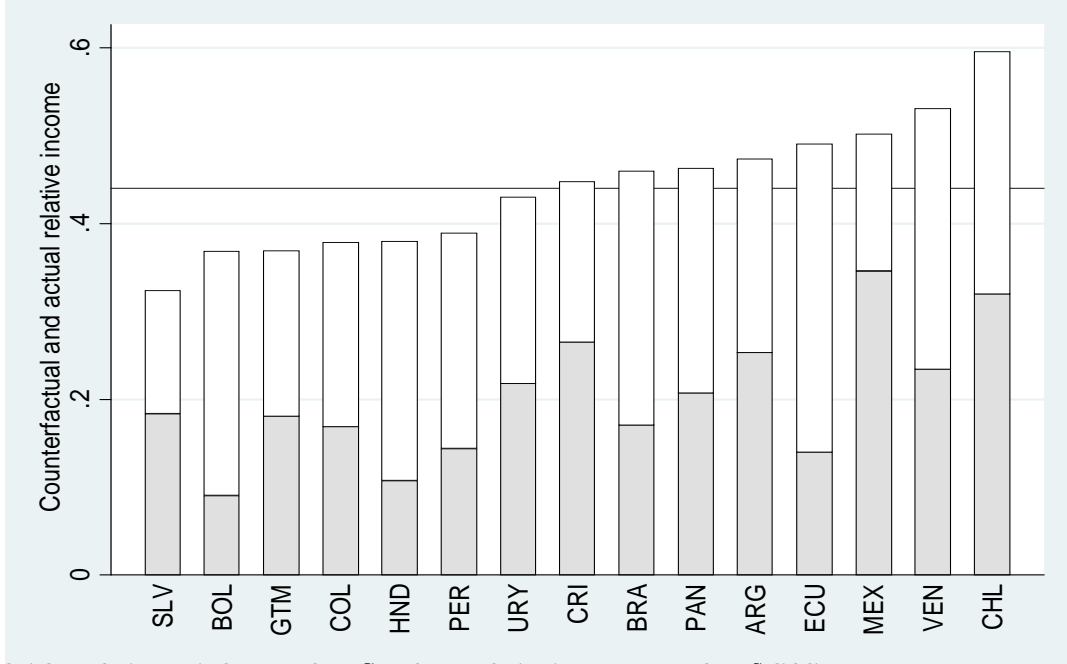
My baseline calibration uses coefficients for mapping years of schooling, health, and cognitive skills into human capital that, taken individually, are lower than those presented in other contributions. In this section I explore the robustness to my results to more commonly-used values. Hence, I set  $\beta_s = 0.10$ ,  $\beta_r = 0.65$ , and  $\beta_t = 0.002$ .

Results from the large sample using this aggressive calibration are shown in Figure 11. Given the larger coefficients, counterfactual incomes are necessarily smaller than under the baseline calibration. Yet quantitatively the difference is not very large. Average relative capital drops to 40%, so still roughly double relative income.

Figure 12 shows the results from the aggressive calibration using the PISA test scores. Including cognitive skills in the calculation of relative capital has a much bigger impact than under the baseline, because the coefficient on cognitive skills is an order of magnitude larger. The average counterfactual relative income falls to 40%, compared with 49% in the baseline calibration (within the same narrow sample). This is a large gain in explanatory power of observables. For many countries, the gap between relative income and relative capital shrinks considerably.

Finally, Figure 13 reports the results from the aggressive calibration when using the

Figure 10: Relative capital, baseline calibration, “regional-test” cognitive skills



Overall height: relative capital per worker. Grey bars: relative income per worker. Solid line: mean.

regional test scores. Recall that these tests tend to show even larger cognitive gaps with the US. Correspondingly, using these tests in combination with the aggressive calibration leads to an even better alignment between relative capital and relative income.

## 6 Implications for Efficiency Gaps

We have seen that, depending on cognitive skill correction, counterfactual income ratios (relative capital) in Latin America tend to be much larger than actual income ratios. This discrepancy implies that Latin America suffers from an *efficiency gap* as much as it suffers from a capital gap.

