Journal of Development Economics 108 (2014) 106–118

Contents lists available at ScienceDirect

Journal of Development Economics

journal homepage: www.elsevier.com/locate/devec

The efficiency of human capital allocations in developing countries

Dietrich Vollrath

University of Houston, 201C McElhinney Hall, Houston, TX 77204, United States

ARTICLE INFO

Article history:
Received 3 June 2013
Received in revised form 11 December 2013
Accepted 23 January 2014
Available online 6 February 2014

JEL classification:
O12
O15
O47
O57
J31

Keywords:
Misallocation
Human capital
Aggregate productivity
Structural transformation
Wage differentials

ABSTRACT

For a set of 14 developing countries I evaluate whether differences in wage gaps between sectors – estimated from individual-level wage data – have meaningful effects on aggregate productivity. Under the most generous assumptions regarding the homogeneity of human capital, my analysis shows that eliminating wedges between wages in different sectors leads to gains in output of less than 5% for most countries. These estimated gains of reallocation represent an upper bound as some of the observed differences in wages are due to unmeasured human capital. Under reasonable assumptions on the amount of unmeasured human capital the gains from reallocation fall well below 3%. Compared to similar estimates made using data from the U.S., developing countries would gain more from a reallocation of human capital, but the differences are too small to account for a meaningful portion of the gap in income per capita with the United States.

1. Introduction

A prominent feature of developing economies is the wide variation in labor productivity between different sectors. Gollin, Lagakos, and Waugh (2013) document that agricultural labor productivity is typically about one-half the level of non-agriculture, echoing the work of Lewis (1954) on dual economies. Moreover, this productivity gap appear appears to be most pronounced in the poorest countries, a fact noted by Kuznetz (1971), and explored further by Gollin et al. (2002), Caselli (2005), and Restuccia et al. (2008). Beyond just agriculture and non-agriculture, McMillan and Rodrik (2011) document that large differences in labor productivity exist across ten broad sectors within each country they study.[1]

If these disparities reflect real differences in the marginal productivity of labor between sectors, then aggregate productivity is lower than its potential, offering a partial explanation for low measured total factor productivity in developing countries. Several papers have attempted to estimate the loss in aggregate productivity from such misallocations. Focusing on only the distinction between agriculture and non-agriculture, Chanda and Dalgaard (2008), Vollrath (2009a), and Cordoba and Ripoll (2009) all suggest that there are substantial losses due to misallocations, while Caselli (2005) and Graham and Temple (2006) find much smaller effects.[2]

The existing literature infers differences in the marginal product of labor and/or human capital between sectors based on aggregate level information on output and labor inputs. Gollin et al. (2013) are the most sophisticated in addressing the measurement issues arising from the use of this data. They employ census and survey data to adjust for differences in education and hours worked between sectors, as well as providing evidence that national accounts data provide a reasonable measure of agricultural value-added.[3] After their adjustments, they continue to find large gaps in human capital productivity between agriculture and non-agriculture.

E-mail address: devollrath@uh.edu.

[1] MacMillan and Rodrik's calculations also show that shifts of labor between sectors are an important part of economic growth, although the contribution is not necessarily positive. In Latin American and Sub-Saharan Africa sectoral shifts were actually a drag on growth in recent decades.

[2] These papers are part of a wider literature looking at variation in factor returns and the allocation of factors across different uses. Banarjee and Duflo (2005) discuss the wide variation in factor prices found within developing countries. Looking specifically at firms, Hsieh and Klenow (2009) find that misallocation of physical capital and labor across firms in India and China lowers output by around 30% relative to the United States. There is a growing literature on productivity differences across firms and their relationship with aggregate productivity. See Foster et al. (2001) and Syverson (2011) for overviews of this literature. Similarly, Restuccia and Rogerson (2008) find large effects of dispersion in firm's revenue productivity on aggregate productivity. Jones (2011) discusses how misallocations of factors of production can lead to lower productivity because of the input/output relationships between industries.

[3] Herrendorf and Schoellman (2011) find that measures of agricultural value-added in the United States are actually inaccurate given the accounting treatment of the income of farm proprietors.

http://dx.doi.org/10.1016/j.jdeveco.2014.01.009
0304-3878/

In this paper I examine the aggregate impact of labor misallocation from the "bottom-up", using individual level wage data to identify wedges between sectors. Given the observed wedges between sectors, I then calculate the hypothetical gain to aggregate productivity from eliminating those wedges. I do this for a set of 14 developing countries.[4] The data are derived from the Living Standard Measurement Surveys organized by the World Bank and are collated by the Rural Income Generating Activity Database produced by Davis et al. (2010). Despite the name the database contains information on urban workers as well as rural.

