NBER WORKING PAPER SERIES

ACCOUNTING FOR CROSS-COUNTRY INCOME DIFFERENCES

Francesco Caselli

Working Paper 10828
http://www.nber.org/papers/w10828

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2004

This is the draft of a chapter on Development Accounting for the forthcoming Handbook of Economic Growth, edited by P. Aghion and S. Durlauf. A previous draft circulated with the title: "The Missing Input: Accounting ..." A data set will be posted at www.economics.harvard.edu. I thank Silvana Tenreyro for comments, and Mariana Colacelli, Kalina Manova, and Andrea Szab`o for research assistance. The views expressed herein are those of the author(s) and not necessarily those of the National Bureau of Economic Research.

Accounting for Cross-Country Income Differences
Francesco Caselli
NBER Working Paper No. 10828
October 2004
JEL No. E2, O1, O2, O3, O4, O5

ABSTRACT

Why are some countries so much richer than others? Development Accounting is a first-pass attempt at organizing the answer around two proximate determinants: factors of production and efficiency. It answers the question ``how much of the cross-country income variance can be attributed to differences in (physical and human) capital, and how much to differences in the efficiency with which capital is used?" Hence, it does for the cross-section what growth accounting does in the time series. The current consensus is that efficiency is at least as important as capital in explaining income differences. I survey the data and the basic methods that lead to this consensus, and explore several extensions. I argue that some of these extensions may lead to a reconsideration of the evidence.

Francesco Caselli
Harvard University
Department of Economics
Littauer Building
Cambridge, MA 02138
and NBER
caselli@harvard.edu

Contents

1 Introduction

This chapter is about development accounting. It is widely known, and will be found again to be true here, that cross-country differences in income per worker are enormous. Development accounting uses cross-country data on output and inputs, at one point in time, to assess the relative contribution of differences in factor quantities, and differences in the efficiency with which those factors are used, to these vast differences in per-worker incomes. Hence, it is the same idea of growth accounting (illustrated by Jorgenson's chapter in this Handbook), with cross-country differences replacing cross-time differences.

Conceptually, development accounting can be thought of as quantifying the relationship

$$\text{Income} = F(\text{Factors, Efficiency}). \tag{1}$$

Like growth accounting, this is a potentially useful tool. If one found that Factors are able to account for most of the differences, then development economics could focus on explaining low rates of factor accumulation. There would of course be ample scope for controversy over the policies better suited to engineering higher investment rates in various types of capital, but there would be consensus over the fact that the intermediate goal of development policy is to engineer those higher rates. Instead, should one find that Efficiency differences play a large role, then one would have to confront the additional task of explaining why some countries extract more output than others from their factors of production. Experience suggests that this additional question is the hardest to crack.

The consensus view in development accounting is that Efficiency plays a very large role. A sentence commonly used to summarize the existing literature sounds something like "differences in efficiency account for at least 50% of differences in per capita income." The next section of this chapter (Section 2) will survey the existing literature, replicate its basic findings, and update them with more recent data. Looking at a cross-section of 93 countries in the year 1996, I confirm that standard procedures assign to Efficiency the role of the chief culprit.

Operationally, the key steps in development accounting are: (1) choosing a functional form for F, and (2) accurately measuring Income and Factors. Efficiency is backed out as a residual. As for the Solow residual, this residual is a "measure of our ignorance" on the causes of poverty and under-development. And, as in growth accounting, one potentially promising research strategy is to try to "chip away" at this residual by improving on steps (1) and (2), i.e. by looking at alternative functional forms, and by attempting a more sophisticated measurement of Income and Factors. For example, one could try to include information on quality differences in the capital stock – instead of relying exclusively on quantity.

The bulk of this chapter aims at outlining strategies for such a chipping away.[1] It investigates the potential for different functional forms, and different ways of estimating inputs and outputs, to reduce the measure of our ignorance. Rather than reaching firm conclusions, it tries to classify ideas into more or less promising. Its contribution is to formulate sentences such as "improvements in the measurement of x are unlikely to significantly reduce the unexplained component of per-capita income differences," or "the unexplained component is somewhat sensitive to the measurement of z, so this is a potentially fruitful area for further research."

The experiments I perform fall in five broad categories. The first is a fairly mechanical set of robustness checks with respect to the choice of parameters in the basic model used in the literature, as well as with respect to possible measurement errors in output, labor, and years of schooling. I conclude that none of these robustness checks seriously calls into question the conclusions from the current consensus (Section 3).

Second, I consider extensions of the basic development-accounting framework aimed at improving the measurement of human capital. In most development-accounting exercises differences in human capital stem exclusively from differences in the quantity of schooling. One set of extensions I consider exploits cross-country data on school resources and test scores as proxies for the quality of education, and then uses these quality indicators to augment the quantity-based measure of human capital. I find that taking into account schooling quality leads to trivially small reductions in the measure of our ignorance. Another extension replicates existing work that augments human capital by a proxy of the health status of the labor force. There is some indication that this may lead to a significant reduction in the unexplained component of income, but I argue that the bulk of the variance most likely remains unexplained. All the measures of human capital considered are built on the assumption that the private return to human capital accurately describes its social return. I conclude this section with a brief discussion of why and how one may want to try and relax this assumption (Section 4).

Third, I turn to the measurement of physical capital. Here I review contributions that highlight enormous cross-country variation in the composition of the stock of equipment. A simple model shows how to relate variation in capital composition to unobserved quality differences in the capital stock. How much of the responsibility for efficiency differences can be assigned to these differences in the quality of capital depends on parameters that are very hard to pin down, but the potential is extremely large. I therefore conclude that the composition of capital should be a key area of future research. I also glance at vintage-

[1] The analogy in spirit with Jorgenson's monumental contribution in growth accounting – some of which is collected in Jorgenson (1995a, 1995b) is obvious, but it stops there: the reader should expect nothing like the same level of depth, comprehensiveness, and insight.

capital models, but argue that they hold little promise for development accounting, as well as at the distinction between private and public investment, which is instead potentially quite important. (Section 5).

The most innovative contributions of the chapter are represented by the fourth and fifth sets of extensions. In the former I explore the role of the sectorial composition of output. The large differences in overall efficiency that are found at the aggregate level could reflect large differences in efficiency within each sector of the economy, but they could also be due to the fact that some countries have more of their inputs in intrinsically less productive sectors than others. I explore this idea by looking at an agriculture/non-agriculture decomposition (poor countries have as much as 90% of their workforce in agriculture, rich countries as little as 3%), but find that only a small fraction of the overall variation in efficiency is due to differences in sectorial composition: Efficiency differences appear to be a *within industry* phenomenon (Section 6).

The last set of exercises explores a radical departure from the standard framework, and finds radically different answers. In the standard framework efficiency differences are factor neutral: if a country uses physical capital efficiently, it also necessarily uses human capital efficiently. I argue that this is a pretty restrictive assumption, and propose a simple generalization of the basic framework where cross-country efficiency differences are allowed to be non neutral. Stunningly, I find that, when non neutrality is allowed for, the data say that poor countries use physical capital more efficiently than rich countries (while rich countries use human capital more efficiently). Furthermore, when the development-accounting exercise is performed in a context of non neutral efficiency differences the conclusions on the contribution of these differences to cross-country income inequality become very fragile. In particular, if the elasticity of substitution between physical and human capital is low enough, observed differences in factor endowments become able to explain the bulk of the income distribution. I therefore conclude that the most important outstanding question in development accounting may well be what this elasticity of substitution is (Section 7).

Before plunging into the data and the calculations, it is worthwhile to stress the limits of development accounting. Development accounting does not uncover the ultimate reasons why some countries are much richer than others: only the proximate ones. Like growth accounting, it has nothing to say on the causes of low factor accumulation, or low levels of efficiency. Indeed, the most likely scenario is that the same ultimate causes explain both. Furthermore, it has nothing to say on the way factor accumulation and efficiency influence each other, as they most probably do. Instead, it should be understood as a diagnostic tool, just as medical tests can tell one whether or not he is suffering from a certain ailment, but cannot reveal the causes of it. This does not make the test any the less useful.

2 The Measure of Our Ignorance

The key empirical result that motivates this chapter is that in a simple framework with two factors of production, physical and human capital, a large fraction of the cross-country income variance remains unexplained. This result has been established by a variety of authors using a variety of techniques. Knigth, Loayza, and Villanueva (1993), Islam (1995), and Caselli, Esquivel and Lefort (1996), for example, used panel-data techniques to estimate (1). They all found that, after controlling for factor accumulation, country-specific effects played a large role in output differences, and interpreted these fixed effects as picking up differences in efficiency. King and Levine (1994), Klenow and Rodriguez-Clare (1997), Prescott (1998), and Hall and Jones (1999), instead, used a calibration approach, and found that plausible parametrizations of (1) had limited explanatory power without large efficiency differences. These studies used cross-country national-account data on inputs and outputs, but Hendricks (2002) was able to reach similar conclusions by using earnings of migrants to the United States, and Aiyar and Dalgaard (2002) by using a dual approach involving factor prices rather than quantities. All these papers were inspired by – and written in response to the challenged posed by – the seminal contribution of Mankiw, Romer, and Weil (1992).[2]

In this section I revisit the basic development-accounting finding. Because I want to set the stage for a variety of extensions of the basic model, I adopt the calibration approach, which offers more flexibility in experimenting with different parameter values and functional forms.[3]

I adopt as the benchmark Hall and Jones' production function, according to which a country's GDP, Y, is

$$Y = AK^{\alpha}(Lh)^{1-\alpha}, \tag{2}$$

where K is the aggregate capital stock and Lh is the "quality adjusted" workforce, namely the number of workers L multiplied by their average human capital h. α is a constant. Clearly this is a special case of (1), where the residual A represents the efficiency with which factors are used. It is also clear that A corresponds to the standard notion of Total Factor Productivity (TFP), so until further notice I will speak of efficiency and TFP interchangeably.

In per-worker terms the production function can be rewritten as

$$y = Ak^{\alpha}h^{1-\alpha}, \tag{3}$$

[2]However, there are some pre-1990s antecedents. In particular, the nine–country studies of Denison (1967), and Christensen, Cummings, and Jorgenson (1981).

[3]An earlier survey of the material covered in this section is provided by McGrattan and Schmitz (1999). See also Easterly and Levine (2001) for a review of development accounting as well as other evidence for cross-country efficiency differentials.

where k is the capital labor ratio ($k = K/L$). We want to know how much of the variation in y can be explained with variation in the observables, h and k, and how much is "residual' variation, i.e. must be attributed to differences in A. Clearly to answer this question we need, besides data on y, data on k and h, as well as a value for the capital share α.

2.1 Basic Data

The basic data set used in this chapter combines variables from two sources. The first is version 6 of the Penn World Tables [PWT6 - Heston, Summers, and Aten (2002)], i.e. the latest incarnation of the celebrated Summers and Heston (1991) data set. From PWT6 I extract output, capital, and the number of workers. The second is Barro and Lee (2001), which I use for educational attainment. Several additional data sources will be brought to bear for specific exercises in later sections, but the data we construct here will be crucial to everything we do.

Previous authors have mostly used version 5.6 of the Penn World Tables (PWT56). They have therefore attempted to explain the world income distribution as of the late 1980s. By using version 6.0 I am able to update the basic result to the mid-90s.

I measure y from PWT6 as real GDP per worker in international dollars (i.e. in PPP - this variable is called RGDPW in the original data set).[4]

I generate estimates of the capital stock, K, using to the perpetual inventory equation

$$K_t = I_t + \delta K_{t-1},$$

where I_t is investment and δ is the depreciation rate. I measure I_t from PWT6 as real aggregate investment in PPP.[5] Following standard practice, I compute the initial capital stock K_0 as $I_0/(g+\delta)$, where I_0 is the value of the investment series in the first year it is available, and g is the average geometric growth rate for the investment series between the first year with available data and 1970. The rationale for this choice is tenuous: $I/(g+\delta)$ is the expression for the capital stock in the steady state of the Solow model. We will see below if results are very sensitive to this assumption, or for that matter to the others I am about to make, such as the one for δ, which – following the literature – I set δ to 0.06. To compute

[4]Some authors subtract from the PWT measure of GDP the value-added of the mining industry, because not doing so would result in some oil-rich countries being among the most productive in the world. This rationale is inherently dubious (then why not substracting the value added of agriculture and forestry, that also use natural resources abundantly?). More importantly, since a similar correction is not feasible for the capital stock, this procedure must result in hugely downward biased estimates of the TFP of these countries. I apply no such correction here.

[5]Computed as RGDPL*POP*KI, where RGDPL is real income per capita obtained with the Laspeyres method, POP is the population, and KI is the investment share in total income.

k, I divide K by the number of workers.[6]

To construct human capital I take from Barro and Lee (2001) the average years of schooling in the population over 25 year old. Following Hall and Jones (1999) this is turned into a measure of h through the formula:

$$h = e^{\phi(s)},$$

where s is average years of schooling, and the function $\phi(s)$ is piecewise linear with slope 0.13 for $s \leq 4$, 0.10 for $4 < s \leq 8$, and 0.07 for $8 < s$.[7] The rationale for this functional form is as follows. Given our production function, perfect competition in factor and good markets implies that the wage of a worker with s years of education is proportional to his human capital. Since the wage-schooling relationship is widely thought to be log-linear, this calls for a log-linear relation between h and s as well, or something like $h = \exp(\phi_s s)$, with ϕ_s a constant. However, international data on education-wage profiles (Psacharopulos, 1994) suggests that in Sub-Saharan Africa (which has the lowest levels of education) the return to one extra year of education is about 13.4 percent, the World average is 10.1 percent, and the OECD average is 6.8 percent. Hall and Jones's measure tries to reconcile the log-linearity at the country level with the convexity across countries.

s is observed in the data every five years, most recently in 2000. Since s moves slowly over time, a quinquennial observation can plausibly be employed for nearby dates as well.

I treat a country as having "complete data" at date t if it has an uninterrupted investment series between 1960 and t, *and* it has an observation for s in 1995.[8] With this definition, there are 93 countries with complete data in 1995 and 1996, 89 in 1997, and 86 in 1998 (and 0 thereafter). Hence, I focus on 1996 as the most recent year that affords the largest sample. In this sample, for more than half of the countries the investment series starts in 1950.[9]

As is well known, per-capita income differences are enormous. The richest country in the sample (USA) has income per worker equal to 57,000 1996 international dollars, while the poorest (Zaire, today's Democratic Republic of the Congo) has 605 – a ratio of 94. The ratio between the 90th (France) and the 10th percentile (Togo) of the income distribution, a measure of dispersion I'll use prominently in the rest of the paper, is 20. The log-variance, another measure I'll rely on heavily, is 1.26.

[6] Obtained as RGDPCH*POP/RGDPW, where RGDPCH is real GDP per capita computed with the chain method.

[7] Specifically we have $\phi(s) = 0.134 \cdot s$ if $s \leq 4$, $\phi(s) = 0.134 \cdot 4 + 0.101 \cdot (s-4)$ if $4 < s \leq 8$, $\phi(s) = 0.134 \cdot 4 + 0.101 \cdot 4 + 0.068 \cdot (s-8)$ if $8 < s$.

[8] Availability of data on income and labor force are not binding given these constraints.

[9] Nicaragua in 1979 has $I < 0$, which we deal with by re-setting it to $I = 0$. Haiti has missing data on I in 1966, which we deal with by imputing the average of 1965 and 1967.

For the last ingredient required by equation (3), α, I (implicitly) use US time-series data on the capital-share, whose long-run (and roughly constant) average value is 1/3. All these data choices will be subject to scrutiny in the rest of the chapter – indeed, this scrutiny is one of the chapter's contributions.

2.2 Basic Measures of Success

With data on k, h, and y, and a choice for α, equation (3) is one equation in the unknown A. In particular, after defining $y_{KH} = k^{\alpha}h^{1-\alpha}$, we can rewrite (3) as

$$y = A\, y_{KH}, \tag{4}$$

where both y and y_{KH} are measurable. I will refer to y_{KH} as the *factor-only model*.

Throughout this chapter I will pursue the following version of the development-accounting question: how successful is the factor-only model at explaining cross-country income differences? In other words, I will compare the (observed) variation in y_{HK} to the (observed) variation in y. Clearly, this means that I am asking the following question. Suppose that all countries had the same level of efficiency A: what would the world income distribution look like in that case, compared to the actual one?

To perform this assessment, I will look at two alternative measures. The first one is in the tradition of variance decompositions. From (4) we have

$$\text{var}[\log(y)] = \text{var}[\log(y_{KH})] + \text{var}[\log(A)] + 2\text{cov}[\log(A), \log(y_{KH})]. \tag{5}$$

Now notice that if all countries had the same level of TFP we would have var$[\log(A)]$ =cov$[\log(A), \log(y_{KH})]$ = 0. Hence, a first measure of success of the factor-only model is

$$success_1 = \frac{\text{var}\left[\log(y_{KH})\right]}{\text{var}\left[\log(y)\right]}.$$

In our data the counterfactual variance, var$[\log(y_{KH})]$, takes the value 0.5. Since the observed variance of $\log(y)$, var$[\log(y)]$, is 1.25 this approach leads to the conclusion that the fraction of the variance of income explained by observed endowments is $success_1$ =0.4.

While $success_1$ is nicely grounded in the tradition of variance decomposition, it has the well-known drawback that variances are sensitive to outliers. A measure that is less sensitive to outliers is a measure of the inter-percentile differential. Define x^p the value of the p_{th} percentile of the distribution of x. My second measure of success of the factor-differences model is

$$success_2 = \frac{y_{KH}^{90}/y_{KH}^{10}}{y^{90}/y^{10}},$$

i.e. it compares what the $90_{th} - to - 10_{th}$ percentile ratio would be in the counterfactual world with common technology, to the actual value. In the data the value of y_{KH}^{90}/y_{KH}^{10} is 7.

Since y^{90}/y^{10} is 20, according to the percentile ratio the fraction of the cross-country income dispersion explained by observables is $success_2$ =0.35.

I summarize the baseline experiment in Table 2.2. Clearly by both measures of success the dispersion of y_{KH} is much less than the dispersion of y, and this is the basic fact that motivates this study.[10]

var[log(y)]	1.246	y^{90}/y^{10}	20
var[log(y_{KH})]	0.501	y^{90}_{KH}/y^{10}_{KH}	7
$success_1$	0.40	$success_2$	0.35

Table 1: Baseline Success of the Factor-Only Model

Before proceeding, it is useful to check that these results are consistent with the slightly different data used in previous studies. Using the Hall and Jones data set $success_1$ is 0.40, as with ours, and $success_2$ is 0.34 (vs. 0.35 with ours). As is evident, the different decade, country coverage, and methodology in assembling the PWT does not lead to important changes in this basic finding.

2.3 Alternative Measures Used in the Literature

$success_1$ essentially asks what would the dispersion of (log) per-capita income be if all countries had the same level of efficiency, A, and then compares this counter-factual dispersion to the observed one. Klenow and Rodriguez-Clare (1997) propose the alternative measure:

$$success_{KR} = \frac{\text{var}[\log(y_{KH})] + \text{cov}[\log(A), \log(y_{KH})]}{\text{var}[\log(y)]},$$

which differs from $success_1$ for the covariance term in the numerator. In terms of equation (5) $success_{KR}$ is equivalent to a variance decomposition in which the contribution from the covariance term is split evenly between A and y_{KH}. Because in the data cov[log(A), log(y_{KH})] is positive (0.26) the Klenow and Rodriguez-Clare measure assigns a greater role to k and h than the simple ratio of variances: $success_{KR}$ is 0.61. Here I do not emphasize this measure because it does not answer the question: what would the dispersion of incomes be if all countries had the same A? As Klenow and Rodriguez-Clare explain, it asks the different question: "when we see 1% higher y, how much higher is our conditional expectation of y_{KH}?" which in my opinion is not as intuitive.

[10] Of course variation in y_{KH} – even though much less than variation in y – is economically significant and interesting in its own right. For recent studies shedding light on the sources of variation in k and h see, e.g., Bils and Klenow (2000), Hsieh and Klenow (2003), and Gourinchas and Jeanne (2003).

Klenow and Rodriguez-Clare (1997) and Hall and Jones (1999) also work with a different version of the expression for per-capita income, because they rewrite (3) as

$$y = \left(\frac{k}{y}\right)^{\frac{a}{1-\alpha}} hA^{\frac{1}{1-\alpha}},$$

i.e. in terms of the capital-output ratio instead of the capital-labor ratio. In other words their counterfactual income estimates based on factor-differences is $\tilde{y}_{KH} = \left(\frac{k}{y}\right)^{\frac{a}{1-\alpha}} h$ (instead of $y_{KH} = k^{\alpha}h^{1-\alpha}$). I find y_{KH} more intuitive and cleaner, as $\tilde{y}_{KH}$ is not invariant to differences in A (since A affects y), and is therefore less appropriate for answering the question: "what would the income distribution look like if all countries had the same A?". Indeed, it is easy to see that $\tilde{y}_{KH} = y_{KH}A^{-\alpha}$. Whether var[log($\tilde{y}_{KH}$)] is greater or less than var[log(y_{KH})] depends on the relative magnitudes of (appropriately weighted) var[log(A)] and cov[log(y_{KH}),log(A)], with log($\tilde{y}_{KH}$) getting less credit the (relatively) larger is the covariance. Intuitively, when A and y_{KH} covary a lot, if the latter is very small the former is also very small, so that $\tilde{y}_{KH}$ does not vary as much. In practice this is indeed what happens: when using $\tilde{y}_{KH}$ the factors only model looks even more unsuccessful than when using y_{KH}: $success_1$ is as low as 0.26, and $success_2$ is 0.21. Notice that relative to Klenow and Rodriguez-Clare (1997) we have made two methodological changes whose effects go in opposite directions: omitting the covariance term from $success_1$ lowers the explanatory power of factors, while writing y in terms of the capital-labor ratio increases it. This is why we end up with results that are in the same ball park.

It is worth noting that Hall and Jones' production function, equation (2), is substantially more restrictive than the one used by some of the other authors in the literature. In particular Mankiw, Romer, and Weil (1992) and Klenow and Rodriguez-Clare (1997) work with $Y = K^{\alpha}H^{\beta}L^{1-\alpha-\beta}$. Equation (2) is the special case where $\beta = 1-\alpha$. The great advantage of the Hall and Jones' formulation is that it generates the log-linear relation between wages and years of schooling that we exploited to calibrate h.[11] Since wage data do seem to call for log-linear wage-education profiles, Hall and Jones' restriction may be justified.

2.4 Sub-samples

It may be interesting to take a look at the values that the success measures take in sub-sample of countries. This is done in Table 2, where I report $success_1$ – as well as its two component parts – for the sub-samples of countries below and above the median per worker income; in and out of the OECD; and for the various continents. I also for convenience repeat the

[11] With the Mankiw, Romer, and Weil formulation the wage of a worker with s years of schooling is $w(s) = w_L + w_H h(s)$, where w_L is the wage paid to "raw" labor and w_H is the wage per unit of human capital.

full-sample values. I do not report $success_2$ because the small sample sizes make this variable hard to interpret.