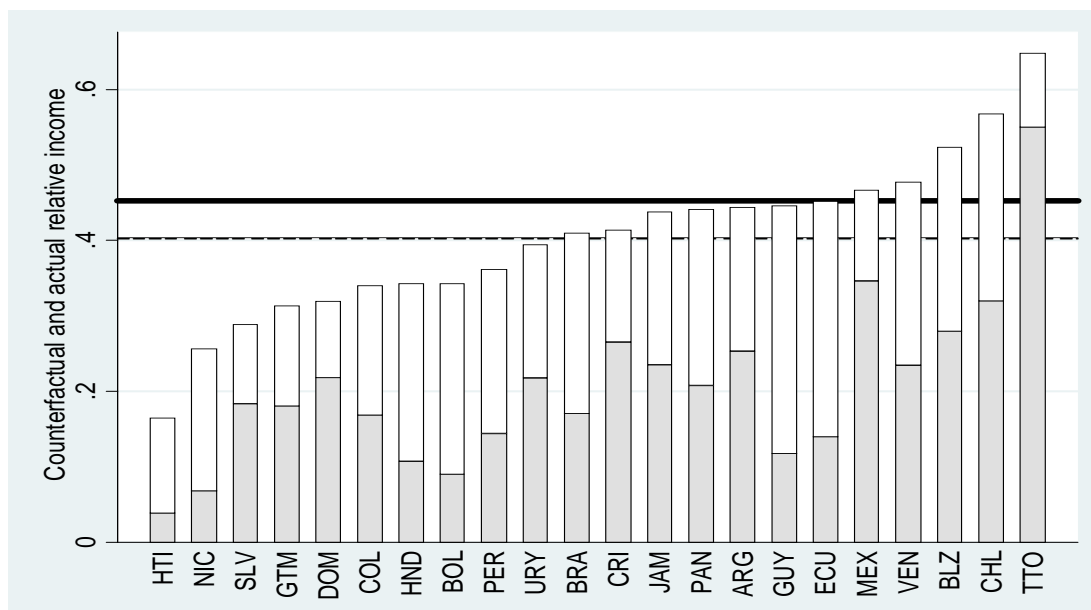
We can quantify efficiency gaps by noting, from (1) and (2), that

$$\frac{A_i}{A_{US}} = \frac{y_i/y_{US}}{\tilde{y}_i/\tilde{y}_{US}}.$$

Hence, Latin American efficiency gaps can be directly gleaned from Figures (8)-(13) by simply dividing the height of the shaded bars by the overall height of the bars.

In Table 1 I report the sample averages of the implied efficiency gaps, for the various cognitive skill correction - calibration combinations. For completeness I also report

Figure 11: Relative capital, aggressive calibration, no cognitive-skill correction



Overall height: relative capital per worker. Grey bars: relative income per worker. Dashed line: broad sample mean. Light solid line: intermediate sample mean. Heavy solid line: narrow sample mean.