In the raw data there exist substantial wedges between sectors in the wage per day within almost every country I study. Fig. 1 shows the average wage in each sector relative to the average wage for each of the 15 surveys I use (there are two surveys for Nicaragua). The nine sectors shown for each country are a standard ISIC (revision 2) breakdown, with the Miscellaneous sector omitted. Agriculture (represented by the dark circles) tends to have wages well below the country average, dipping to only 50% of average wages in Ecuador and Nigeria. Sectors that have consistently high wages are Finance and business services (the open squares) and Utilities (the dark squares), with wages 1.5–2.5 times higher than average in each country.

Of course, much of the variation in wages between sectors in the countries in Fig. 1 reflects human capital differences, rather than differences in the wage paid per unit of human capital. I will explain the nature of the data and the precise estimation below, but Fig. 2 plots the residual wage per day in each sector after I have removed the influence of human capital using a simple Mincerian regression for each country that includes education, age, and occupation. As can be seen, there is much less variation in Fig. 2 within each country.

Regardless, there is still noticeable variation. In Ghana, for example, the wage paid to a unit of human capital in mining is 2.5 times the average across all workers in that country. In Nigeria and Ecuador the wage rate in agriculture is still roughly 50% of the average wage. Tajikistan exhibits several sectors with wages nearly 2 times the national average.

Do the wage differences in Figs. 1 and 2 imply a significant aggregate productivity loss within developing economies? An answer to that question requires finding the hypothetical productivity level when wedges between sector-level wages are eliminated. For that calculation I start by making several assumptions designed to maximize the productivity gain from reallocation. First, I assume that all units of human capital within a country are perfect substitutes. This implies there is no loss of productivity for a unit of human capital when it shifts sectors. Secondly, I assume that there is no unmeasured human capital. This implies that the wage wedges seen in the figures reflect real differences in wages, rather than differences in unmeasured skills between sectors.

Despite these very strong assumptions, my calculations show that productivity after removing the wage wedges would rise by less than 5% for 11 of the 14 countries. For 13 of the 14, the productivity gain is less than 11%, and the gain only reaches 15% in Tajikistan. For comparison purposes, the implied productivity gain from reallocation across sectors in the United States, calculated using data from the Current Population Survey in 2000, is approximately 1.8%. While the U.S. appears to have a more efficient allocation, the difference between it and the developing countries is not terribly large. Moreover, the misallocation of human capital does not explain much of the gap in income per capita between these countries and the United States. As an example, income per capita in the U.S. in 2004 was about 22 times that in Nigeria. If wage wedges were eliminated between sectors in Nigeria, that ratio would only fall to about 19.5. For other countries the explanatory power of misallocation across sectors is even smaller.

[4] The countries are Albania, Bangladesh, Bulgaria, Ecuador, Ghana, Guatemala, Indonesia, Malawi, Nepal, Nicaragua, Nigeria, Panama, Tajikistan, and Vietnam. The surveys from Ecuador, Nicaragua, and Vietnam are from the late 1990s, and the rest are from the 2000s.

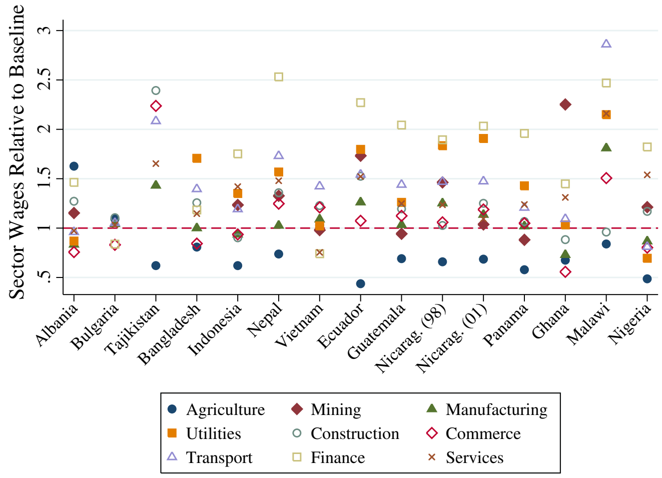

Fig. 1. Average earnings per worker per day, by sector, by country. Note: The figure shows, for each country, the average wage per day in a sector relative to the average wage for the entire country $(1+\tau_j^W)$, without any controls for human capital. Author's calculations using the RIGA database of Davis et al. (2010). See text for details.