Sub-sample	Obs.	var[log(y)]	var[log(y_{KH})]	$success_1$
Above the median	47	0.174	0.112	0.643
Below the median	46	0.586	0.257	0.438
OECD	24	0.077	0.047	0.606
Non-OECD	69	1.001	0.378	0.376
Africa	26	0.843	0.271	0.322
Americas	25	0.341	0.175	0.513
Asia and Oceania	25	0.652	0.292	0.448
Europe	17	0.128	0.033	0.255
All	93	1.246	0.501	0.400

Table 2: Success in sub-samples

Obviously the variation in log income per worker is smaller the smaller and more homogenous the sub-samples. Perhaps more interestingly, it is also smaller in sub-samples that tend to be richer on average (Above the median, OECD, Europe and Americas). It is indeed remarkable that, within the four continental groupings, the greatest variation in living standards is observed in Africa, a continent that is often depicted as flattened out by unmitigated and universal blight.

The success of the factor-only model is higher in the above the median and in the OECD samples than in the below the median and non-OECD samples, respectively. Hence, it is easier to explain income differences among the rich than among the poor. Furthermore, as indicated by comparison with the results for the full sample, it is easier to explain income differences among the rich than between the rich and the poor – while it is roughly as easy to explain within-poor differences as rich-poor differences. At the continental level, success is highest in the Americas, with roughly 50% of the log income variance explained, and lowest in Europe, with 25%. The latter result is entirely driven by the inclusion of the lone eastern European country (Romania), whose very high level of human capital makes it difficult to explain its very low income [Caselli and Tenreyro (2004) generalize this finding to a broader sample]. When Romania is excluded the success of the factor-only model for Europe is virtually perfect. In sum, the factor-only model works the worst where we need it most: i.e. when poor countries are involved.[12]

[12]Results for Asia are virtually unchanged when I do not aggregate it with Oceania.

3 Robustness: Basic Stuff

The rest of this chapter is essentially about the robustness of the findings reported in Table 2.2. In this section I start out with a set of relatively straightforward and somewhat plodding robustness checks. In particular, I look at some of the parameters of the basic model as well as at some issues of measurement error. Subsequent subsections deviate from the benchmark increasingly aggressively.

3.1 Depreciation Rate

The effect of varying the depreciation rate in the perpetual-inventory calculation is to change the relative weight of old and new investment. A higher rate of depreciation will increase the relative capital stock of countries that have experienced high investment rates towards the end of the sample period. Poorer countries have in general experienced a larger increase in investment rates over the sample period, but the relative gain is very small, so it is unlikely that higher or lower depreciation rates will have a considerable impact on our calculation.[13] In Figure 1 I compute and plot $success_1$ and $success_2$ for different values of δ. Clearly, the sensitivity of the factor-only model to changes in δ is minimal.

3.2 Initial Capital Stock

The capital stocks in our calculations depend on the time series of investment (observable) and on assumptions on the initial capital stock, K_0, which is unobservable. Does the initial condition for the capital stock matter? One way to approach this question is to compute the statistic

$$\frac{(1-\delta)^t K_0}{(1-\delta)^t K_0 + \sum_{i=0}^{t}(1-\delta)^i I_{t-i}}, \tag{6}$$

i.e. the surviving portion of the guessed initial capital stock as a fraction of the final estimate of the capital stock. For $t = 1996$ the average across countries of this statistic is .01, with a maximum of .08 (Congo). This is *prima facie* evidence that the initial guess has very small "persistence". However, this statistic is considerably negatively correlated with per capita income in 1996 (correlation coefficient -.24), indicating that our estimate of the capital stock is more sensitive to the initial guess in the poorer countries in the sample. This may be troublesome because if we systematically overestimated the initial (and hence the final) capital stocks in poor countries, we will bias downward the measured success of the factor-only model. Furthermore, it is not implausible that our guess of the initial capital stock

[13] We computed the average investment rate in the sub-periods 1969-1972 and 1993-1996. Then we subtracted these two averages, and correlated the resulting change in investment with real GDP per worker in 1996. The result is a modest -.01.

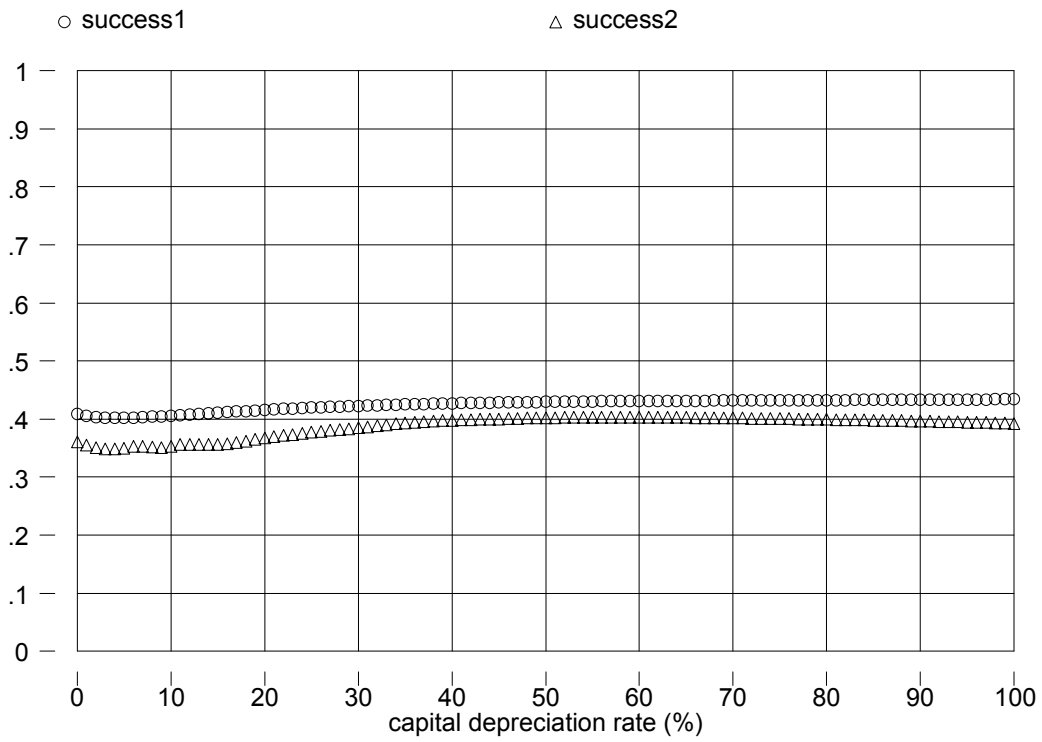

Figure 1: Depreciation Rate and Success

will be too high for poor countries. While rich countries may have roughly satisfied the steady state condition that motivates the assumption $K_0 = I_0/(g + \delta)$, most of the poorer countries almost certainly did not. Indeed it is quite plausible that their investment rates were systematically lower before than after date 0 (i.e. before investment data became available for these countries).[14]

A first check on this problem is to focus on a narrower sample with longer investment series. If we focus only on the 50 countries with complete investment data starting in 1950, we should be fairly confident that the initial guess plays little role in the value of the final capital stock. In this smaller, and probably more reliable, sample we get $success_1 = 0.4$, and $success_2 = .54$. Hence, the ratio of log-variances is unchanged relative to the full 93-country sample, but the inter-percentile ratio shows a considerable improvement. Clearly, though, as the sample size declines the inter-percentile ratio becomes less compelling as a measure of

[14]Circumstantial evidence that this may be the case is that a regression of the growth rate of total investment between 1950 and 1960 on per-capita income in 1950 yields a statistically significant negative coefficient.

dispersion, so on balance these results – though inconclusive – are reasonably reassuring.[15]

Another strategy is to attempt to set an upper bound on the measures of success, by making extreme assumptions on the degree to which the capital stock in poor countries is mismeasured. One such calculation assumes persistent growth rates in investment, I (as opposed to persistent investment levels). For example, we can construct a counter-factual investment series from 1940 to 1950 by assuming that the growth rate of investment in this period was the same as in the period 1950-1960. For countries with investment data starting after 1950 we can use the growth rate of investment in the first ten years of available data, and project back all the way to 1940. We can then use the perpetual inventory model on these data [always with $K_0 = I_0/(g+\delta)$], and measure success. On the full sample this yields $success_1 = .41$ and $success_2 = .35$, and on the sub-sample with complete I data starting in 1950 it yields $success_1 = 0.4$ and $success_2 = 0.54$, i.e. no change.[16]

Another experiment is to estimate the initial capital stock by assuming that the factor-only model adequately explained the data at time 0. Suppose that we trusted the estimate $K_0 = I_0/(g+\delta)$ for the United States (where date 0 is 1950), and consequently for all other dates. Then for any other country we could estimate K_0 by solving the expression

$$\frac{Y_0}{Y_{US}} = \left(\frac{K_0}{K_{US}}\right)^{\alpha} \left(\frac{L_0 h_0}{L_{US} h_{US}}\right)^{1-\alpha},$$

where 0 is now the first year for which this country's and the US' data on investment, GDP, and human capital is available. Note that everything is observable in this equation except K_0 (tough this does require us to construct new estimates of the human-capital stock for years prior to 1996). Clearly this procedure implies enormous variance in K_0, and this variance should persist to 1996, giving the factor-only model a real good shot at explaining the data. On the full sample this yields $success_1 = 0.43$ and $success_2 = 0.36$, and on the sub-sample with complete I data starting in 1950 it yields $success_1 = 0.45$ and $success_2 = 0.54$. Hence, even when the initial capital stock is constructed in such a way that the factor-only model fully explained the data at time 0, the model falls far short in 1996. I conclude from this set of exercises that improving the initial capital stock estimates is not likely to lead to major revisions to the baseline result.

[15] In this 50-country sample the variance of log income is 1.04, and the inter-percentile ratio is 7.6, i.e. according to both measures there is less dispersion than in the full sample (1.2 and 20), but much more so for the inter-percentile differential.

[16] When this method is used only to fill-in data between 1950 and 1961 it yields $success_1 = 0.40$ and $success_2 = 0.35$.

3.3 Education-Wage Profile

By assuming decreasing aggregate returns to years of schooling the Hall and Jones method dampens the variation across countries in human capital, thereby potentially increasing the role of differences in technology. More generally, our measure of human capital may obviously be quite sensitive to the parameters of the function $\phi(s)$.

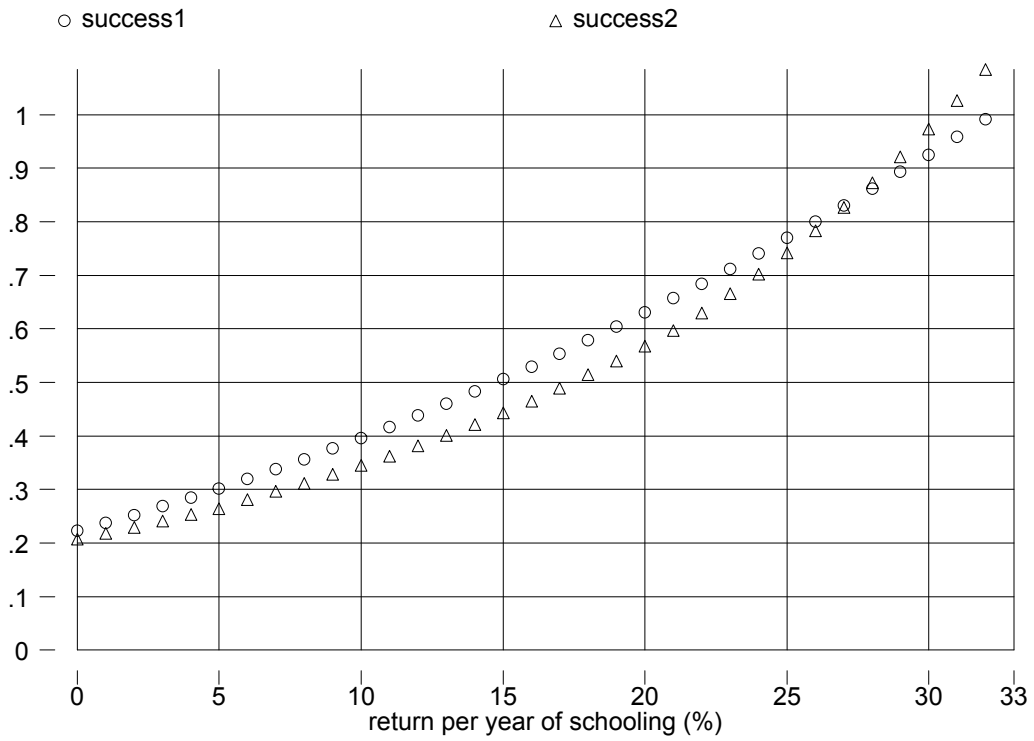

Figure 2: Returns to schooling and Success

One way of checking this is to assume a constant rate of return, or $\phi(s) = \phi_s s$, and experiment with various values of the (constant) return to schooling ϕ_s. Since countries with higher per-capita income have higher average years of education, the factor-only model will be the more successful the steeper is the education–wage profile. Figure 2 confirms this by plotting $success_1$ and $success_2$ as functions of ϕ_s.

While higher assumed returns to an extra year of education do lead to greater explanatory power for the factor-only model, only returns that are implausibly large lead to substantial successes. For $success_1$ ($success_2$) to be 0.75 the return to one year of schooling would have to be around 24% (25%). As already mentioned, in the Psacharopulos (1994)

survey the average return is about 10%. The highest estimated return is 28.8% (for Jamaica in 1989), but this is a clear outlier since the second-highest is 20.1% (Ivory Coast, 1986). These tend to be OLS estimates. Instrumental variable estimates on US data are 17-20 percent at the very highest [Card (1999)] - sufficient for our measures of success to just clear the 50 percent threshold. But the IV estimates tend to be lower in developing countries. For example, Duflo (2001) finds instrumental-variable estimates of the return to schooling in Indonesia in a range between 6.8% to 10.6% – and roughly similar to the OLS estimates. It seems, then, that independently of the return to schooling, the variation in schooling years across countries is too limited to explain very large a fraction of the cross-country variation in incomes.[17]

3.4 Years of Education 1

De la Fuente and Doménech (2002) survey data and methodological issues that arise in the construction of international educational attainment data, such as the average years of education in the Barro and Lee data set. Their conclusion – perhaps not surprisingly – is that such series are rather noisy, and that this explains in part why human-capital based models often perform rather poorly. For several OECD countries they also construct new estimates that take into account more comprehensive information than is usually exploited, and find that for this restricted sample their measure substantially improves the empirical explanatory power of human capital.

To see if incorrect measurement of s is a likely culprit for the lack of success of the factor-only model, I compute our success statistics for the sub-sample covered in the De la Fuente and Doménech (2002) data set, first with our baseline data, and then with the new figures provided by these authors (in their Appendix 1, table A.4) for 1995. This data is available for only 16 of our 93 countries. In this 16-country sample, with our baseline (Barro and Lee) schooling data, I obtain $success_1 = 0.44$ and $succeess_2 = 0.83$. In the same sample with the De la Fuente and Doménech data $success_1 = 0.47$ and $success_2 = 0.86$. The differences seem small.

This result is not particularly surprising because De la Fuente and Doménech (2002) show that the discrepancies between their measures and the ones in the literature are (i) stronger in first differences than in levels; and (2) stronger at the beginning of the sample than at the end. Indeed, for the 16-country sample in 1995 the correlation between the De la Fuente and Doménech (2002), and the Barro and Lee (2001) data I use in the rest of the paper is 0.73. Incorrect measurement of s is not the reason why the factor-only model performs poorly.

[17] This discussion, of course, assumes away human-capital externalities. I return to this in Section 4.5.

3.5 Years of Education 2

So far we have used Barro and Lee's data on years of schooling in the population over 25 years old. This may be appropriate for rich countries with a large share of college graduates. But it is much less appropriate for the typical country in our sample. Barro and Lee (2001) also report data on years of schooling for the population over 15 years of age. These data can be combined with the data on the over-25 as follows.

First, note that we can write

$$s_{15} = \frac{1}{N_{15}} \left[(N_{15} - N_{25}) s_{<25} + N_{25} s_{25} \right],$$

where s_{15} and s_{25} are the average years of education in the population over 15 and over 25 years of age, respectively (the data), and $s_{<25}$ is the (unknown) average years of education in the population between 15 and 25 years of age. N_{15} and N_{25} are the sizes of the population over 15 and over 25. With data on N_{15} and N_{25}, then, this is one equation in one unknown that can easily be solved for $s_{<25}$. With an estimate of $s_{<25}$ at hand, one could then produce a new measure of s as

$$s = L_{<25} s_{<25} + L_{25} s_{25},$$

where $L_{<25}$ and L_{25} are estimates of the proportions of above and below 25 year olds in the economically active populations.

I take data for N_{15}, N_{25}, $L_{<25}$, and L_{25} from LABORSTA, the data base of the International Labor Organization (ILO).[18] This alternative measure of s can be constructed for 90 countries. With it, our success measures are 0.391 and 0.345, respectively. Hence, this potentially improved measure of human capital *worsens* its explanatory power. The reason is not hard to see: poor countries experienced much faster growth in schooling than rich countries. This means, in particular, that the education gap is much smaller for the cohort less than 25 years of age. Hence, bringing the education of this cohort into the picture reduces the cross-country variation in human capital.

3.6 Hours Worked

So far I have measured L as the economically active population, a measure that basically coincides with the labor force. Of course, the number of hours worked – the concept of labor input we would ideally like to use in our calculations – may be far from proportional to this measure, both because of cross-country differences in unemployment, and because the

[18] For both the population at large and the economically active subset the data is available at 10-year intervals from 1950 to 2000, with the lucky exception that there is also an observation for 1995, which I use. The population is broken down in 5-year age intervals, so it's a no brainer to aggregate up to the numbers we need.

average employed worker may supply different amount of hours in different countries, for example because of different wages or different opportunity cost of work (in terms of forgone leisure).

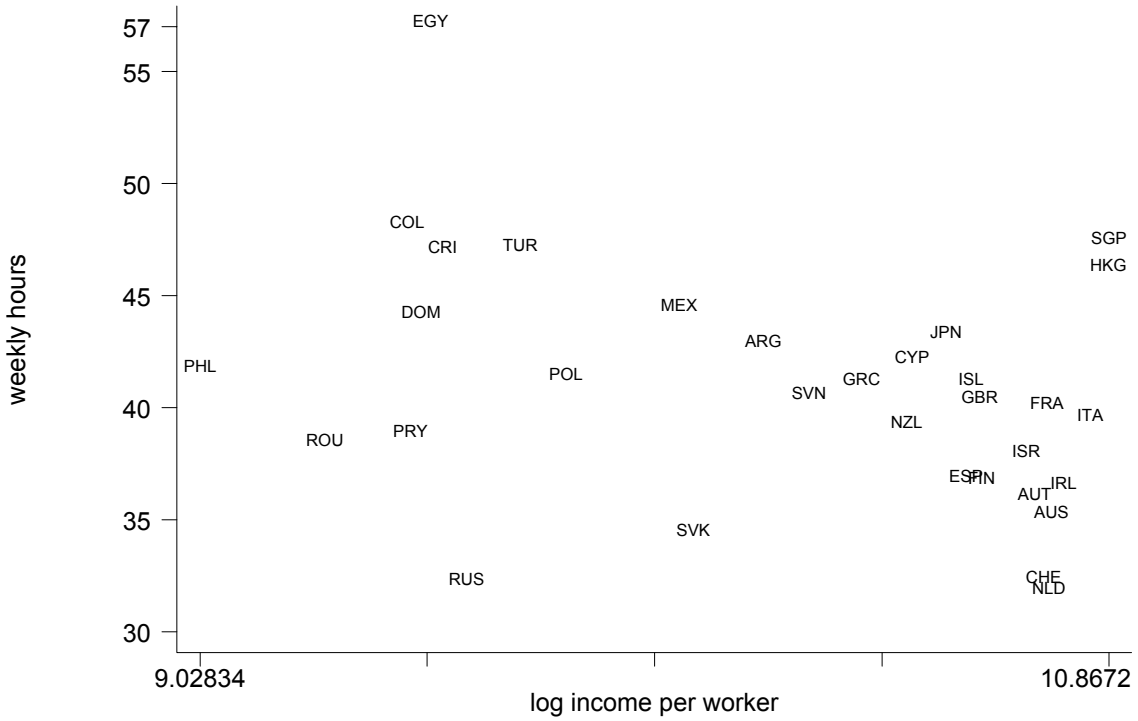

Figure 3: Hours Worked around the World

LABORSTA includes data on weekly hours for 41 countries in 1996, and in this sample the variance of hours worked is indeed very large: from 57 (Egypt), to 26.6 (Moldova).[19] However, the particular cross-country pattern of these hours does not go in the direction that favors the factor-only model. Figure 3, where I plot weekly hours against log income per worker (for countries with data on both variables), clearly shows that workers work fewer hours in high-income countries.[20] This implies that – if anything – TFP differences are under-estimated!

In principle, this effect may be compensated – and possibly reversed – by higher

[19]For the subset of 28 countries that are both in the ILO sample and in our baseline sample the maximum is the same, and the minimum is 31.7 (Netherlands).

[20]The coefficient of a regression of log weekly hours on log per-worker income implies that a one percent increase in per worker income lowers weekly hours by about 0.1% – which is sizable.

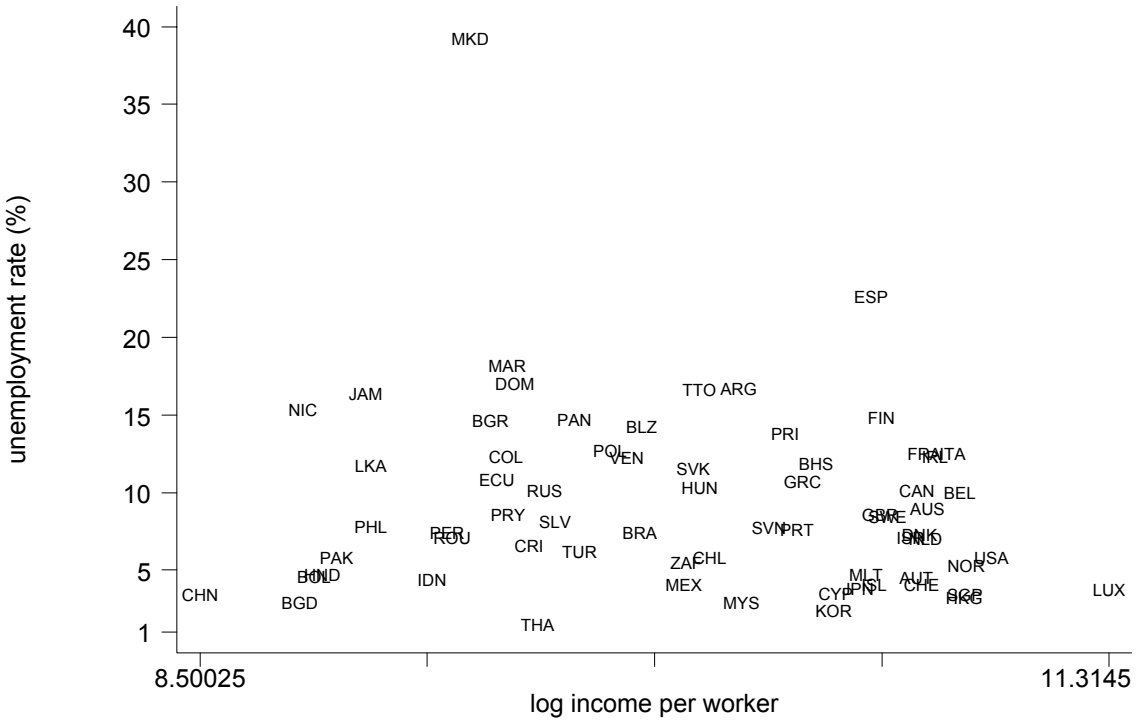

Figure 4: Unemployment Rates Around the World

unemployment rates in poorer countries. Figure 4 plots data from the World Bank's World Development Indicators (WDI) on unemployment rates against log per capita income, in 1996.[21] Contrary to common perceptions, unemployment rates are not higher in poorer countries. It therefore seems unlikely that further pursuing differences in hours worked may lead to a significant improvement in the explanatory power of the factor-only model.