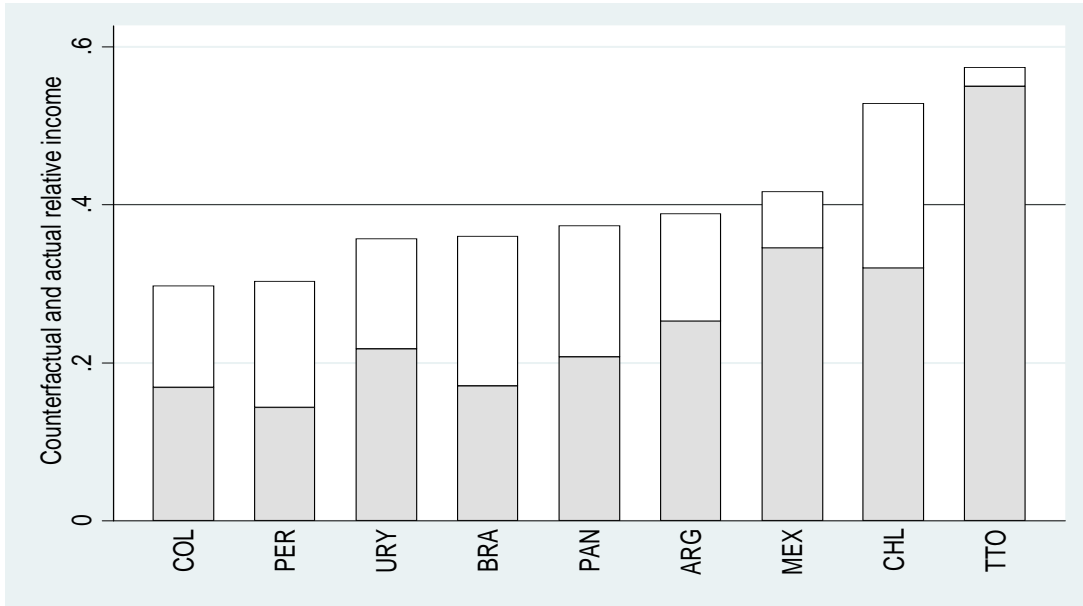
the corresponding averages for relative income and relative capital, as well as labor-force weighted means.

Using the baseline calibration, average relative capital and average relative efficiency are almost identical, irrespective of sample/cognitive skill correction/weighting. One way to put this is that capital gaps and efficiency gaps contribute equally to Latin American income gaps. When using the aggressive calibration, the relative importance of capital gaps increases, particularly when adding the cognitive-skill corrections. Still, even under the most aggressive scenario average Latin American relative efficiency is only 60% of US efficiency.<sup>22</sup>

In order to fully appreciate the importance of these efficiency gaps it is crucial to note that, under almost any imaginable set of circumstances, physical (specifically, reproducible) and human capital accumulation respond to a country's level of efficiency. The higher  $A$

<sup>22</sup>In the narrow sample it is probably best to focus on the labor-force weighted results, as the unweighted ones give disproportionate weight to Trinidad and Tobago.

Figure 12: Relative capital, aggressive calibration, PISA cognitive skills



Overall height: relative capital per worker. Grey bars: relative income per worker. Solid line: mean.

the higher the marginal productivity of capital, leading to enhanced incentives to invest in equipment and structure, schooling, etc. While quantifying this effect is difficult, most theoretical frameworks would lead one to expect it to be large. Hence, it is legitimate to conjecture that a significant fraction of the capital gap may be due to the efficiency gap.<sup>23</sup>