In comparison to the prior literature, the implied productivity gains I find here are quite small. Hsieh and Klenow (2009) report gains to manufacturing TFP of around 100% for China and India from reallocating factors between firms. Vollrath (2009a) finds gains on the order of 150% for some sub-Saharan African countries from reallocating labor between agriculture and non-agriculture. The gains I find here do not necessarily contradict those findings. Those papers use aggregate data and assumptions about production functions to identify (and then remove) marginal product wedges between sectors or firms. In this paper I am looking only at *wage* wedges between sectors. It is quite possible that while wage gaps (and the gains from eliminating them) may be small, marginal product gaps (and the gains from eliminating them) could still be large. This would be the case if there was an additional wedge between the marginal product of labor and the wage *within* each sector. Eliminating wage gaps is thus a subset of the total productivity gains available from eliminating marginal product gaps.

Furthermore, I am calculating the *static* gains from reallocation. That is, I am holding physical capital and total factor productivity (TFP) constant in each sector in making my calculations, and hence the marginal product of human capital falls as more human capital flows into a sector.

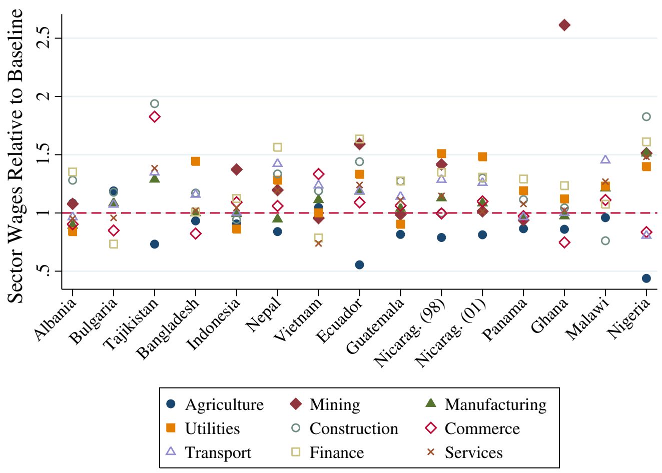

Fig. 2. Average earnings per worker per day controlling for human capital, by sector, by country. Note: The figure shows, for each country, the average wage per day in a sector relative to the average residual wage for the entire country $(1+\tau_j^W)$, after controlling for human capital characteristics (education, age, occupation, and occupation-specific education returns). Author's calculations using the RIGA database of Davis et al. (2010), see text for details.

In a *dynamic* setting, physical capital would accumulate in response to the more efficient allocation of human capital, slowing the fall of the marginal product. It is also possible that TFP would respond, perhaps through learning-by-doing or scale effects in some sectors. Hence the dynamic gains from reallocation could be quite large even if the static gains I have calculated remain relatively small. While I do not develop a full dynamic model here, if I parametrically limit the speed at which the marginal product of human capital falls, then gains from reallocation rise appreciably, in several cases reaching the size found in Hsieh and Klenow (2009).

Despite this, the static results are useful when searching for explanations of overall marginal product wedges between sectors. Frictions that lead to different rates of returns (e.g. wages) between sectors are unlikely to be important from an aggregate perspective. Rather, theories that allow for perfect factor markets but involve sorting across sectors based on unobserved skills (as in Lagakos and Waugh, 2013) or differences in time allocations to work (as in Vollrath, 2009b) appear more appropriate. Alternatively, theories that allow for distortions within sectors that drive a wedge between the wage and value marginal product of human capital (e.g. markups from imperfect competition) would be consistent with the findings of this paper.

To reinforce the finding that wage gaps do not imply large productivity losses, note that the gain from reallocating human capital across sectors is likely to be lower than my estimates suggest. I expand on my original estimates by incorporating the possibility that the variation within countries in Fig. 2 partly reflects unmeasured human capital rather than differences in wages between sectors. If half of the observed wage differences are due to unmeasured human capital, then the hypothetical gains from reallocation fall to below 5% for every country, and the explanatory power of misallocation for cross-country income differences falls as well.

One caveat is that the RIGA data contain information only on wage-earners, and exclude the self-employed – such as farmers – who make up a large portion of the labor force in many developing countries. However, in an extension I show that under a variety of assumptions about the level of human capital and earnings of self-employed workers, the implied gains to productivity remain relatively small. However, without more accurate data on self-employed individual earnings it is not possible to be conclusive.

To continue I first outline the method for calculating the hypothetical gain from reallocating human capital between sectors and how to use micro-level data to make that calculation. Following that I go over the data used in the estimations and report the variation in sector-level returns to human capital. With the estimates in hand I am able to calculate the potential gains from reallocation of labor across sectors and discuss the robustness of these results with respect to unmeasured human capital, self-employed workers, and parameter values.