3.7 Capital Share

The exponents on k and h act as weights: the larger the exponent on, say, k, the larger the impact that variation in k will have on the observed variation in y. However, under constant returns to scale these exponents sum to one, so increasing the explanatory power of k through increases in α also means lowering the explanatory power of h. Because k is more variable across countries than h, in general one can increase the explanatory power of the "factor-only" model by increasing α.

[21] ILO data on unemployment generate a very similar picture.

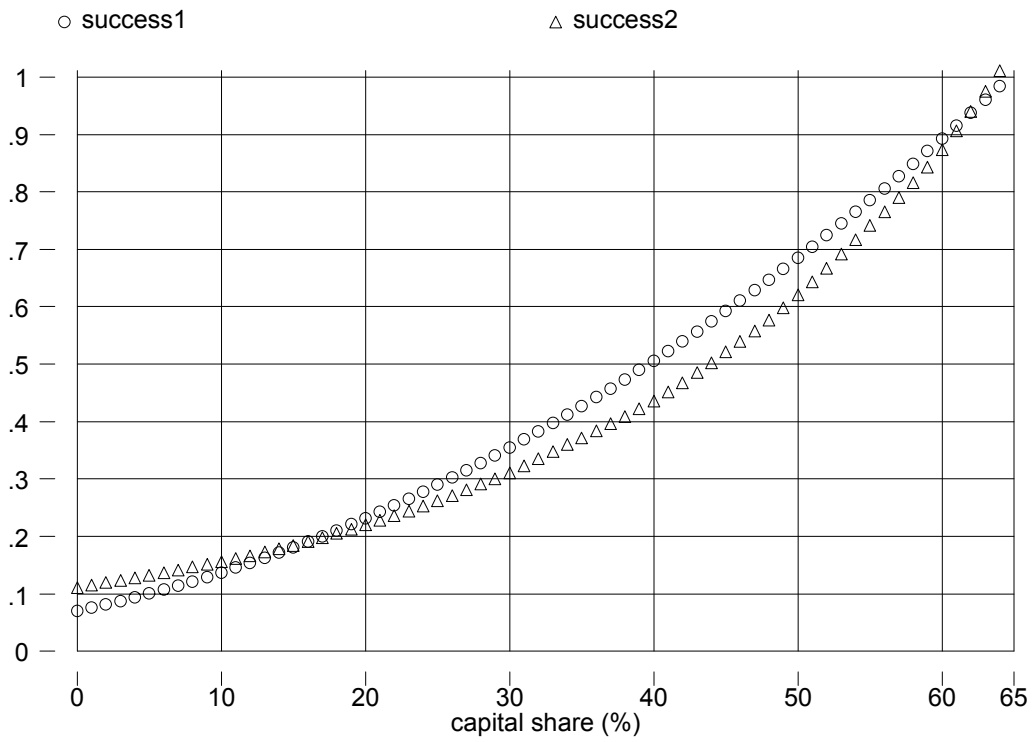

Figure 5: Capital Share and Success

Figure 5 plots $success_1$ and $success_2$ as functions of the capital share α. As predicted, the fit of the factor-only model increases with the assumed value of α. Remarkably, our measures of success are quite sensitive to variations in α. For example, a relatively minor increase of α to 40% is sufficient to bring $success_1$ to 0.5, and a 50% capital share implies success measures in the 0.6 - 0.7 range. Success is almost complete with a 60% capital share. This high sensitivity of the success measure, especially around the benchmark value of $\alpha = 1/3$, imply that the parameter α is a "sensitive choice" in development accounting, and that our assessment of the quantitative extent of our ignorance may change non-trivially with more precise measures of the capital share. Still, as long as the capital share is below 40%, most of the variation in income is still explained by TFP.

4 Quality of Human Capital

We have seen that simple parametric deviations from the benchmark measurements in Section (2) do not alter the basic conclusion that differences in the efficiency with which factors are

used are extremely large. Here and in the next section I subject this claim to further scrutiny, by investigating possible differences in the *quality* of the human and physical capital stocks. For, the measures adopted thus far are exclusively based on the quantity of education and the quantity of investment, but do not allow, for example, one year of education in country A to generate more human capital than in country B. Similarly, they do not allow one dollar of investment in country A to purchase capital of higher quality than in country B.

I conceptualize differences in the quality of human capital by writing

$$h = A_h e^{\phi(s)}.$$

Up until now, I have assumed that A_h is constant across countries. In this section I examine the possibility that A_h is variable.

4.1 Quality of Schooling: Inputs

Klenow and Rodriguez-Clare (1997) and Bils and Klenow (2000) have proposed ways to allow the quality of education to differ across countries. Their main focus is that the human capital of one generation (the "students") may depend on the human capital of the preceding one (the "teachers"). One can further extend their framework to allow for differences in teacher-pupil ratios, and other resources invested in education. For example, one could write:

$$A_h = p^{\phi_p} m^{\phi_m} k_h^{\phi_k} h_t^{\phi_h}, \tag{7}$$

where p is the teacher-pupil ratio, m is the amount of teaching materials per student (textbooks, etc.), k_h is the amount of structures per student (classrooms, gyms, labs, ...), and h_t is the human capital of teachers: the better the teachers, the more students will get out of their years of schooling. More generally, the term h_t might capture externalities in the process of acquiring human capital.

In this sub-section I will try to plug in values for the inputs p, m, k_h, and h_t, and calibrate the corresponding elasticities. Unfortunately, little is known about the latter. Indeed, they are the object of intense controversy in and out of academe. Hence, I will typically look at a fairly broad range of values.

4.1.1 Teachers' Human Capital

I begin by focusing on the last of the factors in (7), h_t. To isolate this particular channel for differences in schooling quality I ignore other sources, i.e. I set $\phi_p = \phi_m = \phi_k = 0$, which is essentially Bils and Klenow's assumption. When we review the evidence on these other ϕs, we'll see that this assumption may actually be quite realistic. If we make the additional

"steady state" assumption that $h_t = h$, we can write

$$h = e^{\phi(s)/(1-\phi_h)},$$

and plugging this into (3) we get:

$$y = Ak^{\alpha}e^{\frac{(1-\alpha)\phi(s)}{1-\phi_h}}. \tag{8}$$

Note that this formulation magnifies the impact of differences in years of schooling, the more so the larger the elasticity of student human capital to teacher's human capital.

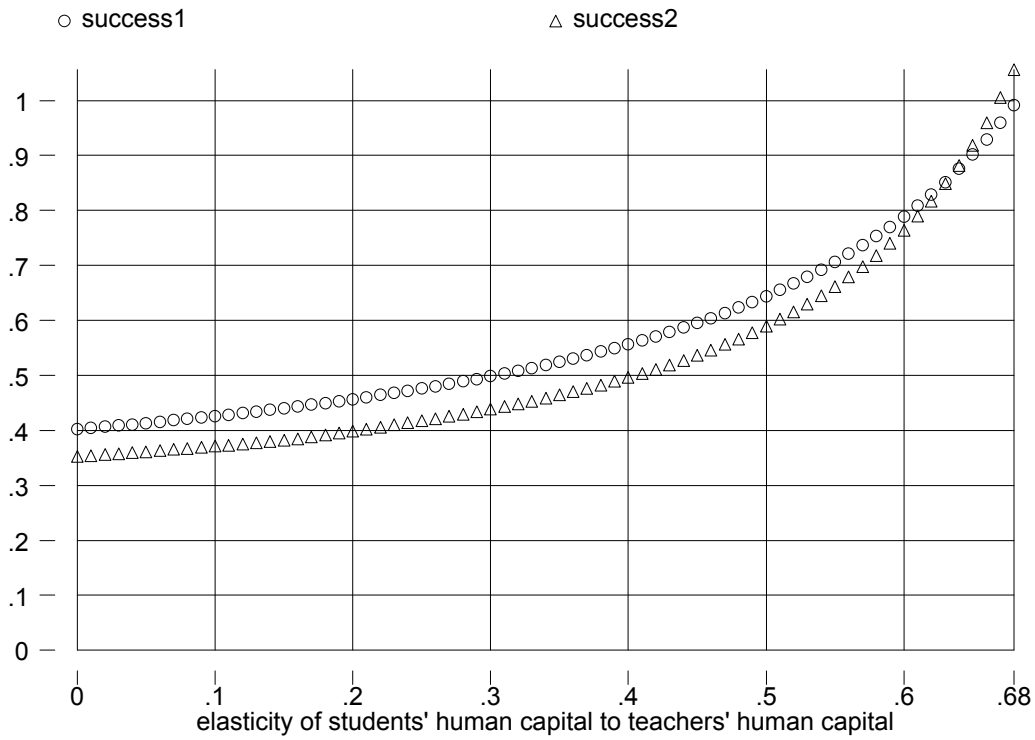

Figure 6: ϕ_h and Success

I continue to choose $\alpha = 1/3$, and the function $\phi(\cdot)$ as described in Section 2. The new, unknown parameter is ϕ_h. In Figure 6 I plot $success_1$ and $success_2$ as functions of this parameter. Note that $\phi_h = 0$ is the baseline case of section 2. At the low values of ϕ_h implied by the baseline case the success measures are fairly insensitive to changes in the elasticity of students' to teachers' human capital. However, the relationship between the success measures and ϕ_h is sufficiently convex that when ϕ_h is 67% success is complete. Coincidentally, 67% is the upper bound of the range of values Bils and Klenow consider "admissible" for ϕ_h, though

clearly this admissibility is purely theoretical: their preferred values are actually in the 0-20% range.[22]

One way to think about what is reasonable for ϕ_h is to compute by how much the teachers human capital effect "blows up" the Mincerian return: from equation (8) we see that with $\phi_h = 0.2$ the "social" return to schooling is 1.2 times the private one; with $\phi_h = 0.4$ it is 1.7 times larger; and with $\phi_h = 0.67$ it is 3 times more. While it is hard to reach a firm conclusion, it would seem that reasonable priors on ϕ_h are inconsistent with large improvements in the fit of the factor-only model.

Turning to possible objective estimates of ϕ_h, the first option is of course to look for estimates of the effect of teachers' years of education on student achievement. This is because under our assumptions differences in teacher's quality are ultimately determined by teachers' years of education. However, Hanushek's (2004) review of the literature concludes that teachers' measurable credentials – including years of education – have no measurable impact on schooling outcomes.[23]

Another way to formulate priors on the possible magnitude of ϕ_h is to look at evidence on the effect of parental education on wages. After all, our simple representative-agent model of human capital is not explicit about the particular way the economy's average level of human capital enhances the learning experience of new members of society. We can legitimately re-interpret h_t, therefore, as the human capital of parents. One recent set of log-wage regressions including the schooling of parents (alongside with an individual's own schooling) is presented in Altonji and Dunn (1996). Depending on data sources, and on whether the regression is estimated for men or women, their coefficient on father's years of schooling ranges from -.5% to 1%, and the coefficient on mother's schooling from less than .1% to about .5%. Note that given our functional form assumption the coefficient of parental education is $\phi_s\phi_h$, where ϕ_s

[22]They compute this upper-bound (roughly) as follows. Given data on schooling years of different cohorts, given a Mincerian wage-years of schooling profile, and given a value for ϕ_h, it is possible to estimate the growth rate of h, and hence the contribution of growth in h to the growth of y. Holding the Mincerian profile constant, the larger ϕ_h, the larger the fraction of growth explained by human capital (for reasons already touched upon in the text). For Bils and Klenow the upper bound for ϕ_h is the value such that growth in human capital explains *all*

of growth - or the value beyond which the residual, growth in TFP, would have to be negative. When the Mincerian profile features decreasing returns, as in our baseline specification, and as in Bils and Klenow's preferred specification, this maximum value for ϕ_h is 0.19; when the Mincerian profile is linear the maximum becomes 0.67. The decreasing returns case allows for a smaller maximum ϕ_h because, towards the beginning of the sample period, many countries with very low education levels have very high Mincerian returns, implying fast growth in human capital.

[23]This does not mean that teachers' quality does not matter, of course. It only means that teacher quality is not related to measurable credentials. This unmeasurable quality effect remains (appropriately) a part of the measure of our ignorance.

is the return to own years of schooling (assumed constant for simplicity). If the return to own schooling, ϕ_s, is in the ball park of 0.10 (as the evidence on Mincerian coefficients roughly implies), and we focus on Altonji and Dunn's upper bound of 0.01 for $\phi_s\phi_h$, we conclude that ϕ_h cannot be more than 0.1. A quick check with Figure 6 reveals that even this upper bound does not support a meaningful boost in the explanatory power of overall human capital.[24]

4.1.2 Pupil-Teacher Ratios

The term $h_t^{\phi_h}$ in equation (7) does not appear to enhance the success of the factor-only model. I now consider the term p^{ϕ_p}. Lee and Barro (2001) report data on the pupil-teacher ratio in a cross-section of countries for various periods since 1960, and separately for primary and secondary schooling. For each country, I focus on the pupil-teacher ratio in the years when the average worker attended school. To pinpoint this year, I need to start with an estimate of the age of the average worker, which I construct from LABORSTA.[25] Then I assume that children begin primary schooling at the age of 6. This implies that the relevant observation for the primary pupil-teacher ratio would be for the year 1996-age+6. Furthermore, using unpublished panel data by Barro and Lee on the duration of primary and secondary schooling, we can determine the relevant observation for the secondary pupil-teacher ratio as 1996-age+6+duration of primary school.

In order to combine the primary and secondary ratios in a unique statistic, I combine the duration of schooling data with our basic data on the average years of schooling of the population over 25 years of age, s, to determine what fraction of schooling time the average worker spent in primary, and what fraction in secondary school. I then construct p by simply averaging the primary and secondary teacher-pupil ratio using as weights the time spent in these two grades, respectively. At the end of all this, I have data on p for 86 of our 93 countries.[26]

[24] Another way to boost the contribution of human capital to income would be to assume that parental/teacher human capital increases the *slope* and not just the intercept of the log-wage – schooling relation. This is indeed Altonji and Dunn's main focus. However, they do not find much evidence in support of this hypothesis.

[25] As already mentioned, LABORSTA breaks down the economically active population in 5-year age intervals, from 10-14, to 60-64, plus a catch-all bracket for 65+. To get at the average age of a worker I simply weighted the middle year of each interval by the fraction of the labor force in that interval. For the 65+ group, I arbitrarily used 68. Of my 93-country sample, this data is available for 91 countries. I imputed average age for the two missing countries (Taiwan and Zaire) through a cross-sectional regression of average age of worker on per-worker income and years of schooling.

[26] Since the pupil-teacher ratio is observed at five-year intervals in practice we "target" the observation closest to the estimated age at which the average worker went to school. With this procedure, in the sample of 86 countries with data on pupil-teacher ratios, the target dates for primary school attendance are 1960 for two countries, 1965 for 40, and 1970 for 44. For secondary school attendance the target dates are 1965 (one

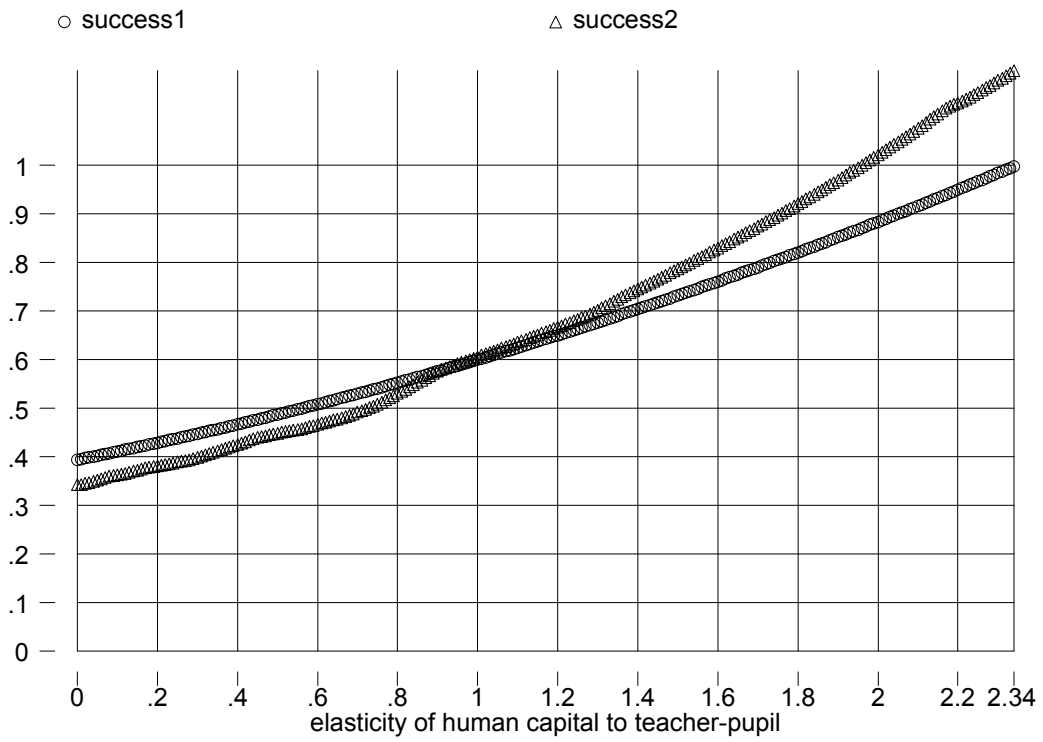

Figure 7: ϕ_p and Success

Figure 7 plots $success_1$ and $success_2$ as functions of ϕ_p. Since richer countries have higher teacher-pupil ratios, clearly a higher elasticity of human capital to this ratio implies a better fit, or greater success. What is a reasonable range of values for ϕ_p? At the low end of the spectrum there is the position taken by Hanushek and coauthors, who conclude that resources – including a large teacher-pupil ratio – have little if any effect on economic outcomes.[27] At the other end of the spectrum, my own reading of the literature indicates that the highest published estimate of ϕ_p is a very sizable 0.5.[28] However, even with this extremely high estimate it is clear that the fit of the model improves modestly, with our success measures barely attaining even the 50% mark.

country), 1970 (25), 1975 (55), and 1980 (5).

[27]In a cross-country context, Hanushek and Kimko (2000) find no evidence that more resources improve schooling quality, and Hanushek, Rivkin and Taylor (1996) and Hanushek (2003) reach the same conclusion upon reviewing the US-based literature.

[28]Card and Krueger (1996). I infer this number from their reported 5% increase in earnings associated with a 10% reduction in class size for white men.

4.1.3 Spending

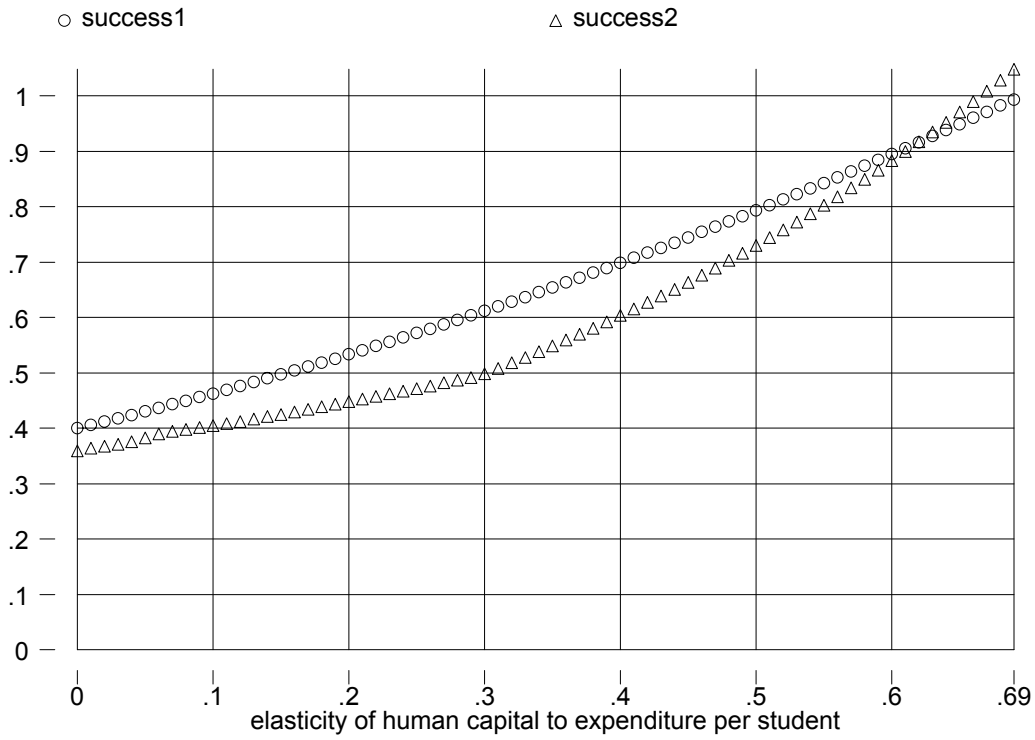

Figure 8: ϕ_{sp} and Success

I do not have direct data on materials, m and structures per student, k_h. Instead, I have – always from Lee and Barro (2001) – a measure of government spending per student in PPP dollars. The bulk of this spending typically goes to teacher salaries, so variation in these data also reflect differences in the number *and possibly the quality* of teachers per student. However, to a certain extent, they may also reflect variation in materials. For the purposes of using these data, it seems sensible, therefore, to replace equation (7) by $A_h = spending^{\phi_{sp}}$, where the dating of the spending observation and the weights given to primary and secondary spending are determined as for the pupil-teacher ratio. For this exercise, I have data for 63 countries, and for this sample the measures of success are plotted in Figure 8. Again, rich countries devote more resources to education per student, so the fit of the model improves with ϕ_{sp}. However, again, there is the Hanushek position in the papers cited above, according to which ϕ_{sp} should be thought of as close to zero. At the other end of the range I have found an estimate of 0.2, which clearly is barely sufficient to even clear

the 50% threshold of explanatory power.[29,30]

4.2 Quality of Schooling: Test Scores

Another way to investigate the potential of quality-of-education modifications to the basic model is to exploit information on the performance of students on reading, science, and math tests in different countries. When students in one country outperform students of another (holding grade constant), we can assume that they have enjoyed schooling of higher quality, whether this higher quality comes from higher teacher-pupil ratios, quality of teachers, other expenditures, or other unobservables specific to the production of human capital. Hanushek and Kimko (2000) find that test scores enter significantly in growth regressions.

To implement this idea I think of A_h as a function of test scores: higher test scores signal higher human capital. Suppose, for example, that the relationship between school quality and test results is given by $A_h = e^{\phi_\tau \tau}$, where τ is the test score.[31] Then, with data on test scores, if we knew ϕ_τ we could construct a new counterfactual measure of y_{KH}, or the output attributable to "observable" factors of production.