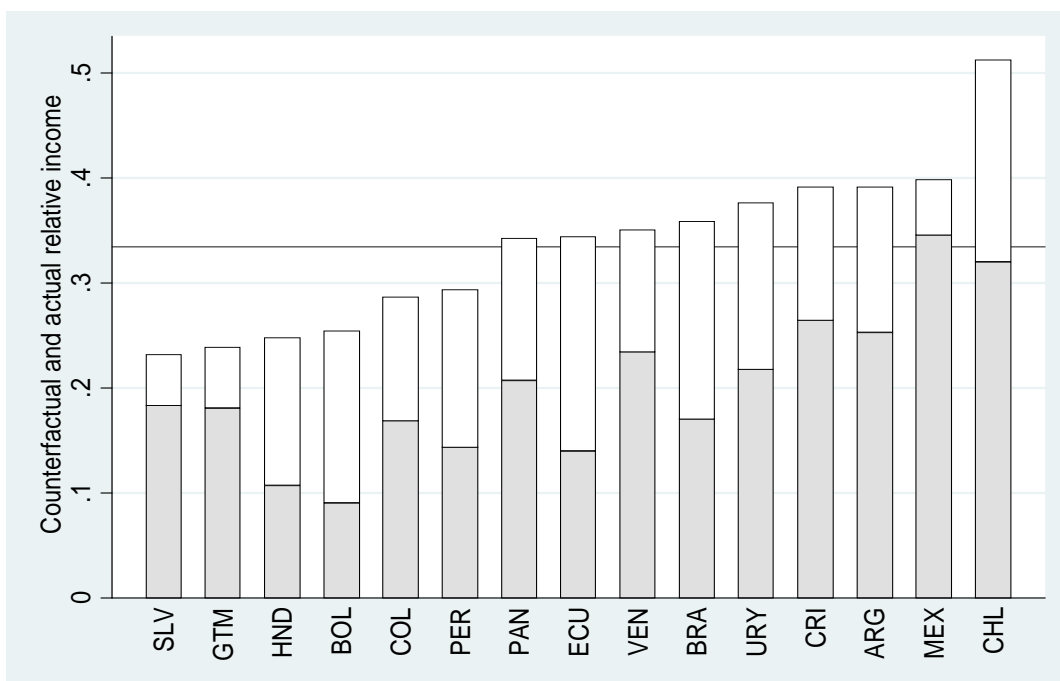
## 7 Implications and Conclusions

There is a large gap in income per worker between Latin America and the USA: Latin American workers are only about one fifth as productive as workers from the United States. A development-accounting calculation reveals that both capital gaps and efficiency gaps contribute to this overall productivity gap. In particular, a Cobb-Douglas aggregate

---

<sup>23</sup>In principle, one might also argue for a reverse direction of causation, with larger physical and human-capital stocks leading to higher efficiency. In particular, this would be true if the model was misspecified, and there were large externalities. But as already mentioned the empirical literature has not to date uncovered significant evidence of externalities in physical and human capital.

Figure 13: Relative capital, aggressive calibration, “regional-test” cognitive skills



Overall height: relative capital per worker. Grey bars: relative income per worker. Solid line: mean.

of observable physical and human capital per worker is roughly in the order of 40% on average of the corresponding US level (capital gap) – implying that the efficiency with which inputs are used in Latin America is in the order of 50% of US levels (efficiency gap). Reducing this efficiency gap would reduce the overall productivity gap both directly, by allowing Latin America to reap greater benefits from its physical and human capital, and indirectly, since much of the capital gap is likely due to the efficiency gap itself: closing the efficiency gap would stimulate investment at rates potentially capable of closing the capital gap as well.

These conclusions are contingent on the quality of the underlying macroeconomic data. There is growing concern about the quality and reliability of the PPP national-account figures in the Penn World Tables and similar data sets [e.g. Johnson et al. (2013)]. Similar concerns apply, no doubt, to our proxies for human capital as well (as already discussed particularly in the context of cognitive skills). It is true that such concerns are most often voiced in the context of implied comparisons of changes, especially over short time spans: cross-country comparisons of *levels* reveal such gigantic differences (as seen above) that they seem unlikely to be entirely dominated by noise. Still, exclusive reliance on these macro data is highly inadvisable.

Table 1: Summary of Results

Sample/Cognitive Skill Measure	Relative GDP	Calibration			
		Baseline		Aggressive	
		Relative Capital	Relative Efficiency	Relative Capital	Relative Efficiency
Broad/None	<b>0.21</b>	<b>0.44</b>	<b>0.44</b>	<b>0.40</b>	<b>0.49</b>
	0.21	0.46	0.45	0.41	0.50
Narrow/PISA	<b>0.26</b>	<b>0.49</b>	<b>0.52</b>	<b>0.40</b>	<b>0.64</b>
	0.22	0.46	0.47	0.37	0.59
Intermediate/“Regional”	<b>0.20</b>	<b>0.44</b>	<b>0.45</b>	<b>0.33</b>	<b>0.60</b>
	0.22	0.46	0.46	0.36	0.60

Bold entries are unweighted sample means. Plain entries are labor-force weighted sample means

Fortunately, it is also increasingly unnecessary. The increasing availability of *firm level* data sets, particularly when matched with employee-level information (e.g. about schooling), provides an opportunity to supplement the macro picture with microeconomic productivity estimates comparable across countries.