2. The gains from reallocation

The earnings gaps observed in Figs. 1 and 2 suggest the possibility that human capital is misallocated and aggregate productivity is lower than it might otherwise be. In what follows I will describe my method for estimating the hypothetical gain to aggregate productivity from reallocating human capital efficiently. To make this estimate I will be making several aggressive assumptions that favor finding a large gain from reallocation in a static setting. Hence my estimates should be seen as an upper bound on static gains to reallocation, and as will be seen later even these upper bounds are empirically small.

2.1. Finding the efficient output level

To assess the efficiency of human capital allocations, I want to compare the hypothetical output of an economy where human capital is optimally allocated to actual observed output. To calculate this hypothetical, I need a model of production, as in the firm-level studies of misallocation in Restuccia and Rogerson (2008) or Hsieh and Klenow (2009), or the sector-level studies of Vollrath (2009a), Cordoba and Ripoll (2009), and Graham and Temple (2006).

Let each sector j operate with a Cobb–Douglas technology of the form

$$Y_j = Z_j H_j^{1-\alpha} \tag{1}$$

where Z_j encompasses both total factor productivity and the capital stock. I will be concerned here only with human capital allocations, ignoring the allocation of physical capital. Further assume that each sector is operated to maximize profits

$$\pi_j = \left(1-\tau_j^R\right) p_j Y_j - w\left(1+\tau_j^W\right) H_j, \tag{2}$$

where p_j is the price of output of sector j. The term τ_j^R is a wedge that applies to sector j's revenues, and hence I'll refer to this as the *revenue-wedge*. The revenue-wedge is similar to that used in Restuccia and Rogerson (2008) and Hsieh and Klenow (2009). It represents some distortion in the revenue-productivity of sector j. In addition, there is a second wedge, $1+\tau_j^W$, that distorts the wage rate facing sector j away from the average wage w; I'll refer to this as the *wage-wedge*.[5]

Profit-maximization by sectors yields the following first-order condition

$$\frac{1-\tau_j^R}{1+\tau_j^W}(1-\alpha) p_j Z_j H_j^{-\alpha} = w. \tag{3}$$

The combination of wedges will skew the amount of human capital used by sector j relative to the undistorted amount. If either τ_j^R or τ_j^W rises (falls), human capital is becoming relatively expensive (cheap) for sector j, and hence the amount of human capital utilized falls (rises). This represents a misallocation as the revenue marginal product of human capital is not equal to the average wage.

I further assume that human capital is perfectly substitutable between sectors. This implies that total human capital is simply $H = \sum_j H_j$ in the economy. To find aggregate output, one additional assumption is necessary regarding prices. I will assume that they are fixed, equivalent to assuming that we are working with a small open economy where all sectors produce output that can be traded internationally. While stark, this assumption acts to maximize the hypothetical output attainable without the wedges, as discussed in Vollrath (2009a). I discuss this assumption and its influence on the results in more detail below. Observed output in this economy is then

$$Y^{obs} = \left(\sum_j \left(p_j Z_j\right)^{1/\alpha} \left[\frac{1-\tau_j^R}{1+\tau_j^W}\right]^{1/\alpha}\right)^{\alpha} H^{1-\alpha}, \tag{4}$$

which follows from using Eq. (3) for each sector to solve simultaneously for the shares H_j/H, and then summing up output across sectors.

Y^{obs} is actual output in the economy, but the two wedges mean that this is below the output-maximizing level. With this structure in place, though, it is possible to consider how much output would rise hypothetically if one or both wedges were removed. In this paper, I am going to examine the gain to output from eliminating the wage-wedges, τ_j^W, which gives a hypothetical output level of, Y^*,

$$Y^* = \left(\sum_j \Omega_j^{1/\alpha}\right)^{\alpha} H^{1-\alpha} \tag{5}$$

where $\Omega_j = (1-\tau_j^R)p_j Z_j$. It represents the productivity of sector j absent the wage-wedge (but with the revenue-wedges still in place). Given

[5] The wage-wedge τ_j^W here is not just an analogue of the second wedge used in Hsieh and Klenow (2009). If I expanded the set-up described here to include capital, then profits would be $\pi_j = (1-\tau_j^R)p_j Y_j - w(1+\tau_j^W)H_j - R(1+\tau_j^K)K_j$. The term τ_j^K is what Hsieh and Klenow use to generate differences in the capital/labor ratio between firms.

this hypothetical output level, I compare it to observed output to find the gain from reallocation:

$$R = \frac{Y^*}{Y^{obs}} = \frac{\left(\sum_j \Omega_j^{1/\alpha}\right)^\alpha}{\left(\sum_j \frac{\Omega_j^{1/\alpha}}{\left(1+\tau_j^W\right)^{1/\alpha}}\right)^\alpha}. \tag{6}$$

This provides a measure of the gain in aggregate productivity from eliminating between-sector differences in the *wage-wedge* only. The numerator is aggregate productivity without the wage-wedge, while the denominator is aggregate productivity with the wage-wedges included.