I use data on test scores provided by Lee and Barro (2001), who for several countries observe data on multiple tests (e.g.: math, science, and reading), and for multiple grades, at different dates. Ideally I would follow the procedure outlined in the previous sub-section, i.e. to "target" the year in which the average worker is presumed to have been in school. Because this data is very sparse, however, and mostly available in recent dates, I will focus on recent observations. This procedure is appropriate if the quality of education has grown over time at roughly similar rates across countries.

The two tests that afford the greatest country coverage – 28 countries with overlapping test, input, and output data– are a math and a science test imparted to 13 year old children between 1993 and 1998. The scores are standardized on a 0-to-100 scale, and I take the simple average of the two test scores.[32] With this summary measure of τ at hand, in Figure 9 I plot our measures of success against ϕ_τ.

The result should be treated with great caution given the very small sample size. Notice for example that, even for $\phi_\tau = 0$, both measures of success are considerably higher

[29] Johnson and Stafford (1973), who run a regression of log hourly wages on log state expenditure per student (and controls), obtaining a coefficient of 0.198. For the reasons discussed by Hanushek and co-authors there is a high presumption of upwad bias in this estimate.

[30] Lee and Barro (2001) also report information on the duration of the schooling year (in days and hours), but these variables – while highly variable – are weakly, and if anything negatively, correlatd with per-capita income, so that they are highly unpromising from the perspective of improving the fit of the model. Similarly, teacher salaries, as a percent of per-capita GDP, are higher in poorer countries.

[31] The reason for the exponential form will be apparent below.

[32] The correlations between the two test scores is 0.87.

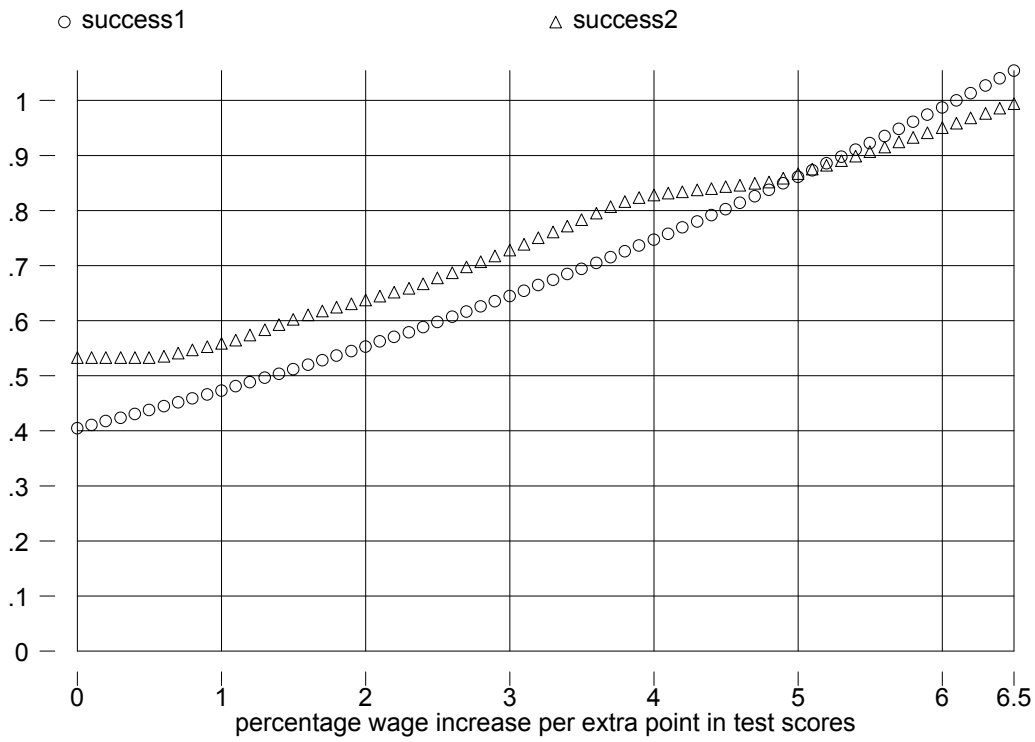

Figure 9: $\phi_\tau(\times 100)$ and Success.

than in the full sample. With that caveat, it is true that students in rich countries perform better in standardized tests, and therefore the success of the model improves with ϕ_τ.

To find a benchmark for ϕ_τ against which to evaluate Figure 9, notice that our assumption on the relationship between test scores and school quality translates into an assumption on the relation between test scores and wages: a unit increase in test scores is associated with a ϕ_τ proportional increase in wages. I have chosen this exponential form because studies of the relationship between test scores and wages tend to report coefficients from regressions of log wages on absolute test scores. For example, the coefficient $\phi_\tau \times 100$ (after rescaling the test data to be in the same units as ours) is reported to be between 0.08 and 0.34 by Murnane, Willet, and Levy (1995); between 0.12 and 0.27 by Currie and Thomas (1999), and between 0.55 and 1.02 by Neal and Johnson (1996) – which is at the high end of the range of available estimates.[33]

[33]Murnane, Willet, and Levy (1995) report the coefficient of the regression of log of weekly wages on math test scores (tested in senior high school) to vary between 0.00004 and 0.00017 (depending on sex and cohorts considered, US data). Since the test results are reported to vary between 2 and 17 points, we assume that

Inspection of Figure 9 given this range of values suggests that using test scores as proxies for schooling quality cannot substantially improve the performance of the factor-only model. The problem is that, given the drastically reduced sample size, it is hard to take a stand on the degree to which this finding generalizes.

I can attain a slight increase in sample size if I drop the requirement that the tests be imparted in roughly the same period and roughly the same subject. If I use all the test scores available from the 1990s, i.e. I average across all tests irrespective of subject, age group, and specific year, our sample size becomes 42 and success is given by figure 10. If I use all available tests, including those from decades before the 1990s, the sample size is 45 and success is shown by figure 11. As we increase the sample size, the potential success of the factor-only model if anything declines.

4.3 Experience

Klenow and Rodriguez-Clare (1997) and Bils and Klenow (2000) also allow for differences across countries in experience levels. Since Mincerian wage regressions indicate that experience increases earnings, it makes sense to correct human capital for the contribution of experience. This correction has two conflicting effects on the explanatory power of human capital. Since workers in rich countries live longer than workers in poor countries, this should boost rich countries' human capital. However, since rich-country workers spend more time in school, a smaller proportion of their time is spent accumulating experience, which reduces their relative human capital.[34]

Klenow and Rodriguez-Clare (1997) find that the net effect is negative: experience is

the test is on a 0–20 scale. When translated to our 0-100 scale this implies the ϕs reported in the text. The Currie and Thomas (1999) results imply that "students who score in the upper quartile of the reading exam earn 20% more than students who score in the lower quartile of the exam, while students in the top quartile of the math exam earn another 19% more. When they control for father's occupation, father's education, children, birth order, mother's age, and birth weight, the wage gap between the top and bottom quartile on the reading exam is 13% for men and 18% for women, and on the math exam it is 17% for men and 9% for women" (Krueger, 2002, p. 25). From here we can infer that ϕ_τ varies between 0.0012 and 0.0027 (dividing the percentage change in the wage by the 75 points that separate the top from the bottom quartile). Neal and Johnson (1996) run a regression of log real yearly wages on standardized AFQT test scores, and find a coefficient between 0.17 and 0.29. Introducing more controls the coefficients are between 0.12 and 0.16. Since the standard deviation of AFQT scores (as reported in the note to their Appendix A.3) is 36.65, this implies that a one-point increase in AFQT scores increases wages by between 0.33 and 0.79 percent. Given that AFQT scores range between 95 and 258, this implies a ϕ between 0.0055 and 0.0102 (treating each of the AFQT points as 1.64 of our 100 points). (Whether AFQT scores are measures of schooling outcome is somewhat controversial). Hanushek and Kimko (2000) use essentially the same international test scores we are using here to explain the earnings of migrants to the US, and obtain $\phi_\tau \times 100$ of approximately 0.2.

[34] A third effect, that adds to the second, is that rich-country workers may retire earlier.

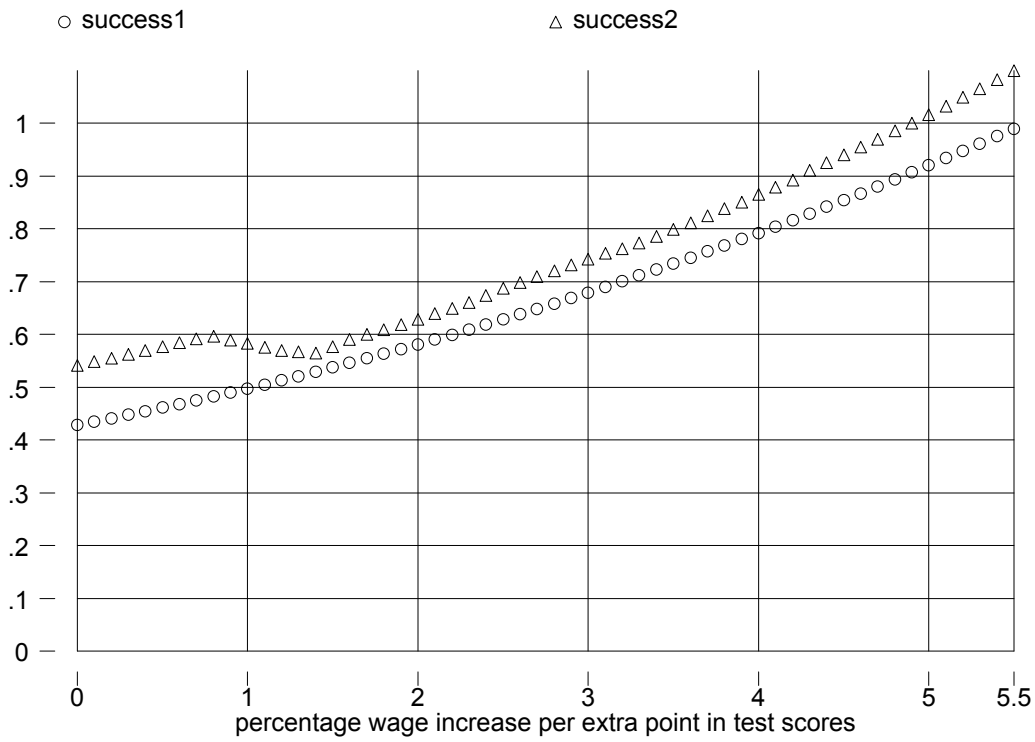

Figure 10: $\phi_\tau(\times 100)$ and Success, all tests in the 90s.

actually higher in poor countries. Hence, in their calculations correcting for experience lowers the explanatory power of the factor-only model. However, in order to compute the average age of workers they rely on UN data on the age structure of the population, while in principle it would be more accurate to look at the age structure of the labor force. Using again the LABORSTA-based measure of the average age of the economically active population in the formula

$$\text{experience=age-schooling-6,}$$

I find that the correlation between experience and per-capita income is -0.29 in our 93-country sample. Therefore, I confirm the Klenow and Rodriguez-Clare conclusion that poor countries have less education but more experience. Adding experience to the model, therefore, will only worsen its explanatory power.[35]

[35]This discussion assumes implicitly that experience enters linearly in the production function for human capital, an assumption we know not to be valid. However, for this cnsideration to overturn the conclusion we just reached, it would have to be the case that poor countries are to the right of the argmax, which seems very unlikely: in my data, the maximum average experience is 27 years. More importantly the discussion

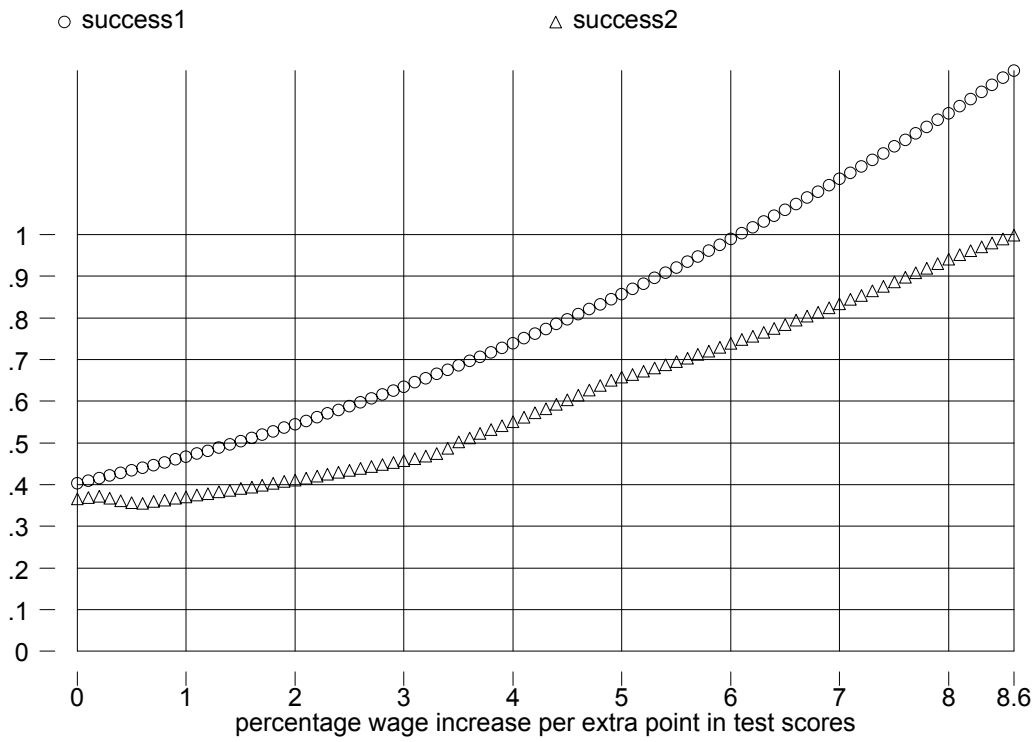

Figure 11: $\phi_\tau(\times 100)$ and Success, all available test scores

4.4 Health

Weil (2001) and Shastry and Weil (2003) point out that there are very large cross-country differences in nutrition and health status, and argue that these differences map into substantial differences in energy and capacity for effort. They find that accounting for health differences across countries increases by one-third the explanatory power of human capital for differences in per-capita income.

Weil (2001) uses as a proxy for health the Adult Mortality Rate (AMR), which measures the fraction of current 15 year old people who will die before age 60, under the assumption that age-specific death rates in the future will stay constant at current levels. In practice, this is a measure of the probability of dying "young," and is therefore a plausible (inverse) proxy for overall health status.

also abstracts from compositional issues. Feyrer (2002) uncovers an economically important and remarkably robust association between a country's productivity and its share of the labor force that is between 40 and 49 years of age. Extending the development accounting framework to capture this effect would be a worthwhile task.

The correction of human capital for health can be implemented through the assumption $A_h = e^{\phi_{amr} AMR}$, where clearly $\phi_{amr} < 0$: a higher adult mortality rate implies a less energetic workforce. I gather cross-country data on AMRs from the WDI, covering 91 of our 93 countries, for the year 1999. I plot success for different values of $-\phi_{amr}$ in Figure 12. Since richer countries have healthier workers, the explanatory power of human capital increases in $-\phi_{amr}$.

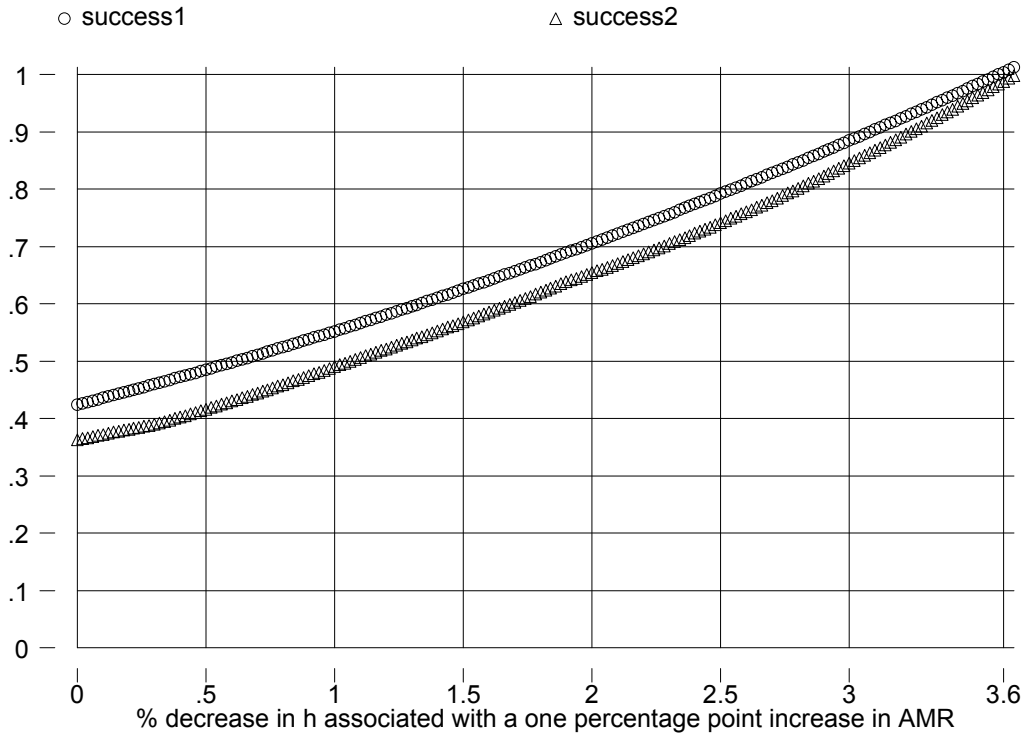

Figure 12: $-\phi_{amr}$ and Success

Weil's preferred value for $-\phi_{amr}(\times 100)$ is 1.68. Conditional on this value, I do confirm his finding that the factor-only model's explanatory power improves considerably - indeed by almost one third, taking us well above the 50 percent threshold of success. This is therefore a very important and promising contribution.

Given his choices of functional form, however, this calibration implies that a one-percentage-point reduction in the probability of dying young is associated with a 1.68 percent increase in human capital, *and hence in wages*. Put another way, reducing the probability of dying before the age of 60 (as of age 15) by 6 percentage points has the same impact on wages as one extra year of schooling. This effect may seem a bit too large to be realistic.

Given the somewhat tortuous – if ingenious – path through which Weil comes up with this calibration, I would tend to consider this number an upper bound.[36]

One can perhaps improve on Weil's exercise by exploiting as a proxy for health the information on average birthweight generated by Behrman and Rosenzweig (2004). The great advantage is that these authors also report estimates (based on within twin-pair regressions) of the economic returns to higher birthweight (as measured by wages). This should provide a more solid base for calibration. There are, however, several shortcoming. First, birthweight may be less strongly correlated with health than the adult mortality rate. Second, the cross-section of mean birthweights refers to new borns in 1989, so it captures (a correlate of) the health of a cohort of workers that was not even in the labor force (aside from the most extreme cases of child labor) as of the date of our development accounting exercise (1996). Hence, one needs to assume a very high degree of persistence in cross-country differences in birthweights in order to put a lot of stock in this exercise. Third, the point estimate of the returns to birthweight are from a sample of US female twins, and one may question their applicability to the population at large.[37]

With those caveats, Figure 13 plots the usual measures of success when we assume $A_h = e^{\phi_{bw} BW}$, where BW is Behrman and Rosenzweig's mean birthweight (in pounds), and ϕ_{bw} is the elasticity of human capital to birthweight. The number of countries is 82. Since birthweight is higher in rich countries a higher value of ϕ_{bw} increases the explanatory power of human capital. However, the value of ϕ_{bw} implied by Behrman and Rosenzweig's log wage regressions is 0.076, which implies a trivially small improvement in the success of the factor-only model. This is substantively the same conclusion of Behrman and Rosenzweig, who report that the variance of $\phi_{bw} BW$ is less than one percent of the variance of $\log(y)$.

In sum, while the results with the adult mortality rate strongly imply that a correction for differences in health status is a first-order requirement in the measurement of human capital, those using birthweight are much less supportive. In light of the shortcomings of both exercises, however, it seems highly worthwhile to try and explore the matter further with more accurate indicators of health and more precisely calibrated parameters.

[36] Weil uses published micro-level estimates from three developing countries to infer the elasticity of human capital to *height*. He then uses time series data from Korea and Sweden to estimate a relationship between height and the AMR. He then combines these two pieces of information to infer the elasticity of human capital to the AMR. In essence, he is using the AMR to predict height, and then applies to the predicted height the microeconomic estimate of the effect of height on wages.

[37] Fourth, Behrman and Rosenzweig also find some evidence of non-linearity in the birthweight-log wage relationship, so to perform a really accurate exercise one should use data on each country's entire birthweight distribution. Of course that is easily available for only a handful of countries.

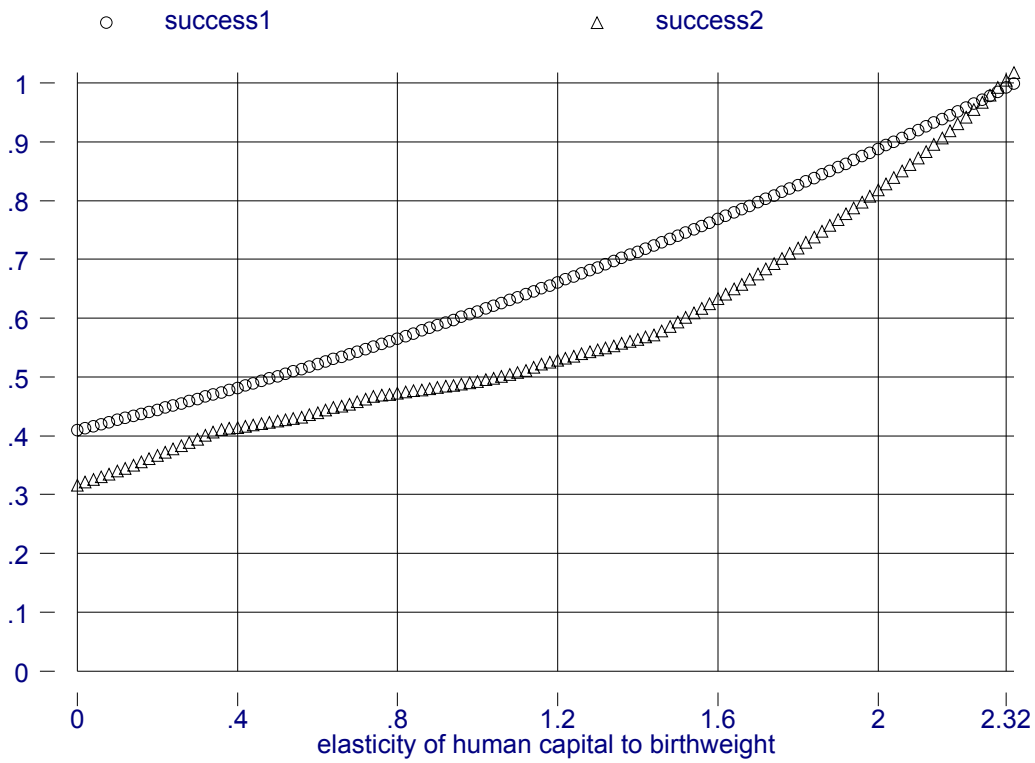

Figure 13: ϕ_{bw} and Success

4.5 Social vs. Private Returns to Schooling and Health

Some additional important caveats about the nature of the calculations above is in order before I "set aside" human capital. Recall that the function $\phi(s)$ that we have used to map years of schooling into human capital was calibrated on estimates of *private* rates of return. Similarly, attempts at calibrating the health-human capital relation rely on observed private returns to health. But, as pointed out by various authors, and especially forcefully by Pritchett (2003), these private returns may bear little relationship to the social (or aggregate) return to education, which is of course what one would like to plug in our calculations.