The benefit of producing such micro productivity estimates is by no means limited to permitting to check the robustness of conclusions concerning average capital and efficiency gaps - though this benefit alone is sufficient to make such exercises worthwhile. An additional benefit is to uncover information on the within country *distribution* of physical capital, human capital, and efficiency. A relatively concentrated distribution would suggest that efficiency gaps are mostly due to aggregate, macroeconomic factors that affect all firms fairly equally (e.g. impediment to technology diffusion from other countries). A very dispersed distribution, with some firms close to the world technology frontier, would be more consistent with allocative frictions that prevent capital and labor to flow to the more efficient/talented managers.

More generally, firm-level data is likely to prove essential in the quest for the determinants of the large efficiency gaps revealed by the development-accounting calculation. After all, (in-)efficiency is – by definition – a firm-level phenomenon. Most of the most plausible



possible explanations for the efficiency gap are microeconomic in nature – whether it is about firms unable to adapt technologies developed in more technologically-advanced countries, failures in the market for managers and/or capital, frictions in the matching process for workers, etc. It seems implausible that evidence for or against these mechanisms can be found in the macro data. Yet understanding the sources of the Latin American efficiency gap is unquestionably the most urgent task for those who want to design policies aimed at closing the Latin American income gap.

## References

Barro and Lee (2013): "A New Data Set of Educational Attainment in the World, 1950-2010." *Journal of Development Economics*.

Caselli, F. (2005): "Accounting for cross-Country Income Differences," in Philippe Aghion and Stephen Durlauf (eds.), *Handbook of Economic Growth*, Volume 1A, 679-741, Elsevier.

Caselli, Francesco, and Antonio Ciccone (2013): "The Contribution of Schooling in Development Accounting: Results from a Nonparametric Upper Bound." *Journal of Development Economics*.

Caselli, Francesco and Coleman, John Wilbur II, 2006. "The World Technology Frontier." *American Economic Review*, 96, pp. 499-522.

Caselli, Francesco, and James Feyrer, 2007. "The Marginal Product of Capital." *Quarterly Journal of Economics*, 122, pp. 535-568.

Gollin, Douglas, 2002. "Getting Income Shares Right." *Journal of Political Economy*, 110, pp. 458-474.

Gundlach, Rudman, and Woessman (2002): "Second Thoughts on Development Accounting," *Applied Economics*, 34, 1359-69.

Hall, Robert and Charles Jones, 1999. "Why Do Some Countries Produce So Much More Output Per Worker Than Others?" *Quarterly Journal of Economics*, 114, pp. 83-116

Hanushek, Eric, and Ludger Woessman (2012a): "Schooling, Educational Achievement, and the Latin American Growth Puzzle", *Journal of Development Economics* 99 (2), 497-512,

Hanushek, Eric A.,Woessmann, Ludger, (2012b). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth*.

Hanushek, Eric and Lei Zhang (2009): "Quality-Consistent Estimates of International Schooling and Skill Gradients," *Journal of Human Capital*, 2009, 3(2), 107-43.

Iranzo, Susana and Giovanni Peri, 2009. "Schooling Externalities, Technology, and Productivity: Theory and Evidence from US States." *Review of Economics and Statistics*, 91, pp. 420-431

Johnson, Simon, William Larson, Chris Papageorgiou, and Arvind Subramanian (2013): "Is Newer Better? Penn World Table Revisions and Their Impact on Growth Estimates,"

*Journal of Monetary Economics*, March 2013.

Klenow, Peter, and Andres Rodriguez-Claire, 1997. "The Neoclassical Revival in Growth Economics: Has It Gone Too Far?". In NBER Macroeconomic Annual, MIT Press

Vogl, Tom S. (2014): "Height, Skills, and labor market outcomes in Mexico," *Journal of Development Economics*, 107, 84-96.

Weil, David, 2007. "Accounting for the Effect of Health on Economic Growth." *Quarterly Journal of Economics*, 122, pp.1265-1306.

Woessman, Rudiger (2003): "Specifying Human Capital," *Journal of Economic Surveys*, 17, 3, 239-270.

World Bank (2012): *The Changing Wealth of Nations*.