As Pritchett points out, the social return to education may be higher or lower than the private one. Most growth theorists instinctively think about the former case, as they have in mind models with positive spillovers from human capital. However, Pritchett's review of the evidence is typical in finding very little empirical support for positive externalities.[38] On the other hand, various versions of the education-as-signalling-device model, as well as models

[38]See, e.g., Heckman and Klenow (1997) and Acemoglu and Angrist (2000).

of rent seeking, imply that the social return to education is lower than the private return.[39] This possibility is quite compelling. Note, however, that our calculations above imply that if we uniformly lower the social rate of return to education, cross-country schooling inequality will explain *even less* of income inequality than it does in our benchmark calculation (see Figure 2).

Pritchett, however, also convincingly argues that the extent of rent seeking, and therefore the extent to which the social return is below the private return, is much larger in poor countries. For example, in many poor countries the government employs an overwhelmingly large share of college graduates. This is sometimes the result of guaranteed-employment rules that commit the government to find employment to anyone with a tertiary degree. In contrast, in rich countries most college graduates work in the private sector. Since standard rent-seeking arguments imply that the government sector is intrinsically likely to make less efficient use of resources, this implies that on average the social return to education will be lower in poor countries. This effect is of course reinforced by the fact that poor countries are notoriously more prone to corruption and rent seeking than rich ones.

This *will* help. If the social rate of return to education (and health) is allowed to be higher in rich countries, then the variance of h will increase, and with it the explanatory power of the model. How important this could be quantitatively is hard to say, but by all means it would be worth finding out. A first exploratory step may be to break down the labor force into government-employees and private-sector workers. One may then retain the parameterization of the benchmark case for the private sector workers, but assume lower returns for government employees.

5 Quality of Physical Capital

5.1 Composition

Recent research by Eaton and Kortum (2001) has shown that most of the world's capital is produced in a small number of R&D-intensive countries, while the rest of the world generally imports its equipment. This suggests that, for most countries, (widely available data on) imports of capital of a certain type are an adequate proxy for overall investment in that type of equipment. Caselli and Wilson (2004) exploit this observation to investigate cross-country differences in the composition of the capital stock.[40]

Their results – a partial summary of which is shown in Table 3 – are startling: different

[39] See, e.g., Murphy, Shleifer, and Vishny (1991), or Gelb, Knight, and Sabot (1991).

[40] This idea is also used in Caselli and Coleman (2001a). The equipment-import data are extracted from the Feenstra (2000) data set.

types of equipment constitute widely varying fractions of the overall capital stock across countries. For each of the nine equipment categories Caselli and Wilson work with, the share in total investment in 1995 has minima in the low single digits, and maxima that vary between 20 percent and 80 percent! The standard deviations of investment shares are always large relative to the cross-country means. Furthermore, this enormous heterogeneity is systematically related to per capita income, as the correlations with income of the various investment shares are large in absolute values.

	Fabricated Metal Products	Non-Electrical Machinery	Office Computing Accounting	Electrical Equipment	Comm. Equipment	Motor-Vehicles	Other Transp.	Aircraft	Prof. Goods
Mean	.08	.21	.06	.14	.11	.24	.03	.05	.07
STD	.06	.08	.05	.07	.05	.10	.04	.09	.03
Min	.01	.03	.01	.01	.01	.01	.00	.00	.01
Max	.55	.48	.41	.59	.37	.55	.34	.88	.23
Corr Y	-.25	-.14	.53	.27	.20	-.32	-.41	.14	.33
R&D	202	887	1170	848	2280	1810	57	1880	801

Table 3: The Composition of the Capital Stock

To begin to see why this vast heterogeneity in the composition of equipment may matter for development-accounting, Table 3 also reports global, cumulated R&D expenditures in the various equipment categories.[41] The wide variation in R&D spending across types reinforces the impression of equipment heterogeneity across countries: since equipment shares vary so much, so does the embodied-technology content of the aggregate capital stock. If the R&D content of equipment determines its quality, i.e. its productivity per dollar of market value, one begins to suspect that the quality of capital – and not only its quantity – may vary across countries.

Furthermore, these differences are systematic, since richer countries appear to employ high R&D capital much more than poor countries. A simple way to see this is to look at Figure 14, from Wilson (2004). For equipment types with a high R&D content the share in overall investment is positively correlated with output per worker, while the opposite is true for low-tech equipment types. Could it be that rich countries use higher quality equipment, and that this higher quality accounts for some of the residual TFP variance?

To see how this may work it is useful to write down a very simple model. Imagine that final output Y is produced combining various intermediate inputs, x_p, according to the

[41] This is sometimes referred to as the "R&D stock" of a certain type of capital, as it is computed by cumulating past R&D spending with the perpetual inventory method. The R&D spending data come from the ANBERD data base. See Caselli and Wilson (2004) for details.

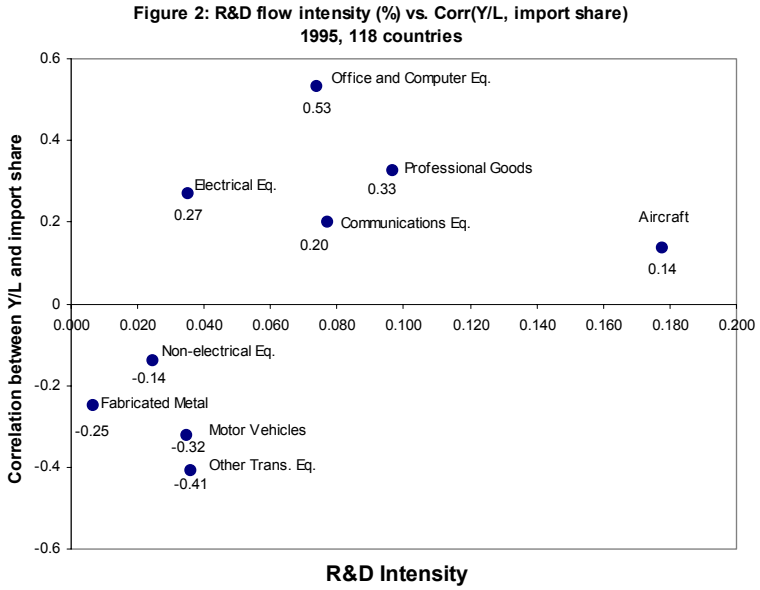

Figure 14: Capital Composition and Income

CES production function[42]

$$Y = B\left[\sum_{p=1}^{P}(x_p)^{\gamma}\right]^{\frac{1}{\gamma}} \qquad \gamma < 1,$$

where B is a disembodied total factor productivity term. Intermediate-good p is produced combining equipment and labor:

$$x_p = A_p\left(h_p L_p\right)^{1-\alpha}\left(K_p\right)^{\alpha} \qquad 0 < \alpha < 1, \tag{9}$$

where K_p measures the quantity of equipment (in current dollars) used to produce intermediate-input p, $h_p L_p$ is human-capital augmented labor in sector p, and A_p is the productivity of sector p.

The key assumption is that capital is heterogeneous: there are P distinct types of capital, and each type is product specific, in the sense that intermediate p can only be produced with capital of type p. In other words, an intermediate is identified by the type of equipment that is used in its production.[43] The assumption that $\gamma < 1$ implies that – in producing aggregate output – all these activities are imperfect substitutes.

[42]Production functions such as this one have been the staple of recent developments in growth theory.

[43]For example, for equipment-type "trucks," the corresponding intermediate good x (say, "road transportation") is the one obtained by combining workers with trucks. For equipment-type "computers," the

The productivity term A_p is product specific. Product variation in A allows for the possibility that one dollar spent on equipment of type p may deliver different amounts of efficiency units if instead spent on type p'. For example, the embodied-technology content of good p may be greater because the industry producing equipment of type p is more R&D intensive.[44]

A number of simplifying assumptions allow one to write a simple formula that brings the idea of the "quality of capital" in sharp relief. In keeping with the representative-agent spirit of the rest of the chapter say, then, that human capital per worker is constant across sector, i.e. $h_p = h$, and that labor is free to flow across sectors so that the marginal product of labor is the same for all equipment types. Then one can show that the output equation can be rewritten as[45]

$$Y = (K)^{\alpha} (hL)^{1-\alpha} B \left(\sum_{p=1}^{P} (A_p)^{\frac{\gamma}{1-(1-\alpha)\gamma}} (\xi_p)^{\frac{\alpha\gamma}{1-(1-\alpha)\gamma}} \right)^{\frac{1-(1-\alpha)\gamma}{\gamma}}, \tag{10}$$

where of course K is the total market value of the capital stock, L is total labor, and ξ_p is the share of capital of type p in the total capital stock, or $\xi_p = K_p/K$.

Up until now we have been writing $Y = (K)^{\alpha} (hL)^{1-\alpha} A$, and we have struggled with the fact that A seems to play an enormously important role in determining output differences. The last equation neatly shows the relationship between A and the composition of the capital stock: if different types of capital have different productivities, then the observed wild variation in equipment shares ξ implies that the quality of capital – over and above its quantity K – can vary across countries and can account for a portion of the unexplained variation.

Caselli and Wilson (2004) propose a regression-based approach to make inferences on the various A_ps. Unfortunately, even with knowledge of the A_ps it is virtually impossible to

corresponding intermediate good is "computing services," etc. Hence, our intermediates do not easily map into industries or sectors (computers are used in most industries), but rather into the various types of activities (transport, computing, etc.) required to generate output within each sector.

[44] Since, for simplicity, we have written an aggregate production function that is symmetric in the services provided by different types of capital, product variation in A_p may also reflect differences in the various capital types' shares in aggregate output.

Caselli and Wilson also allow the productivity term A_p to vary across countries, for a given equipment type p. The idea behind country variation is that equipment of type p may be more complementary with the characteristics of country i, an idea that is strongly supported by their empirical results.

In order to make sure that cross-country differences in K_p measure physical differences in installed capital we need to assume that the law of one price holds. This is plausible, since we know that most capital is imported from a few world producers. If the law of one price does not hold, however, cross-country differences in A_p may also reflect price differences.

[45] See Caselli and Wilson (2004) for details.

bound the amount of income variance that the ξ_ps can explain. This is because the last term of equation (10) is exceedingly sensitive to the value of γ, and there seem at the moment to exist no reliable approach to the calibration of this elasticity. If γ is sufficiently low, i.e. capital types are sufficiently poor substitutes for one another, the quality of capital accounts for *all* of the unexplained component of income differences. However, whether such values are reasonable or not our current state of knowledge cannot say. I conclude therefore that further research on the composition of capital is an important priority for development accounting.

5.2 Vintage Effects

Solow's (1959) paper on vintage capital formalized the idea that technological progress is embodied in capital goods. Jovanovic and Rob (1997), Parente (2000), Mateos-Planas (2000), and Rodriguez-Clare (1996) have noted that this could potentially enhance the explanatory power of cross-country differences in investment rates. The idea, of course, is that low investment rates will be associated with lower adoption of new technology. Indeed, some of these authors have argued that versions of the capital-embodied model greatly outperform the homogenous-capital model in accounting for cross-country income differences.

Formally, the vintage-capital model could be described by the formulas:

$$\begin{aligned} Y_t &= (L_t h_t)^{1-\alpha} \sum_{i=0}^{t} A_{t-i} K_{t-i}^{\alpha}, \\ K_{t-i} &= (1-\delta)^i \tilde{I}_{t-1}, \end{aligned} \tag{11}$$

where $\tilde{I}_{t-1}$ is investment at time $t-1$ *in terms of the consumption good*. Consistent with Solow's idea, this model has the property that capital installed by sacrificing one unit of consumption at date s yields $A_s(1-\delta)^{t-s}$ efficiency units of capital at time t, while capital installed by sacrificing one unit of consumption at time t yields A_t efficiency units. If $A_s < A_t$ we have that the earlier sacrifice in consumption contributes less to output today not only because of physical depreciation, but also because of the older vintage. To appreciate the potential consequences of this note that the same comparison in a homogenous-capital (i.e. disembodied technical change) model would be between $A_t(1-\delta)^{t-s}$ and A_t, i.e. differences in efficiency units obtained with the same sacrifice of consumption would only be due to physical depreciation.

As explained by Greenwood, Hercowitz, and Krusell (1997), under certain conditions the formulation above is equivalent to

$$\begin{aligned} Y_t &= (L_t h_t)^{1-\alpha} \sum_{i=0}^{t} K_{t-i}^{\alpha}, \\ \tilde{K}_{t-i} &= (1-\delta)^i q_{t-i} \tilde{I}_{t-i}, \end{aligned}$$

which has the following interpretation. Instead of one unit of consumption producing equal amounts of capital of increasing quality at different dates, one unit of consumption produces increasing amounts of capital of the same quality. Because the aggregate implications of growth in A are isomorphic to those of growth in q, the two formulations can be equivalent representations of the idea of embodied technological progress.

The second version of the model suggests, however, that – at least in principle – the estimates of the physical capital stocks we have been using until now do actually already reflect embodied technical change. This is because the real investment series we construct from PWT6 is a series for real investment in terms of the investment good, and not in terms of the consumption good. In other words, the PWT6 investment data that we use are data on $I_s = q_s \tilde{I}_s$, and not on $\tilde{I}_s$. Therefore, vintage effects – or at least those vintage effects that show up in a reduced relative price of investment goods – should already be accounted for.

As a very rough check on this argument, I have run a cross-country OLS regression of output per quality-adjusted worker on a distributed-lag function of *depreciated* investments (in units of the investment good). I.e. the left hand side variable was $Y_t/(L_t h_t)^{1-\alpha}$, and the right hand side variables where I_t, $(1-\delta)I_{t-1}$, $(1-\delta)^2 I_{t-2}$ ... I experimented with 5, 10, and 20 lags of investment. The homogeneous-capital hypothesis – or, much more accurately, the hypothesis that all vintage effects are adequately captured by investment-good prices in PWT6 – is equivalent to all the coefficients taking the same value, irrespective of the vintage. The "vintage effects" hypothesis would predict that coefficients on recent lags of investment would be systematically larger than those on older lags. The result was somewhat inconclusive, in that both hypotheses were rejected: all coefficients were not statistically the same, but neither they fell monotonically with the lag of investment. A possible explanation is that the price deflators in PWT work well enough to remove systematic vintage effects, but the remaining i.i.d. measurement error occasionally makes some vintages look more productive than others. In any case there is little indication that vintage-based models will significantly improve on the benchmark.

5.3 Further Problems with K

The investment series I have used to estimate the capital stock is an aggregate of private and public investment expenditures. As Pritchett (2000) very convincingly argues, however, elementary logic and vast anecdotal experience suggest that many governments' investment efforts are much less productive than private ones. There is an infinite supply of examples where government investments have not produced anything tangible (non-existent highways, industrial complexes that have never been completed, etc.). Furthermore, even when public investments do materialize, the resulting structures and machinery may be run less efficiently

than under private management.[46]

As for Pritchett's (2003) criticism of schooling-based measures of human capital based on private returns, his (2000) criticism of what he derogatorily but accurately calls CUDIE (cumulated depreciated investment expenditure) may help shedding light on the puzzles we are concerned with, both because governments tend to account for a larger share of production, employment, and capital ownership in poorer countries, and because less-accountable poor-country governments are likely to be disproportionately less efficient (relative to the private sector) than rich country ones. Hence, there are good reasons to expect the government to play an especially detrimental role in the productivity of investment in poor countries. This implies that the "effective" variance of K is larger than in the baseline model.[47]

As I suggested in the previous section for the analogous problem with human capital, a first pass at investigating this issue would be to try to separate out public from private investment, and apply different weights in the perpetual-inventory calculation, which would become

$$K_t = I_{private,t} + \gamma I_{public,t} - \delta K_{t-1}.$$

One could then try to re-do development accounting with this modified capital measure (possibly for various values of γ). Unfortunately, I have not been able to identify reliable and updated PPP breakdowns of the investment series into private and public capital.[48]

Perhaps a cleaner exercise, but also even more ambitious, would be to try to completely net out the government from the development accounting exercise. I.e. subtract the government's share from aggregate output, capital ownership, and employment, and perform the development-accounting exercise on the residual (private) inputs and output. This confronts the same data limitations as the exercise described in the previous paragraph, and the additional problem of coming up with a reliable PPP government share of GDP.[49]

[46]The 1994 World Bank's *World Development Report* documents substantial cross-countryy differences in the efficiency with which public infrastructure is used. Hulten (1996) uses these inefficiency indicators in a growth-regression exercise.

[47]While the emphasis is on the role of the government, Pritchett gives various other reasons why CUDIE is problematic. One reason he does not list is that poor countries tend to experience more frequent destructions of capital dur to natural disasters [Caselli and Malhotra (2004)] or wars. Since the perpetual inventory model assumes a constant rate of depreciation common for all countries, this implies again that the benchmark numbers underestimate the variability of k.

[48]However, both the WDI and the IMF's Government Finance Statistics (GFS) have data on government capital formation in domestic currency. One could therefore potentially construct a non-government investment series, and – in principle – use PWT deflators to turn this series into a real investment series. Whether one would get anything sensible out of this is another matter.

[49]One can of course not simply subtract G from Y, in PWT, because G is government consumption of goods and services, and not government production. In the NIPA government output is defined as the sum of factor payments, i.e. compensation to general-government employees plus general-government consumption

6 Sectorial Differences in TFP

Up to here in this chapter I have treated a country's GDP as if it was produced in a single sector, i.e. as if GDP measured the physical output of a homogeneous good. The basic message has been that it is impossible to explain cross-country differences in income without admitting a large role for differences in TFP. It is tempting to jump to the conclusion that these TFP differences signal the existence of barriers to technology adoption in less developed countries, or other frictions that broadly make some countries "function" less well than others. But large differences in TFP could also be the result of variation in the weights in GDP of sectors with different sectorial-level productivity – even when these sectorial productivities are identical across countries. In this case we would want to focus on barriers to the mobility of factors across sectors, instead of barriers to the mobility of technology or work practices across countries.[50]

6.1 Industry Studies

There is a tradition of productivity comparisons at the industry, or even at the firm, level. Particularly illustrative of the advantages of this approach are a series of reports published by the McKinsey Global Institute. These studies focus on painstaking comparisons of the production functions (broadly construed) of narrowly defined industries (from automobile, to beer, to retail banking) in a few industrialized economies (mostly US, Japan, Germany, and the UK). Baily and Solow (2001) present a thoughtful survey of the achievements, as well as the shortcomings, of these studies (as well as extensive references).[51]

Briefly, even within narrowly defined industries, and even among countries at very similar levels of development, total factor productivity presents remarkable variation. A similar conclusion, for somewhat more aggregated manufacturing industries (but more countries), is reached by Harrigan (1997, 1999). Hence, industry-level studies suggest that aggregate TFP differences are not solely due to differences in the weights of high- and low-TFP sectors. Since these studies are often limited to industries in a few highly developed economies, however, one should be cautious before assuming that the same causes drive the low TFP levels of less developed economies.

Besides confirming that TFP differences exist also at the industry level, the McKinsey

of fixed capital. Something like these categories are reported in the IMF's Government Financial Statistics. Unfortunately, the spottiness of these data forced me to abandon this particular enterprise.

[50] Here we are arguing that differences in sectoral composition may account for some of the differences in the *level* of GDP. Koren and Tenreyro (2004) show that these same differences account for a substantial fraction of differences in its *volatility*.

[51] A precursor to these intra-industry cross-country productivity comparisons is the three-country study of Conrad and Jorgenson (1985). See also Wagner and Van Ark (1996).

researchers are often also able to shed some tentative light on their sources. In particular, they highlight differences in working practices, and they are sometimes also able to link inefficient practices to the regulatory environment. In general, the studies point to a link between the degree of competition domestic producers are exposed to (as affected by the amount of regulation), and the efficiency with which they organize their labor input. This is of course an important, and plausible, finding. Further support for this view comes from the work of Schmitz (2001) on the North-American iron-ore industry, which shows convincingly that the efficiency of labor practices is very responsive to the degree of product-market competition.[52]

6.2 The Role of Agriculture

As mentioned, existing cross-country comparisons of sectorial TFP tend to be limited to small sets of developed countries. The goal of this section is therefore to provide a rough, preliminary assessment of the sectorial-composition interpretation of TFP differences that extends to developing countries as well. In particular, I will focus on an agriculture-nonagriculture split of GDP. The main reason for looking at this particular breakdown is easily inferred from Figure 15: in the poorest countries of the world virtually everyone works in agriculture, and in the richest virtually nobody does. It is obvious that this is the most important source of variation in the composition of GDP around the World. Another reason for focusing on agriculture is that I have no PPP output data for other sectors. Finally, the agriculture-nonagriculture dualism has traditionally played a central role in the history of thought on economic development.[53]

The main purpose of this section, then, is to assess the hypothesis that (i) agriculture is an intrinsically low TFP sector, and (ii) poor countries' low aggregate TFP is due in substantial measure to their high shares of agriculture. In this subsection, however, I start by taking a preliminary (and perhaps somewhat digressive) look at basic data on agricultural

[52]Another wonderful industry-level comparison of cross-country productivity differences is presented by Clark (1987), who examines the productivity of cotton mills around the world in the early years of the twentieth century. He shows that, assuming constant capital-labor ratios, the textile industries of Britain and New England would have had a huge cost disadvantage relative to India, Japan, and many other countries. Yet, British cotton textiles dominated export markets. Clark shows that the various countries' industries used identical equipment, and that the expertise to organize and run the mills could not have differed too much. Rather, the source of the productivity differences boils down to the fact that each English worker was willing to tend to a much larger number of machines. In low-productivity countries workers were idle most of the time. Why this was so remains a bit of a mystery, and one shouold be cautious in assuming that this finding would still hold up one century later. Nevertheless, Clark's findings reinforce the case that labor practices may be an important source of observed differences in productivity.

[53]Some of the classics are Fisher (1945), Clark (1940), Rostow (1960), Nurske (1953), Lewis (1954), Kuznets (1966), and Jorgenson (1961).

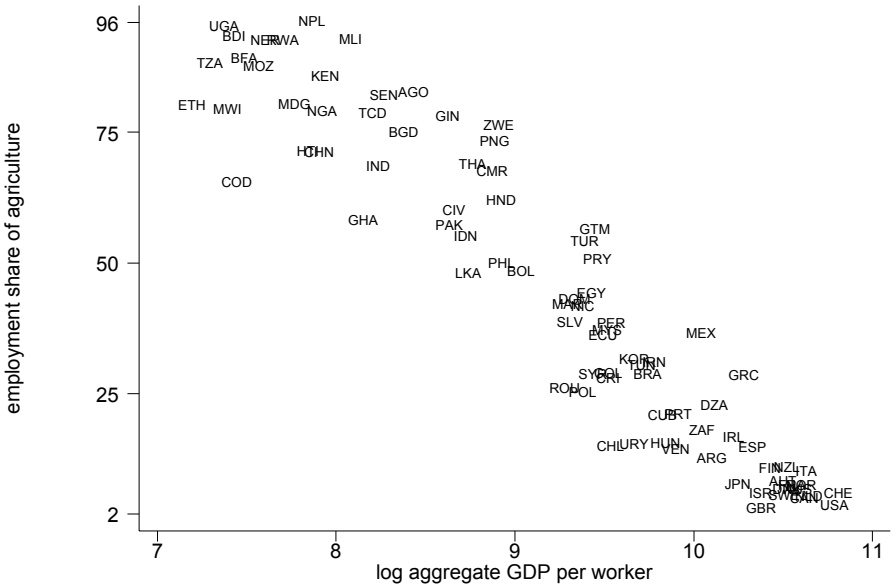

Figure 15: The Importance of Agriculture

GDPs, non-agricultural GDPs, and agricultural labor shares. What we find should provide further motivation for asking the development-accounting question with disaggregated data.

I begin by writing per-worker GDP in PPP as

$$y = P_A y_A l_A + P_{\overline{A}} y_{\overline{A}} l_{\overline{A}}, \tag{12}$$

where P_A ($P_{\overline{A}}$) is the international (PPP) price of agricultural (non-agricultural) goods, y_A ($y_{\overline{A}}$) is the per-worker output of the agricultural (non-agricultural) sector, and l_A ($l_{\overline{A}}$) is the agricultural (non-agricultural) share of employment. We already know that l_A varies dramatically across countries. What about $P_A y_A$ and $P_{\overline{A}} y_{\overline{A}}$?

The Food and Agriculture Organization (FAO) has collected and published cross-country data on *producer prices* in agriculture for a large number of countries between 1970 and 1990 [Rao (1993), see also Restuccia, Yang, and Zhu (2003)]. This permits the construction of PPP exchange rates for agriculture, and therefore of PPP comparisons of agricultural output. Going even further, the FAO researchers also assembled some data on agricultural inputs for the year 1985, and this allowed them to generate a cross-section of PPP agricultural GDPs. Furthermore, the methodology followed in the FAO study deliberately follows the methods of Summers and Heston (1991). Hence, the estimates of PPP agricultural GDP in the FAO data set are comparable to the aggregate PPP numbers of PWT6. The FAO data set also obviously contains information on agricultural employment (which was used for

Figure 15).

A difficulty that needs to be addressed, however, is that the FAO numbers for PPP agricultural GDP do not directly map into the quantity $P_A y_A$ in equation (12). The reason is as follows. The FAO PPP GDPs aggregate the quantities of the various agricultural products by a set of "international prices," that are essentially weighted averages of each countries's prices. The same is done in the PWT for all goods and services. However, the two systems use a different normalization for the international prices - i.e. they have the same relative prices of agricultural products but different absolute levels. Hence, the PPP agricultural value-added coming from the FAO data set cannot simply be plugged into equation (12) with PWT aggregate value added on the left hand side.[54]

I try to solve this problem as follows. It is well known that – because they are quantity-weighted – international-dollar prices in the PWT closely resemble rich-country, and especially US, prices (Hill 2000). Hence, for the US, we should have

$$\frac{P_{A,US}Y_{A,US}}{Y_{US}} \approx \frac{P^D_{A,US}Y^D_{A,US}}{Y^D_{US}},$$

where $P^D_{A,US}Y^D_{A,US}/Y^D_{US}$ is the agricultural share in GDP at *domestic prices*, which can be obtained from the WDI. In these equations Y_{US}, and $P^D_{A,US}Y^D_{A,US}/Y^D_{US}$ are known (the former from PWT, and the latter from WDI). Hence, we can solve for the US value of $P_{A,US}Y_{A,US}$. Now recall that the FAO estimates for PPP agricultural GDP differ from the (implicit) PWT estimates only by a constant of normalization. It should follow that if we rescale all of the FAO agricultural GDP numbers such that the US value coincides with $P_{A,US}Y_{A,US}$ (as just calculated) we have an estimate of the contribution of agriculture to PWT GDP that we can plug into equation (12).

As already pointed out by Restuccia, Yang, and Zhu (2003), the most striking feature of the FAO data is that variation in agricultural value-added per worker (in PPP) dwarfs the variation in aggregate value-added per worker. In the largest sample with data on both agricultural and aggregate GDP per worker (80 countries) the inter-percentile range in agri-

[54]Formulas may help here. Suppose that there is only one agricultural good and one nonagricultural good. Call P_A and $P_{\overline{A}}$ their respective prices in PWT international dollars: the PWT's unit of account. The normalization used in PWT is that $P_A y_{A,US} + P_{\overline{A}} y_{\overline{A},US} = P_{A,US}\ y_{A,US} + P_{\overline{A},US}\ y_{\overline{A},US}$, where $P_{A,US}$ ($P_{\overline{A},US}$) is the price of A ($\overline{A}$) in the US (this is what Summers and Heston mean when they say, somewhat opaquely, that the the PPP of the US is 1). Instead, the normalization in the FAO data set is $P^F_A\, y_{A,US} = P_{A,US}\, y_{A,US}$, where P^F_A is the price of the agricultural good in FAO's international dollar. In this two-good example this obvioulay implies $P^F_A = P_{A,US}$, so that $P^F_A\, y_A = P_A\, y_A$ only if we have $P_A = P_{A,US}$, as well. But the above-described normalization of PWT prices does not assure this at all, and indeed it would be true only by coincidence. I guess one could put this into PWT-speak, and say that the fact that the PPP of GDP is normalized to 1 for the US, does not imply that the PPP of individual sectors (such as agriculture) is normalized to 1 as well.

cultural GDP is 45 and the log-variance is 2.15. The corresponding numbers for aggregate GDP are 19 and 1.09, respectively.[55] Real agricultural GDP per worker is plotted in Figure 16 against real aggregate GDP per worker.

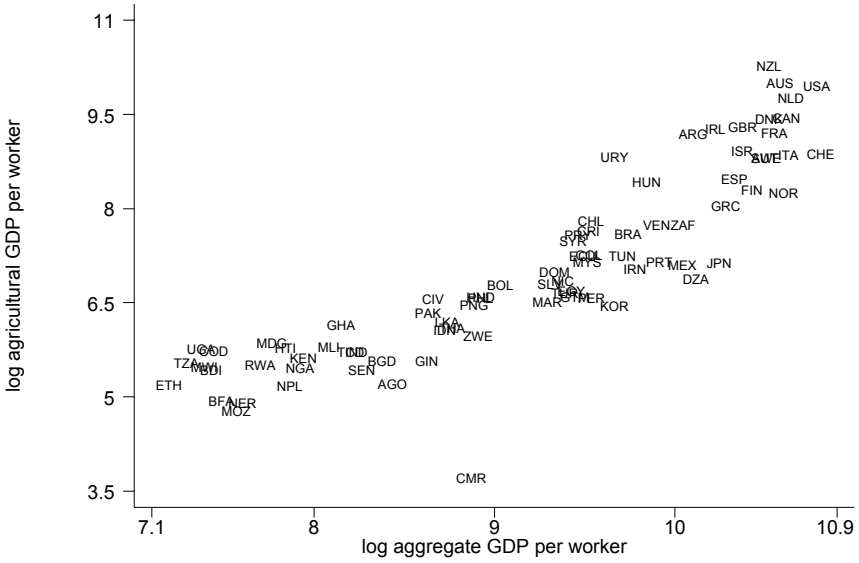

Figure 16: Labor Productivity in Agriculture

Subtracting real agricultural GDP from aggregate GDP, it is also possible to back out non-agricultural value-added per worker.[56] The (not surprising but nonetheless) very important finding is that differences in labor productivity in the non-agricultural sector are much smaller than differences in aggregate labor productivity (and, a fortiori, in agricultural labor productivity). The inter-percentile range is only 3.69 (compared with 19 for aggregate GDP and 45 for agriculture) and the log-variance is 0.29 (compared with 1.10 and 2.15). Figure 17 plots non-agricultural value-added per worker against aggregate value-added per-worker. Comparison of Figures 16 and 17 shows that labor productivity is generally higher outside than inside agriculture, and this is much more true for developing countries, an observation previously made by Gollin, Parente, and Rogerson (2000).[57]

[55]Due to the high persistence of the World's income distribution these last two numbers are very close to the corresponding numbers for our benchmark 1996 sample (20 and 1.2).

[56]This is where the measurement problem described before becomes important: the log-variance of agricultural GDP is obviously insensitive to the price-normalization adopted, but non-agricultural GDP is computed as a residual from equation (12), so it is crucial that $P_A y_A$ and y are in the same units.

[57]Only Australia and New Zealand have higher productivity in agriculture than in nonagriculture. The average log-difference between nonagricultural and agricultural output per worker in the entire 80-country sample is 2.23; among the poorest 20 it is 3.26; among the richest 20 it is 0.97.

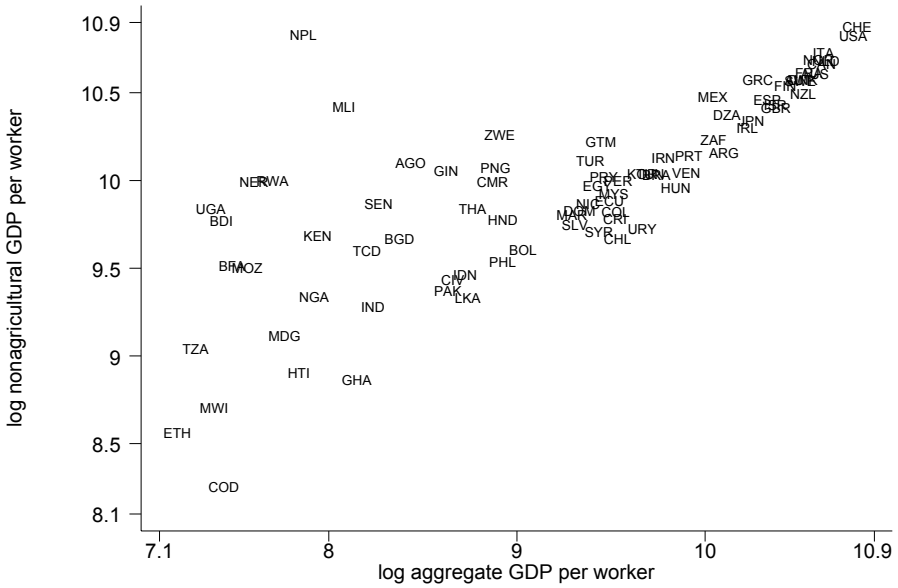

Figure 17: Labor Productivity outside of Agriculture

Recalling now from Figure 15 that the third component of equation (12), the employment share l_A, ranges from almost 0 percent in the richest countries to almost 100 percent in the poorest, we conclude that poor countries have most of their labor force in the sector where they are particularly unproductive.[58,59]

We can summarize this first overview of the sectorial data by saying that there are three proximate reasons for poor countries' poverty: their much lower labor productivity

[58] Attempts at explaining this apparent deviation from comparative advantage abound. It may be that the non-agricultural sector has greater skill requirements, so that low human-capital economies are constrained in the supply of non-agricultural workers (Caselli and Coleman, 2001b); or it could be that investment distortions push producers into the home (agricultural) sector (Gollin, Parente, and Rogerson, 2000); or it could be that economies are subject to a "subsistence constraint," such that resources cannot start moving out of agriculture until agriculture is sufficiently productive to generate a surplus that will feed the industrial class (Gollin, Parente, and Rogerson, 2001; Restuccia, Yang, and Zhu, 2003); or it could be that some countries are "trapped" in agriculture by a coordination failure (long tradition; most recently Graham and Temple, 2001).

[59] Given the huge employment shares of agriculture in figure 15 one would guess that in most developing countries agriculture would account for an equally vast share of GDP. In fact, the agricultural share of GDP is always below 40 percent. This is a consequence of the disproportionately low productivity of agriculture in low-income countries. Also note that the PPP agricultural share in GDP is both much less variable across countries, and – for most countries – lower than the domestic-currency agricultural share in GDP. This is because – perhaps contrary to common wisdom – the relative price of agricultural goods is higher in poor countries than in the US.

in agriculture; their somewhat lower labor productivity outside agriculture; and their larger share of employment in the sector that – on average – is less productive.

To quantify these effects Table 4 presents income-dispersion statistics (log-variance and inter-percentile range) in the data (first row), and under alternative counterfactual assumptions on industry-level productivity and labor shares. Counterfactual 1 is that all countries have the US level of agricultural GDP per worker, but their own level of non-agricultural GDP per worker and agricultural labor share. Counterfactual 2 is that all countries have the US-level of non-agricultural GDP per worker, but their own level of agricultural GDP per worker and agricultural labor share. Counterfactual 3 is that all countries have the US agricultural labor share, but their own level of agricultural and non-agricultural GDP per worker.[60]

Variable	log-variance	int. range
actual real output per worker	1.10	19
counterfactual 1: US y_A, own $y_{\overline{A}}$ & l_A	0.04	1.6
counterfactual 2: US $y_{\overline{A}}$, own y_A & l_A	0.58	7.1
counterfactual 3: US l_A, own y_A & $y_{\overline{A}}$	0.30	3.8

Table 4: Counterfactual World Income Distributions

The results are stunning. The figures in the second row imply that if poor countries achieved the same level of agricultural labor productivity as the US, world income inequality would virtually disappear! This is of course a reflection of the convergence of US agricultural incomes to US non-agricultural incomes [documented in Caselli and Coleman (2001b)], as well as the huge agricultural share of employment in many of the poorest countries. However, the other two counterfactual experiments also generate large declines in dispersion. Because agriculture is generally much less productive than non-agriculture, reducing the agricultural employment share to US levels would reduce income inequality by an enormous two thirds (third row). And cross-country non-agricultural productivity differences, while much less than agricultural ones, are still sufficiently large that income inequality would fall by about one half if poor countries were as productive outside of agriculture as the US (second row).[61]

[60] A similar calculation is reported by Restuccia, Yang, and Zhu (2003). Calculations in this spirit can also be found in Caselli and Coleman (2001b) (for US regions), and Gollin, Parente, and Rogerson (2001).

[61] Of course, this discussion abstracts from the changes in world-wide agricultural relative prices that such changes would bring about.

6.3 Sectorial Composition and Development Accounting

The previous subsection establishes that there are very large *within industry* cross-country differences in output per worker. Indeed, the agricultural GDP differences are substantially larger than the aggregate GDP ones. Furthermore, these cross-country differences in industry GDPs are seen to potentially "account" for a large fraction of the cross-country dispersion in aggregate income. We can now return to the original question: are these large differences in agricultural GDP attributable to the amounts of observable inputs employed in agriculture by the various countries, or are they the result of industry-level cross-country TFP differences? This is of course "the" development-accounting question.

To try to answer this question, we need assumptions on the industry production functions, as well as ways of measuring industry-level inputs. I will assume that each of the two sectors produces according to a Cobb-Douglas technology. In agriculture, the factors of production are capital, labor, and land (T). In non-agriculture, they are capital and labor:

$$Y_A = A_A \left(K_A\right)^{\alpha_A} \left(L_A h_A\right)^{\beta_A} \left(T_A\right)^{1-\alpha_A-\beta_A}, \tag{13}$$

$$Y_{\overline{A}} = A_{\overline{A}} \left(K_{\overline{A}}\right)^{\alpha_{\overline{A}}} \left(L_{\overline{A}} h_{\overline{A}}\right)^{1-\alpha_{\overline{A}}}. \tag{14}$$

The goal now is to construct counter-factual agricultural and non-agricultural output data,

$$\begin{aligned} y_{A,HK} &= k_A^{\alpha_A} h_A^{\beta_A} t_A^{1-\alpha_A-\beta_A} \\ y_{\overline{A},HK} &= k_{\overline{A}}^{\alpha_{\overline{A}}} h_{\overline{A}}^{1-\alpha_{\overline{A}}}. \end{aligned}$$

These counter-factual data answer the question: what would the world distribution *of agricultural (non-agricultural) output per worker* look like if all countries had the same agricultural (non-agricultural) total factor productivity?

Assume that the rates of return on capital must be equalized across sectors – a plausible arbitrage condition. This is easily seen to imply

$$\alpha_A \frac{P_A^D Y_A}{K_A} = \alpha_{\overline{A}} \frac{P_{\overline{A}}^D Y_{\overline{A}}}{K_{\overline{A}}},$$

where P_A^D ($P_{\overline{A}}^D$) is the *domestic* producer price of agricultural (non-agricultural) goods, which will generally differ from the PPP price, and is the price the domestic investor cares for (unless he produces for the export market). The quantities $P_A^D Y_A$ and $P_{\overline{A}}^D Y_{\overline{A}}$ are, of course, agricultural and non-agricultural output *in domestic prices,* and they are observable from WDI.[62] Hence, combining this equation with

$$K_A + K_{\overline{A}} = K, \tag{15}$$

[62] In fact, we use the share of agriculture in GDP in domestic prices from the WDI. This is sufficient for the calculation below since only the ratio $\frac{Y_A^D}{Y_{\overline{A}}^D}$ is needed.

where K is the (observable) total capital stock, we can back out K_A and $K_{\overline{A}}$, and hence (with the FAO data on labor shares) k_A and $k_{\overline{A}}$.[63]

It is harder to come up with numbers for h_A and $h_{\overline{A}}$. Caselli and Coleman (2001b) show that there is a very systematic tendency for agriculture to be one of the least skill intensive sectors in the economy. For example, for each of the years between 1940 and 1990 they rank the roughly 120 industries featured in the US *Census of Population* by percentage of workers with an elementary degree or less, and in each of these years agriculture consistently ranks in the bottom 10. This suggests that it may not be unreasonable to set $h_A = 1$ in all countries, i.e. that the agricultural work force is made up by workers with no education. It is then easy to compute human capital per worker outside of agriculture. In particular, if s is the average years of schooling in the labor force, and a fraction l_A has no education and work in agriculture, then the remaining fraction $1 - l_A$ must have years of education $s_{\overline{A}} = s/(1-l_A)$. $h_{\overline{A}}$ is then computed with the "standard" formula linking years of education to human capital by way of Mincerian returns.

For t_A I simply plug in WDI data on each country's endowment of arable land. For the parameters, I use essentially the same calibration as in Caselli and Coleman (2001b), which is in turn based on the work of Jorgenson and Gollop (1992). The labor share in the USA is about 60% in both farming and non-farming. The capital share in agriculture is about 21%, and the remainder is absorbed by land.

For the 65 countries for which we can construct $y_{A,HK}$ and $y_{\overline{A},HK}$ *and* for which we have measures of y_A and $y_{\overline{A}}$, the success of the factor-only model is as reported in Table 5. Once again, the results are striking. Briefly, the factor-only model explains virtually nothing of the observed per-capita income variance in agriculture: it's entirely a story of TFP differences, even more so than for aggregate GDP. Conversely, physical and human capital inputs do a better than usual job at explaining per-worker output differences outside of agriculture. This may be plausible, as knowledge flows are probably more effective in manufacturing or services than in agriculture. Still, the TFP scales are still tipped against the developing countries.

While informative, this exercise does not yet answer the development accounting question of how much PPP income per worker variation would be observed if all countries

[63] Instead of using a no-arbitrage condition for capital, sector specific capital inputs may be recovered using an indifference condition for rural-urban labor flows. This condition would involve, among other things, an "urban wage premium" compensating for costs of skill acquisition [as, for example, in Caselli and Coleman (2001b)] or for the lower probability of finding a job, as in the celebrated model of Harris and Todaro (1970). See Temple (2003) for an interesting calibration of the Harris-Todaro model aiming to assess the output costs of the labor-market rigidities that lead to economic dualism.

sector	success 1	success 2
agriculture	0.15	0.09
nonagriculture	0.65	0.79

Table 5: Success within sectors

had the same technology. This counter-factual PPP income per worker would be

$$y^i_{HK} = P_A A_A y^i_{A,KH} l^i_A + P_{\overline{A}} A_{\overline{A}} y^i_{\overline{A},KH} l^i_{\overline{A}} \tag{16}$$

where P_A and $P_{\overline{A}}$ are the international-dollar prices of agricultural and non-agricultural goods, and A_A and $A_{\overline{A}}$ are the efficiency levels *in some reference country.*

Because of its additive nature, implementing equation (16) calls for an explicit choice of values for A_A and $A_{\overline{A}}$ to keep constant as we vary country factor endowments and sectorial composition – summarized by $y^i_{A,KH} l^i_A$ and $y^i_{\overline{A},KH} l^i_{\overline{A}}$. (We did not have to choose a benchmark in the purely multiplicative framework of the one-sector model, because in that framework the common level of efficiency disappears when taking the log-variance or the inter-percentile ratio). Furthermore, the results will be sensitive to which country (i.e. what particular choice of As) is chosen as a reference. Hence, the development accountant must decide whose country's technology will be assigned the role of the benchmark technology in the counterfactual exercise.

A certain degree of arbitrariness is inevitable in this choice. Nevertheless, a somewhat plausible argument can be made that it makes sense to hold constant the technology of the richest country in the sample; in our case, the USA. This is because in a sense this is the most successful country, so it is interesting to know how the world income distribution would change if all countries shared the industry TFPs of the most successful among them.

For the reference country (the US) it is true by definition that

$$y_A = P_A A_A y_{A,KH},$$

and

$$y_{\overline{A}} = P_{\overline{A}} A_{\overline{A}} y_{\overline{A},KH}.$$

Since y_A and $y_{\overline{A}}$ are known, we can back out reference values for $P_A A_A$ and $P_{\overline{A}} A_{\overline{A}}$ and compute the counter-factual in (16). The result of this exercise, which is also the main result of this section, is 0.37 for $success_1$, and 0.35 for $sucess_2$. In words, once again, this means that, if all countries had the same *industry-level* TFPs as the US, but their observed allocation of measurable factors to agriculture and non-agriculture, the world distribution of income

would be about one third as unequal as it actually is. Given that – for this sub-sample – the corresponding success measures are 0.48 and 0.42 when the sectorial composition of GDP is not taken into account, I conclude that taking account of differences in sectorial composition actually *decreases* the share of cross-country income inequality that we can explain with a country's factor endowments.

This result should have been expected, by now. In the previous subsection we have seen that the dispersion in agricultural incomes per worker is a critical "source" of dispersion in per-capita income. Table 5 shows, however, that almost all of the variation in agricultural income comes from differences in agricultural TFP. It is not surprising, therefore, that we find that – even allowing for differences in output composition – factor endowments still don't work as the main cause of GDP differences.

One possible way to enhance the quantitative role of sectorial considerations is explored in a highly innovative paper by Graham and Temple (2001). Instead of assuming, as here, that both agriculture and non-agriculture have constant returns to scale, they follow a long tradition in development economics in hypothesizing that the former is characterized by decreasing returns and the latter by increasing returns. As is well known these assumptions tend to generate multiple equilibria, and it is therefore possible to try to explain large cross-country income differences with the argument that poor countries are in "low," i.e. high agriculture, i.e. low returns, equilibria; while rich countries are in industrialized equilibria and therefore benefit from the increasing returns. The difficulty here is to figure out in the data which countries are in the bad and which ones are in the good equilibrium. The contribution of Graham and Temple is to show a very ingenious way of solving this problem. They find that multiple equilibria explain a relatively large fraction of per capita income differences. The lingering question is whether the significant departures from constant returns to scale required for their result are plausible.[64]

7 Non-Neutral Differences

7.1 Basic Concepts and Qualitative Results

In all of the previous sections we have assumed that all differences in efficiency across countries are TFP differences, as summarized by the multiplicative factor A. This implies that we view differences in efficiency as *factor neutral*: some countries simply use all of their inputs more efficiently than others. This is of course a restriction on the set of possible efficiency differences. Caselli and Coleman (2004) have begun exploring a more general view, that allows for the possibility that differences in technology show up as differences in the efficiency with

[64] See also Chanda and Dalgaard (2003) for another contribution that argues for a large role of agriculture.

which specific factors – as opposed to all factors proportionally – are used, or even that some countries use some factors more efficiently, and some less efficiently, than others. In other words, a more general view of technology differences where such differences are *not factor neutral*.[65]

Extending the development-accounting exercise to allow for factor non neutrality in efficiency differences is the object of this section. While Caselli and Coleman consider a three-factor production function (capital, skilled labor, and unskilled labor), here I will stick to the two-factor world (human and physical capital) of the rest of the chapter. The first step we need to take to proceed in this direction is to replace the Cobb-Douglas restriction – which implicitly rules out non-neutrality – with a more general production function where non-neutral differences can be contemplated. The simplest such generalization is provided by the CES formula:

$$Y = [\alpha (A_k K)^\sigma + (1-\alpha)(A_h L h)^\sigma]^{1/\sigma} \qquad \alpha \in (0,1),\, \sigma < 1. \tag{17}$$

In (17) A_k and A_h are the efficiency units delivered by one unit of physical capital and one unit of quality-adjusted labor, respectively. If A_k is higher in one country than in another, we say that the former country uses capital more efficiently. If A_h is greater, the country uses human capital more efficiently. The parameters σ and α are constant across countries. σ governs the ease of substitution between physical and human capital. The elasticity of substitution is

$$\eta = 1/(1-\sigma).$$

The Cobb-Douglas case of the previous sections of the paper emerges as a limit for σ approaching 0 (η approaching 1). In this case, total factor productivity A converges to $A_k^\alpha A_h^{1-\alpha}$.

In the factor-neutral world explored so far in this chapter, making inference about efficiency differences across countries is a simple matter of solving one equation in one unknown. Inference on non neutral differences is a bit more challenging, as equation (17) has two unknowns: A_k and A_h. The issue, then, is to find a suitable second equation. As in Caselli and Coleman (2003), to do so I assume that factor markets are everywhere competitive. Then, if r is the user cost of capital, and if w is the market price of a unit of human capital, the following equations will hold:

$$\begin{aligned} r &= \alpha y^{1-\sigma} k^{\sigma-1} A_k^\sigma \\ w &= (1-\alpha) y^{1-\sigma} h^{\sigma-1} A_h^\sigma. \end{aligned} \tag{18}$$

Given values of α and σ, and data on y, k, h, r, and w, these two equations can be solved for the two unknowns A_k and A_h.[66]

[65] See also Hsieh (2000) for some observations on this topic.

[66] Alternatively, one could solve the system constituted by one of the factor-pricing equations and equation

Rearranging equations (18) we find the following formulas for the factor-specific efficiency levels

$$\begin{aligned} A_k &= \left(\frac{rk}{y}\frac{1}{\alpha}\right)^{1/\sigma}\frac{y}{k} = \left(\frac{S_k}{\alpha}\right)^{1/\sigma}\frac{y}{k} \\ A_h &= \left(\frac{wh}{y}\frac{1}{1-\alpha}\right)^{1/\sigma}\frac{y}{k} = \left(\frac{S_h}{1-\alpha}\right)^{1/\sigma}\frac{y}{h}, \end{aligned} \tag{19}$$

where S_k and S_h are the shares of physical and human capital in income, respectively. To see what these equations tell us about the way technology differs across countries it is useful to start from the case where factor shares are constant across countries, i.e. S_k and S_h are invariant parameters. Note that this is the assumption we have maintained so far throughout the paper, where we have set S_k to α (and consequently S_h to $1-\alpha$). Under this assumption, these equations have very intuitive implications: a high output-capital ratio implies that capital is used efficiently, and the same for human capital.

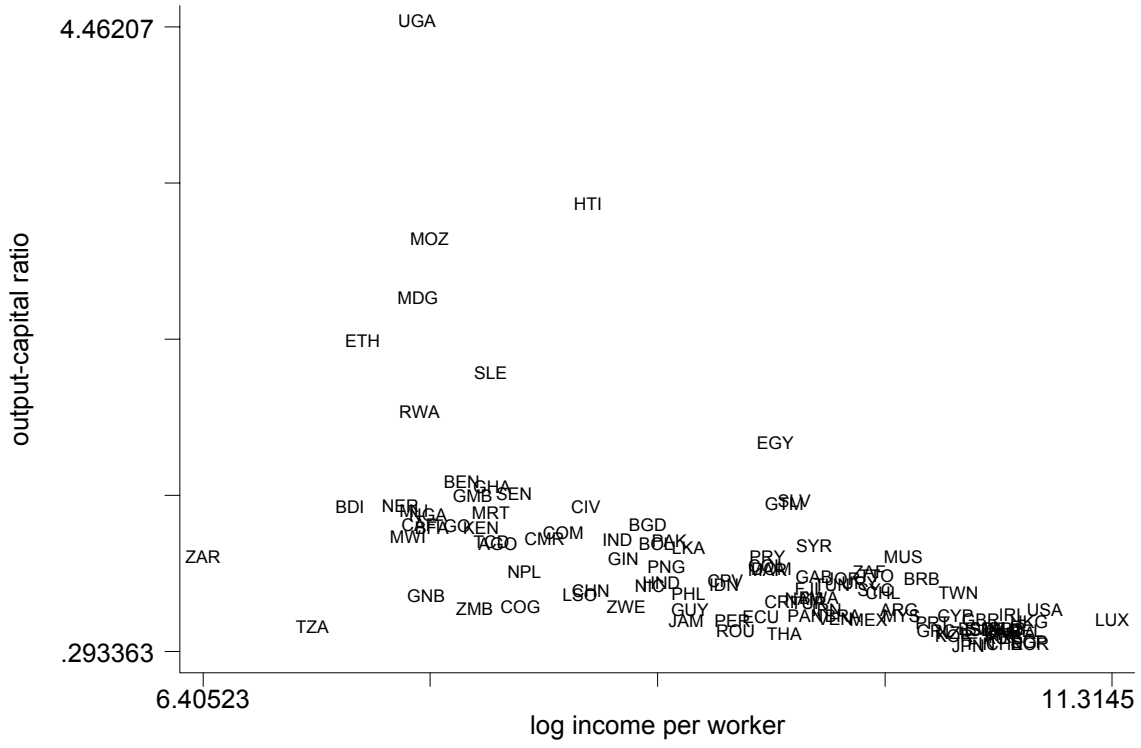

Figure 18: Distribution of y/k

(17). The result would be identical. The properties of the constant returns to scale production function, combined with national-account identities, imply that from any two of these equations the third follows.

Figure 18 plots the output-capital ratio y/k against the log of per-capita income, and Figure 19 does the same for the output-human capital ratio, y/h. As is well known, the output-capital ratio is decreasing in income. The output-human capital ratio, instead, is increasing. Hence, if we continue to assume that factor shares are constant across countries, but we allow for non-neutrality in technology differences, we reach the startling conclusion that rich countries use human capital more efficiently than poor countries, but they use physical capital less efficiently.[67]

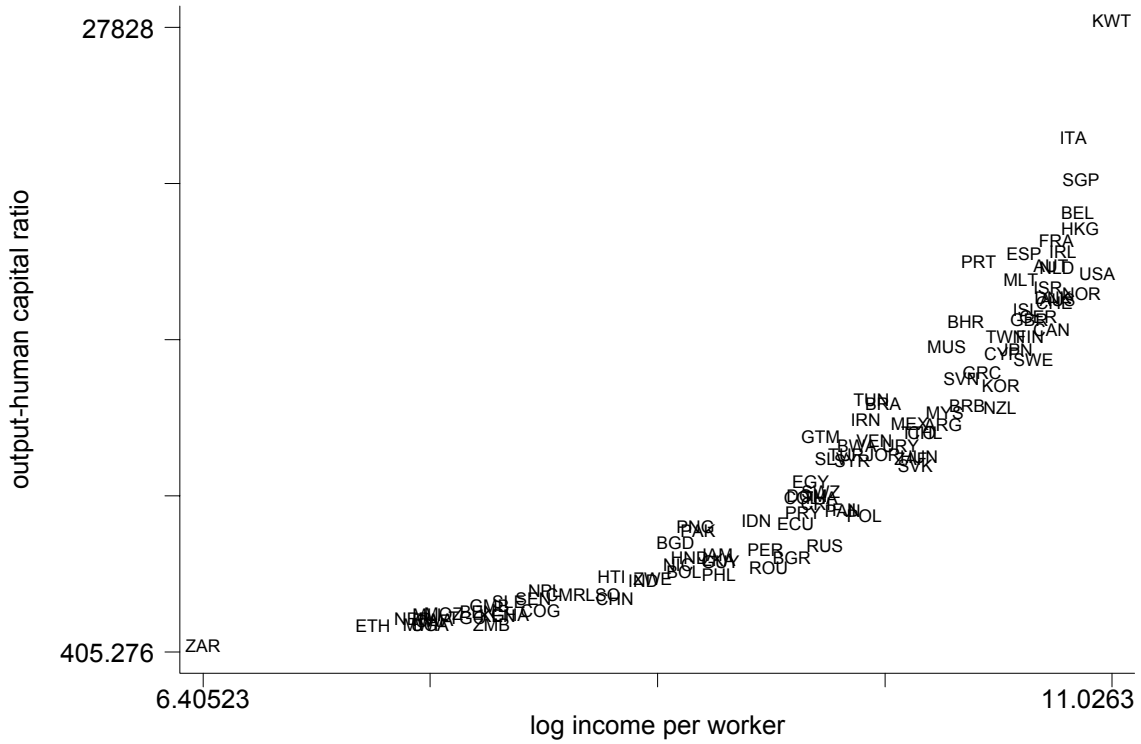

Figure 19: Distribution of y/h

Consider now relaxing the assumption of constant factor shares. Clearly our conclusions would be unchanged if the factor shares, while not constant, were not systematically related with income. Our knowledge of cross-country patterns in factor shares is somewhat

[67]Notice that, in the neutral world of the first part of the chapter, one way of writing total factor productivity is as

$$A = \left(\frac{y}{k}\right)^{\alpha} \left(\frac{y}{h}\right)^{1-\alpha}.$$

Hence, our conclusion there that rich countries are more efficient was based on the fact that the increasing pattern of $(y/h)^{1-\alpha}$ more than compensates for the decreasing pattern in $(y/k)^{\alpha}$.

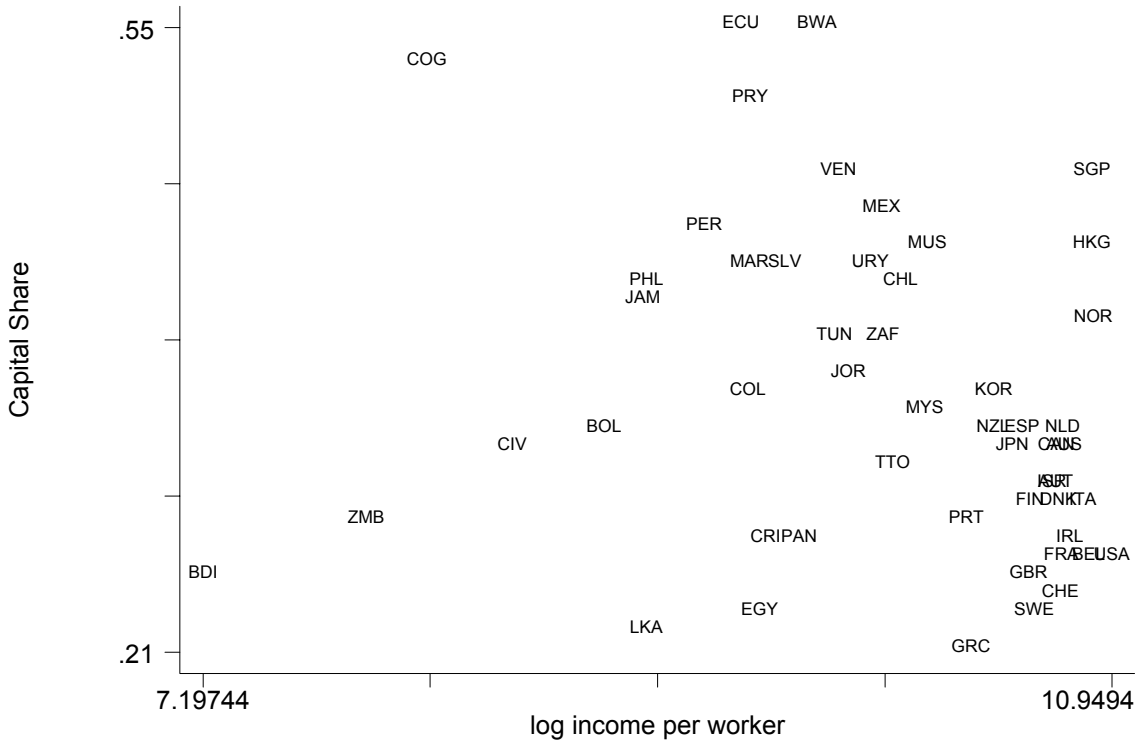

Figure 20: Distribution of S_k

limited. The only thing we are quite sure of is that in the US this share has historically been rather stable, at around 1/3. When it comes to cross-country comparisons, however, we are on shakier ground. Traditionally, the capital share – as measured in the national accounts – is calculated as a residual after employee compensation has been taken out. With this method, S_k is generally found to be higher in poor countries than in rich countries. Recently, however, Gollin (2002), and Bernanke and Gurkaynak (2001) have convincingly criticized the construction of the traditional estimates of the capital share, and have provided revised estimates that – among other things – attempt to include the labor component of self-employment income in the labor share. These estimates are plotted in Figure 20.[68] Figure 20 shows essentially no systematic pattern of cross-country variation in capital shares.

[68]The numbers are from Table X in Bernanke and Gurkaynak (2001). We follow their advice and use the value in column "Actual OSPUE" whenever available; "Imputed OSPUE" when "Actual OSPUE" is unavailable but "Imputed OSPUE" is; and "LF" when the two OSPUE measuresd are unavailable. Of course Bernanke and Gurkaynak (2001) are reporting labor income, so our measure is 1 minus the numbers in the table.

This supports our preliminary finding: the efficiency of capital is higher in poor countries, and the efficiency of (quality adjusted) labor is higher in rich ones!

Unfortunately, the data set on capital shares is small – only 54 observations – and developed economies are over-represented. Furthermore, many untested assumptions have been used to develop these estimates. Hence, the conclusion that capital shares are not systematically related to labor productivity is not iron tight. What would it take then to reverse the startling result that poor countries are more efficient users of capital?

If factor shares vary systematically with per-worker income, then it becomes critical to know what is the elasticity of substitution $\eta = 1/(1-\sigma)$. Suppose that S_k is higher in rich countries. If $\sigma > 0$ (i.e. $\eta > 1$, or capital and human capital are good substitutes relative to the Cobb-Douglas case), then A_k may conceivably become increasing in income [if $(S_k)^{1/\sigma}$ grows "faster" than y/k falls]. In this case, however, since $S_h = 1 - S_k$ the result on A_h could also possibly be overturned. If $\sigma < 0$ (or $\eta < 1$) the results from the constant-share case would be reinforced. Symmetrically, if S_k is decreasing in income, the negative (positive) correlation between A_k (A_h) and y would be reinforced for $\sigma > 0$ ($\eta > 1$), and weakened (and possibly overturned) if $\sigma < 0$ ($\eta < 1$). These observations are summarized in Table 6. Each cell of the table lists the predicted sign (positive, negative, or ambiguous) for the correlation between A_k and y (first term) and between A_h and y (second term), conditional on the observed patterns of y/k and y/h, under various assumptions on σ, and on the correlation between S_k and y.

Table 6: Predicted Correlations between A_k and y, and A_h and y

	$Corr(S_k, y) > 0$	$Corr(S_k, y) = 0$	$Corr(S_k, y) < 0$
$\sigma > 0$ ($\eta > 1$)	?,?	$-, +$	$-, +$
$\sigma < 0$ ($\eta < 1$)	$-, +$	$-, +$	?,?

The intuition for the way observed factor shares modify our predictions on cross-country efficiency patterns is simple. If $\sigma > 0$ the two factors are good substitutes. Because the two factors are good substitutes, it makes sense to try to increase the usage of the most efficient factor. Hence, when $\sigma > 0$ demand will concentrate on the factor with high efficiency, leading to a high share in income for this factor. Conversely, then, with $\sigma > 0$, when we observe a high income share for factor x we can infer that this factor is efficient. On the other hand, if $\sigma < 0$ the two factors are poor substitutes. In this case, allocative efficiency calls for boosting the overall efficiency units provided by the low-efficiency factor. This increases the income share of this factor. Hence, with $\sigma < 0$, a high income share for factor x signals that this factor is used inefficiently.

In sum, skepticism about the greater capital efficiency of poor countries is authorized if one believes that there is a strong positive correlation between S_k and income *and* $\eta > 1$; *or* if one believes that there is a strong negative correlation between S_k and y *and* $\eta < 1$. We have seen what the data say about S_k (no correlation): what about η?

Hamermesh (1986) provides an exhaustive survey, featuring firm, industry, and country-level studies, both cross-sectional and time series. Unfortunately, he reports a dismayingly wide range of estimates, both greater and less than one. To my knowledge, additional recent contributions have not helped narrowing down the region in which η may fall.[69] Since published estimates of η are neither stable, nor reliable, one could, perhaps, turn to theoretical considerations. There is of course a tradition of arguing that long-run elasticities are higher than short-run ones, and macro-economic higher than micro-economic. Ventura (1997) is a particularly convincing recent example. For our purposes it clearly seems appropriate to focus on a long-run, aggregate interpretation of the elasticity. However, it is not clear that even in this case these arguments put a lower bound on η: even accepting that it is higher than a microeconomic, short-run elasticity, does not necessarily imply that it is, say, greater than 1.

For the countries with available data, we can actually compute the implied values of A_k and A_s from equation (19) for different values of the elasticity of substitution η. For each of these implied set of estimates, in Table 7 we report the coefficients of regressions of $\log(A_k)$ and $\log(A_h)$ on $\log(y)$ (t-statistics in parenthesis).[70] The results from the table confirm that the available data is consistent with the situation in the middle column of Table 6: irrespective of the value of η, rich countries seem to use capital less efficiently than poor ones. However, they use human capital more efficiently. These relationships are almost always statistically significant, except for $\eta = 0.9$. This is not surprising, since $\eta = 0.9$ is "almost" Cobb-Douglas, and in this limiting case A_k and A_h are not independently identifiable. The

[69] Aside from the huge dispersion in existing estimates of η, the non-neutrality approach we follow here points to an intrinsic pitfall in attempting to identifying this parameter. Specifically, empirical investigations of the elasticity of substitution implicitly assume that there is no variation across observations in the relative efficiency of labor and capital. If A_k and A_h vary across observations, then the effective input $A_k k$ and $A_h h$ will be mis-measured, perhaps wildly. I believe this may indeed be the reason why estimates of η are so unstable. I think this point is implicit in the analysis of Diamond, McFadden, and Rodriguez (1978). If the induced measurement error is random, it seems the bias in the estimate of η should be upwards. Intuitively, observations with very different input combinatons will appear to have similar output levels, something that is consistent with a high elasticity of substitution. However, if the As vary systematically, the bias could also be downward. Suppose, for example, that A_x and x are positively correlated across observations. Then the data will tend to understate the true variation in effective input, so that less substitutability will appear to be required to explain the observed variation in output.

[70] In these regressions there are 52 observations for $\log(A_k)$ and 49 for $\log(A_h)$. Also notice that, from equations 19, the As are identified up to the common multiplicative constants $\alpha^{1/\sigma}$ and $(1-\alpha)^{1/\sigma}$.

Table 7: Regressions of $log(A_k)$ and $log(A_h)$ on $log(y)$

Dep. Var.	$\eta = .1$	$\eta = .5$	$\eta = .9$	$\eta = 1.1$	$\eta = 1.5$	$\eta = 2$	$\eta = 50$
$\log(A_k)$	-.32	-.27	-.15	-.89	-.48	-.43	-.37
	(5.98)	(4.01)	(.39)	(1.99)	(3.62)	(4.44)	(5.67)
$\log(A_s)$	.80	.74	.20	1.55	1.01	.95	.88
	(28.41)	(17.56)	(.81)	(5.33)	(12.72)	(17.18)	(25.47)

coefficients are also sizable, with a 10 percent increase in income per worker being associated with up to a 9 percent decline in A_k, and even larger increases in A_h.

Everything considered, the result that poor countries use capital more efficiently than rich ones seems surprisingly robust. Needless to say, it is also rather stunning – especially if one is used to think about the world in TFP (factor neutral) terms. It is reminiscent of the findings of Caselli and Coleman (2003), that poor countries use unskilled labor more efficiently. In that paper, a possible explanation is constructed around the idea of appropriate technology. Perhaps countries face a menu of technology choices - some of which are more complementary with some of their endowments than others. It may be possible to construct a similar explanation for the facts documented in this section. In that explanation, rich countries would choose technologies that are complementary with human capital, while poor countries choose technologies that are complementary with physical capital.

7.2 Development Accounting with Non-Neutral Differences

Development accounting asks how the observed distribution of GDP per worker compares to the distribution that would obtain if all countries had the same technology. As is clear from equation (17), we are again in the situation in which – unlike in the simple TFP framework – the answer to this question will not be insensitive to which particular pair of values of A_k and A_h are plugged in. As we did in the previous subsection, and for the same reasons, I choose as the benchmark country the USA. Hence, I first compute the US A_k and A_h from equations (19), and then I plug these numbers in equation (17) for each country, and obtain measures of success of a model where all countries use the same (US) technology.[71]

Figure 21 plots the success of the factor-only model when all countries use the technology of the United States. Note that our measures of success converge – as they should – to those of the factor-neutral (Cobb-Douglas) model for η converging to 1. Success of the factor-

[71] More precisely, I compute $\alpha A_k^\sigma = S_{k,US}(y_{US}/k_{US})^\sigma$, and $(1-\alpha)A_h^\sigma = S_{h,US}(y_{US}/h_{US})^\sigma$, and plug these expressions in equation (17).

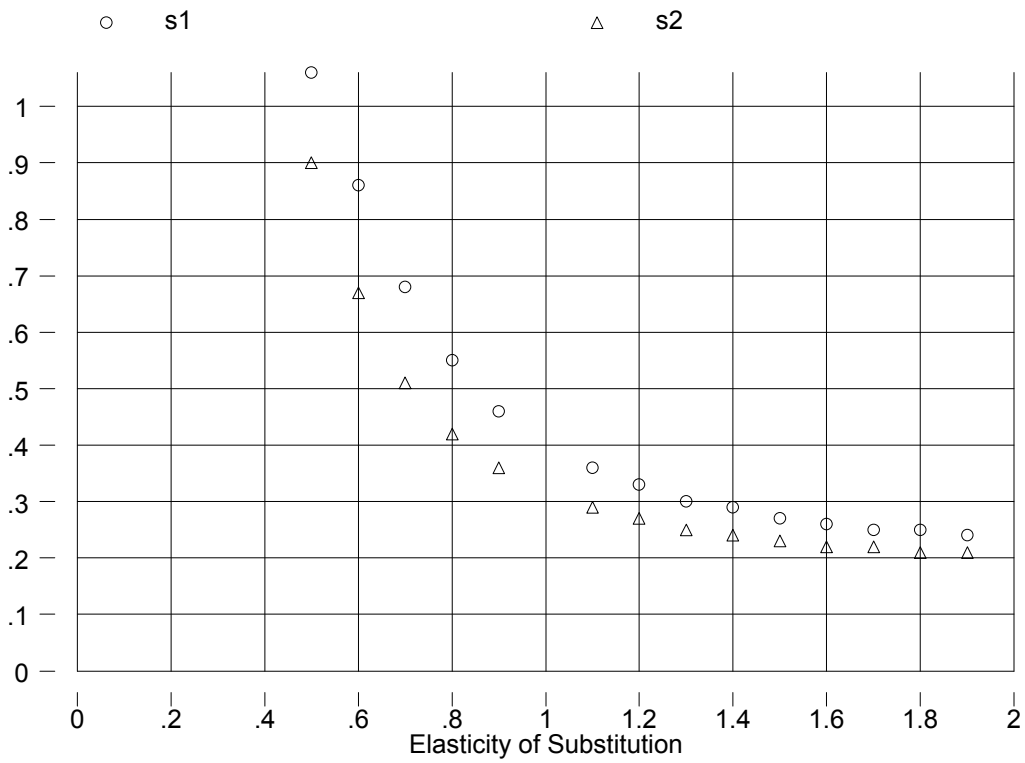

Figure 21: Success with Non-Neutral Technlogy Differences

only model is also systematically decreasing with the value of the elasticity of substitution, η. To see why this is so recall that the lower is η the closer the production function becomes to the Leontief case. Also notice that Figures 18 and 19 imply that poor countries are relatively abundant in human capital: the ratio h/k is decreasing with per-capita income. Finally, recall that the US uses human-capital relatively efficiently. Hence, the ratio $A_{h,US}h/A_{k,US}k$ is extremely high in poor countries, suggesting that much "effective human capital" in these countries would be unproductive, if not entirely wasted, in the limiting Leontief case. This "waste" may explain the low GDP of poor countries without having to invoke low As, i.e. it leads to greater success of the factor-only model. On the other side of $\eta = 1$, we tend to approach the linear production function. The closer we are to this case, the less a disproportionate $A_h h/A_k k$ ratio hurts a country's productivity, so the factor-only model performs less and less well. Something else is required to explain why GDP is so low.

The most remarkable finding of Figure 21, however, is quantitative: namely, not only for elasticities of substitution less than 1 does the model with non neutral technology outperforms the Cobb-Douglas one, but there is a range of elasticities such that the performance

of the model is extremely good. Indeed, an elasticity of 0.5 delivers a perfect fit for the factor-only model! This is remarkable because an elasticity of 0.5 – given our current state of knowledge – is not particularly implausible.[72]

One could object to the exercise we just reported that the counter-factual choice of technology – all countries using the US technology – is not sensible. For, we know from the previous section that – assuming our factor share data to be dependable – countries are observed to use technologies that cannot be ranked: some technologies boost the productivity of physical capital, some others of human capital. Therefore, one may speculate that – given the choice – poor countries would not necessarily choose to use the US technology. Instead, they would choose the technology most appropriate given their factor endowments.

In order to address this point, we now treat the observed (A_k, A_h) combinations from the previous sub-section as a "menu" of available technologies. One could think of this menu as a summary of the world's technical knowledge, each observed (A_k, A_h) pair being a particular blueprint to generate output from physical and human capital. I then re-interpret the counter-factual of no technology differences as one where all countries *have access* to the same menu of technologies; ask what technology from this menu would each country choose; and compute the counter-factual world income distribution when all countries choose their appropriate technology (from the set of available ones). The appropriate technology is the output-maximizing one.[73]

The results from this alternative counter-factual experiment are plotted in Figure 22. Here, for each value of η, I have computed from equations (19) the implied values of A_k and A_h for each country with complete data – including data on the factor shares S_k and S_h. This gave us 49 "observed" technologies. For each country in our 93-country sample, then, I have "chosen" from this menu of 49 the appropriate technology, and I have computed success under the assumption that each country uses this output-maximizing choice. As can be seen, the results preserve the broad qualitative features of those of Figure 21, but quantitatively the factor-only model does much less well. In particular, (almost) complete success only occurs if the technology is Leontief.

The intuition for this change in results is simple. Given the observed wide disparity in factor proportions, when countries can choose their technology appropriately they will in general choose different (A_k, A_h) combinations. In particular, few poor countries will find it optimal to use the technology observed in the USA. Indeed, countries with unfavorable factor endowments will be able to partially remedy by choosing technology appropriately. Hence,

[72] Elasticities less than 0.5 imply an income dispersion under the common-technology assumption that are even greater than the observed ones.

[73] See Caselli and Coleman (2004) for a model where the technology choice is decentralized at the level of firms, and the equilibrium aggregate technology is the GDP-maximizing one.

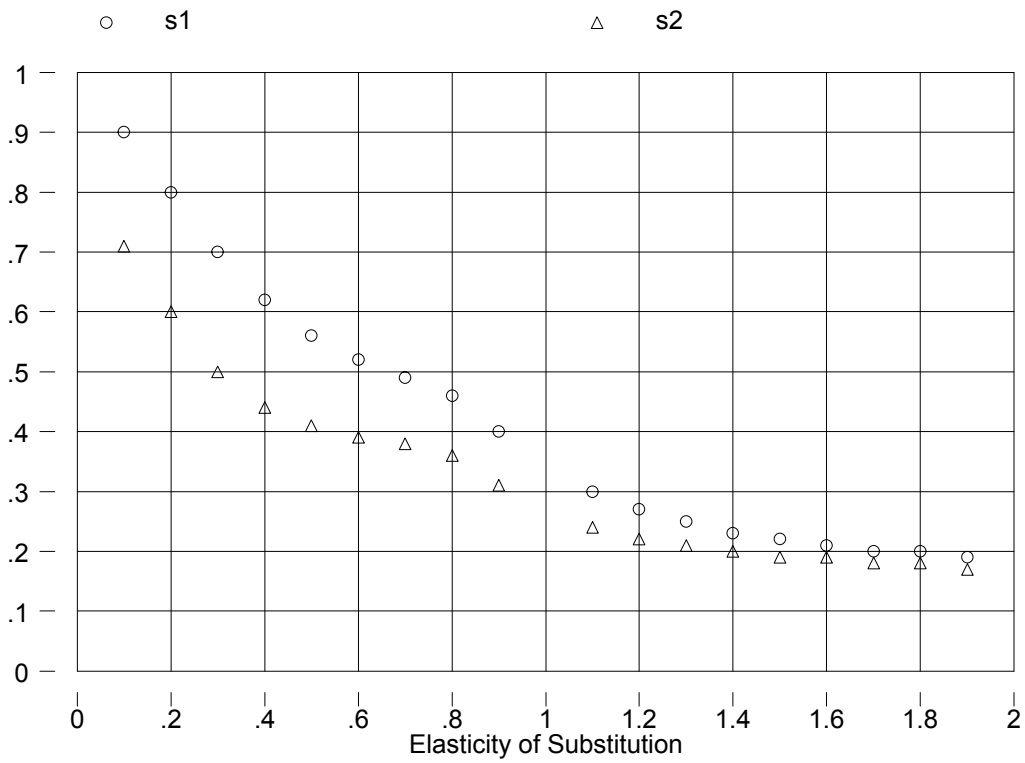

Figure 22: Success with Non-Neutral but Appropriate Technology

factor endowments do not have as much explanatory power for income differences as they do when all countries are forced to use the same (US) technology.

Even when countries choose technology, however, the measures of success are sufficiently sensitive to the choice of the elasticity of substitution that the answer to the development-accounting question can change substantially. It is therefore appropriate to conclude that gathering more information on this elasticity is a high priority for development accounting.

8 Conclusions

Development accounting is a powerful tool to getting started thinking about the sources of income differences across countries. As of now, the answer to the development-accounting question – do observed differences in the factors employed in production explain most of the cross-country variation in income – is: no, way no. This negative answer is robust to attempts to improve the measurement of human capital by allowing for differences in the

quality of schooling and in health status of the population; to attempts to account for the age composition of the capital stock; to sectorial disaggregations of output; and to several other robustness checks.

On the other hand, incomplete knowledge about certain key parameters that describe the relationship between inputs and outputs implies that the jury should be treated as being still out. For one thing, depending on the elasticity of substitution between capital of different types, the observed wild heterogeneity in the composition of the capital stock by type of equipment could turn out to be a key proximate determinant of income differences. For another, depending on the elasticity of substitution between physical and human capital, we may find that all is needed for these factors to explain a large fraction of income inequality is a departure from Cobb-Douglas. Disaggregating the government sector out of the data may also potentially reduce the unexplained component of GDP. There is no deep reason why we should not be able to make progress on these three fronts, so that my assessment of the future of this research enterprise is optimistic.

References

Acemoglu, D., and J. Angrist (2000), "How Large Are Human Capital Externalities? Evidence from Compulsory Schooling Laws", in: B.S. Bernanke, and K. Rogoff, eds., NBER Macroeconomics Annual 2000 (MIT Press, Cambridge) 9-59.

Aiyar, S., and C-J. Dalgaard (2002), "Total Factor Productivity Revisited: A Dual Approach to Levels-Accounting", Working Paper (Brown University).

Altonji, J.G., and T.A. Dunn (1996), "The Effects of Family Characteristics on the Return to Education", The Review of Economics and Statistics 78(4):692-704.

Baily, M.N., and R.M. Solow (2001), "International Productivity Comparisons Built from the Firm Level" Journal of Economic Perspectives 15(3):151-172.

Barro, R.J. and J. Lee (2001): "International Data on Educational Attainment: Updates and Implications," *Oxford Economic Papers*, 53 (3), 541-563 (July).

Behrman, J.R. and M. R. Rosenzweig (2004): "Returns to Birthweight," Review of Economics and Statistics, 86(2): 586-601.

Bernanke, B.S., and R.S. Gurkaynak (2001), "Is Growth Exogenous? Taking Mankiw, Romer, and Weil Seriously", in: B.S. Bernanke, and Kenneth Rogoff, eds, NBER Macroeconomics Annual 2001 (MIT Press, Cambridge)

Bils, M., and Klenow, P.J. (2000), "Does Schooling Cause Growth", The American Economic Review 90(5):1160-1183.

Card, D. (1999), "The Causal Effect of Education on Earnings", in: Ashenfelter, O., and Card, D., eds., Handbook of Labor Economics Volume 3A., (North-Holland, Amsterdam) 1801-1863.

Card, D., and A. Krueger (1996), "Labor Market Effects of School Quality: Theory and Evidence", in: Burtless, G. ed., Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success (Brookings Institute, Wasinghton D. C.).

Caselli, F., and W.J. Coleman (2001a), "Cross-Country Technology Diffusion: The Case of Computers", American Economic Review 91(2):328-335.

Caselli, F., and W.J. Coleman (2001b), "The U.S. Structural Transformation and Regional Convergence: A Reinterpretation", Journal of Political Economy 109(3):584-616.

Caselli, F., and W.J. Coleman (2004), "The World Technology Frontier", Working Paper (Harvard University).

Caselli, F., G. Esquivel, and F. Lefort (1996), "Reopening the Convergence Debate: A New Look at Cross-Country Growth Empirics", Journal of Economic Growth 1(3):363-389.

Caselli, F. and P. Malhotra (2004): "Natural Disasters and Growth: from Thought Experiment to Natural Experiment," working paper, Harvard University.

Caselli, F. and S. Tenreyro (2004), "Is Poland the Next Spain?," working paper,

Harvard University.

Caselli, F., and D. Wilson (2004), "Importing Technology", Journal of Monetary Economics, 51(1), 1-32.

Chanda, A., and C.-J. Dalgaard (2003) "Dual Economies and International TFP Differences", Working Paper (Louisiana State University).

Christensen, L.R., D. Cummings, and D.W. Jorgenson (1981), "Relative Productivity Levels, 1947-1973: An International Comparison", European Economic Review 16(1):61-94.

Clark, C. (1940), "The Morphology of Economic Growth", in: C. Clark (ed.): The Conditions of Economic Progress (Macmillan, London) 337-373.

Clark, G. (1987), "Why Isn't the Whole World Developed? Lessons From the Cotton Mills", Journal of Economic History 47:141-173.

Conrad, K., and D.W. Jorgenson (1985), "Sectoral Productivity Gaps between the United States, Japan and Germany, 1960-1979" in: H. Giersch, ed., Probleme und Perspektiven der weltwirtschaftlichen Entwicklung (Duncker and Humblot, Berlin) 335-347.

Currie, J., and D. Thomas (1999),"Early test scores, socioeconomic status and future outcomes", Working Paper No. 6943 (NBER).

de la Fuente, A., and R. Doménech (2002), "Human Capital in Growth Regressions: How Much Difference Does Data Quality Make? An Update and Further Results", CEPR Discussion Papers No. 3587.

Denison, E.F. (1967), Why Growth Rates Differ (The Brookings Institutions, Washington, D.C.).

Diamond, Peter, Daniel McFadden and Miguel Rodriguez (1978), "Measurement of the Elasticity of Factor Substitution and Bias of Technical Change", in Melvyn Fuss and Daniel McFadden (eds.), Production Economics: A Dual Approach to Theory and Applications, Vol. II, Applications of the Theory of Production, Amsterdam, North-Holland Publishing Company.

Duflo, E. (2001), "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment", American Economic Review 91(4):795-813.

Easterly, W., and R. Levine (2001), "It's Not Factor Accumulation: Stylized Facts and Growth Models", World Bank Economic Review 15(2):177-219.

Eaton, J., and S. Kortum (2001), "Trade in Capital Goods", European Economic Review 45(7):1195-1235.

Feyrer, James (2002): "Demographics and Productivity," working paper, Dartmouth College.

Fisher, A. (1945), Economic Programs and Social Security (Macmillan, London).

Gelb, A., Knight J.B., and R.H. Sabot (1988), "Lewis through a looking glass: Public sector employment, rent seeking, and economic growth," World Bank.

Gollin, D. (2002), "Getting Income Shares Right", Journal of Political Economy 110(2):458-474.

Gollin, D., Parente S.L., and Rogerson R. (2000), "Farm Work, Home Work and International Productivity Differences", Working Paper.

Gollin, D., S.L. Parente, and R. Rogerson (2001), "Structural Transformation and Cross-Country Income Differences", Working Paper (University of California in Los Angeles).

Gourinchas, P-O., and O. Jeanne (2003), "The Elusive Gains from International Financial Integration", Working Paper No. 9684 (NBER).

Graham, B.S., and J. Temple (2001), "Rich nations, poor nations: how much can multiple equilibra explain?", Working Paper No. 76 (Center for International Development , Harvard University).

Greenwood, Jeremy; Hercowitz, Zvi; Krusell, Per. Long-Run Implications of Investment-Specific Technological Change. American Economic Review. Vol. 87 (3). p 342-62. June 1997.

Hall, R.E., and C.I. Jones, (1999), "Why do some countries produce so much more output per worker than others?", The Quarterly Journal of Economics 114(1):83-116.

Hamermesh, D. S. (1986), "The Demand for Labor in the Long Run", in: Handbook of Labor Economics, Vol 1, O. Ashenfelter and R. Layard, eds. (North-Holland Press, Amsterdam).

Hanushek, E.A. (2003), "The Failure of Input Based Schooling Policies", Economic Journal 113(485):F64-98.

Hanushek E.A., and D.D. Kimko (2000), "Schooling, Labor-Force Quality, and the Growth of Nations", The American Economic Review 90(5):1184-1208.

Hanushek, E.A. (2004): "Some Simple Analytics of School Quality," NBER WP 10229.

Hanushek, E.A., S.G. Rivkin, and L.L. Taylor (1996), "Aggregation and the Estimated Effects of School Resources",The Review of Economics and Statistics 78(4):611-627.

Harrigan, J. (1997), "Technology, Factor Supplies, and International Specialization: Estimating the Neoclassical Model", American Economic Review 87(4):475-494.

Harrigan, J. (1999), "Estimation of Cross-Country Differences in Aggregate Production Functions." Journal of International Economics, 47, 267-93.

Harris, J. R. and Todaro, M. (1970). Migration, unemployment and development: a two-sector analysis. American Economic Review, 60, 126-142.

Heckman, J.J., and P.J. Klenow (1997) "Human Capital Policy" Working Paper (Uni-

versity of Chicago).

Heston, Alan; Robert Summers, and Bettina Aten (2002): *Penn World Tables Version 6.1.* Downloadable dataset. Center for International Comparisons at the University of Pennsylvania.

Hill, R. (2000), "Measuring Substitution Bias in International Comparisons Based on Additive Purchasing Power Parity Methods", European Economic Review 44(1):45-162.

Hsieh, C.-T. (2000), "Measuring Biased Technology", Working Paper (Princeton University).

Hulten, Charles R. (1996): "Infrastructure Capital and Economic Growth: How Well You Use It May Be More Important Than How Much You Have." National Bureau of Economic Research, Inc, NBER Working Papers: 5847.

Hsieh, C-T., and P.J. Klenow (2003), "Relative Prices and Relative Prosperity", Working Paper No. 9701 (NBER).

Islam, N. (1995), "Growth Empirics: A Panel Data Approach", The Quarterly Journal of Economics 110:1127-1170.

Johnson, G.E., and F.P. Stafford (1973), "Social Returns to Quantity and Quality of Schooling", Journal of Human Resources 8(2):139-155.

Jorgenson, D. W. (1961), "The Development of a Dual Economy", Economic Journal (71):309-334.

Jorgenson, D.W. (1995a): "Productivity. Volume 1: Postwar U.S. Economic Growth." MIT Press.

Jorgenson, D.W. (1995b): "Productivity. Volume 2: International Comparisons of Economic Growth." MIT Press.

Jorgenson, D.W., and F.M. Gollop (1992), "Productivity Growth in U.S. Agriculture: A Postwar Perspective", American Journal of Agricultural Economics 74 (3):745-750.

Jovanovic and Rob (1997): "Solow v. Solow: Machine Prices and Development," unpublished, New York University.

King, R.G., and R. Levine (1994), "Capital Fundamentalism, Economic Development, and Economic Growth", Carnegie-Rochester Conference Series on Public Policy 40(0):259-292.

Klenow, P.J., and A. Rodriguez-Clare (1997), "The Neoclassical Revival in Growth Economics: Has It Gone Too Far?", in: B.S. Bernanke, and J.J. Rotemberg, eds., NBER macroeconomics annual 1997 (MIT Press, Cambridge) 73-103.

Knight, M., N. Loyaza, and D. Villanueva (1993), "Testing the Neoclassical Theory of Economic Growth", IMF Staff Papers 40(3):512-541.

Krueger, A.B. (2000), "Economic Considerations and Class Size", Working Paper No.

8875 (NBER).

Kuznets, S. (1966), Modern economic growth (Yale University Press, New Haven).

Lee, J-W., and R.J. Barro (2001), "Schooling Quality in a Cross-Section of Countries", Economica 68:465-488.

Lewis, W.A. (1954), "Economic Development with Unlimited Supplies of Labor", Manchester School of Economics and Social Studies 22:139-151.

Lutz, H. (2002), "How Important Is Human Capital for Development? Evidence from Immigrant Earnings", American Economic Review 92(1):198-219.

Mankiw, N.G., D. Romer, and D.N. Weil (1992), "A Contribution to the Empirics of Economic Growth", The Quarterly Journal of Economics 107(2):407-437.

Mateos-Planas, X. Schooling and Distortions in a Vintage Capital Model. University of Southampton - Department of Economics, University of Southampton - Department of Economics. p 40 pages. 2000.

McGrattan, E.R., and J.A. Schmitz, Jr. (1999), "Explaining Cross-Country Income Differences", in: Taylor, J.B., and Woodford, M., eds., Handbook of Macroeconomics Volume 1A., (North-Holland, Amsterdam) 669-737.

Murnane, R.J., J.B. Willett, and F. Levy (1995), "The Growing Importance of Cognitive Skills in Wage Determination", The Review of Economics and Statistics 77(2):251-266.

Murphy, K., Shleifer, A., and R. Vishny (1991): "The Allocation of Talent: Implications for Growth," Quarterly Journal of Economics, 106 (2): 503-530.

Neal, D.A., and W. R. Johnson (1996), "The Role of Premarket Factors in Black-White Wage Differences", Journal of Political Economy 104(5):869-895.

Nurske, R. (1953), Problems of Capital Formation in Underdeveloped Countries (Oxford University Press, New York).

Parente, Stephen L. (2000): Learning-by-Using and the Switch to Better Machines. Review of Economic Dynamics. Vol. 3 (4). p 675-703. October.

Parente, S.L., R. Rogerson, and R. Wright (2000), "Homework in Development Economics: Household Production and the Wealth of Nations", Journal of Political Economy 108(4):680-687.

Rao, P. D.S. (1993), Intercountry comparisons of agricultural output and productivity, Fao Economic and Social Development Paper (FAO, Rome).

Prescott, E.C. (1998), "Needed: A Theory of Total Factor Productivity", International Economic Review 39(3):525-551.

Pritchett, L. (2000), "The Tyranny of Concepts: CUDIE (Cumulated, Depreciated, Investment Effort) Is Not Capital," Journal of Economic Growth 5(4): 361-84.

Pritchett, L. (2003), "Does learning to add up add up? The returns to schooling in

aggregate data," draft chapter for Handbook of Education Economics, Harvard University.

Psacharopulos, G. (1994), "Returns to investment in education: A global update", World Development 22(9):1325-1343.

Restuccia, D., D.T. Yang, and X. Zhu (2003), "Agriculture and Aggregate Productivity: A Quantitative Cross-Country Analysis", Working Paper (University of Toronto, Economics Department).

Rodriguez-Clare, A. (1996): "The Role of Trade in Technology Diffusion," Minneapolis Fed Working Paper 114.

Rostow, W. W. (1960), The Stages of Economic Growth: A Non-Communist Manifesto (Cambridge University Press, Cambridge, MA).

Schmitz, Jr., J.A. (2001), "What Determines Labor Productivity?: Lessons From the Dramatic Recovery of the U.S. and Canadian Iron-ore Industries", Staff Report 286, Federal Reserve Bank of Minneapolis.

Shastry, G.K., and D.N. Weil (2003), "How much of Cross-Country Income Variation is Explained by Health?", *Journal of the European Economic Association*, April-May 1(2-3), 387-396.

Solow (1959): "Investment and Technical Progress," in *Mathematical Methods in the Social Sciences,* (Arrow, Karlin, and Suppes, eds.), Stanford University Press.

Summers, R., and A. Heston, (1991), "The Penn World Table (Mark 5): An expanded set of international comparisons, 1950-1988", The Quarterly Journal of Economics 101(2):327-368.

Temple, J. (2003): "Dualism and Aggregate Productivity," unpublished, University of Bristol.

Ventura, J. (1997), "Growth and Interdependence", The Quarterly Journal of Economics 112(1):57-84.

Wagner, K., and B. van Ark, eds. (1996), International Productivity Differences: Measurement and Explanations (North-Holland, Amsterdam).

Weil, D.N. (2001), "Accounting for the Effect of Health on Economic Growth", Working Paper (Brown University).

Wilson, D.J. (2004), "Cross-Country Differences in Capital Composition", forthcoming in FRBSF Economic Letter.