Setting the problem up with both revenue and wage-wedges allows me to distinguish between my measure of efficiency and those found in prior literature. Restuccia and Rogerson (2008) look at hypothetical gains based on U.S. data, Hsieh and Klenow (2009) look across manufacturing firms in China, India, and the U.S., while Vollrath (2009b) and Cordoba and Ripoll (2009) look across two sectors (agriculture and non-agriculture). In each case, the authors essentially measure the efficiency gain from eliminating *both* the revenue and the wage-wedge. These papers cannot distinguish these two wedges, as they have data on output (or revenues) and stocks of human capital (or total wage bills), but no data on wage rates specifically. In a similar spirit, Gollin et al. (2013) and McMillan and Rodrik (2011) document large gaps in the marginal product of human capital across sectors in developing countries; these gaps are made up of both revenue-wedges and wage-wedges, but there is no way to distinguish between the two given their data.

By using individual-level data I am able to measure wage rates, and so will be able to identify the wage-wedge specifically for each sector. However, I will not be able to separately identify the revenue-wedge for each sector. Hence the efficiency gains I find will be a subset of the total efficiency gains possible through a complete reallocation of factors. While this limits me in some respects, a persistent theme in development economics has been disparities in the returns paid to identical factors of production within countries (see Banarjee and Duflo, 2005 for a review), dating back to Arthur Lewis' dual economy work. This work is generally concerned with the kinds of failures in factor markets that must occur such that factor returns vary by sector. I am providing a measure of the aggregate loss from these failures in a specific market, the one for human capital.

As noted in the introduction, I will find the gains from eliminating wage-wedges, R, to be relatively small. This implies that whatever distortions exist in labor markets do not have large aggregate effects. This also implies that even if there are large aggregate losses from misallocation between sectors, as is found in the firm-level or sector-level studies, those losses are due to revenue-wedges and not wage-wedges. More simply, my findings indicate that large aggregate effects of misallocation appear to be driven by the wedge between marginal product and wage *within* sectors, as opposed to a wedge between wages *across* sectors.

Why are the gains so small? This outcome turns on the dispersion in the productivity terms, Ω_j. Referring back to Eq. (6), the numerator is $(\sum_j \Omega_j^{1/\alpha})^\alpha$, which captures productivity without the wage-wedges. This productivity term increases with the variance of Ω_j. In other words, it is better to have one sector with incredibly high productivity and lots of sectors with low productivity rather than having all sectors with moderate productivity levels. In Figs. 1 and 2 there is dispersion in *wages*, but that doesn't necessarily translate into dispersion in sector-level *productivity*. To see this, rearrange the first-order condition (3) to read

$$w\left(1+\tau_j^W\right) = \frac{(1-\alpha)\Omega_j}{H_j^\alpha}. \tag{7}$$

The dispersion in wages could indicate either dispersion in Ω_j and/or dispersion in H_j. What the data will show is that Ω_j does not actually vary much, and the wage-wedges observed are indicative of variation in the amount of human capital employed in sector j. With little variation in Ω_j, productivity without the wage-wedges is not very high, and hence the implied gain, R, is not very large.

The small gains I find are likely to be an upper bound, given the assumptions that go into calculating R. Human capital is presumed to be perfectly substitutable both within and between sectors, and so when the wage-wedges are removed human capital will flow immediately and costlessly until all human capital earns the same rate w. This suggests that workers will move out of sectors where they were relatively underpaid (and hence over-used) into sectors where they were overpaid (and hence under-used), and that they can immediately and productively be put to use. Agricultural workers going into finance, for example. If agricultural human capital was not a perfect substitute for finance human capital, then the potential gain from reallocation would be smaller. In the extreme, if agricultural human capital was useless in the finance sector, then even if we removed the wage wedges there would be no possible gain from reallocation of human capital, as no one would hire the agricultural workers.

Additionally, as noted previously I have assumed that sector output prices, p_j, are constant. As discussed in Vollrath (2009a) in the context of two sectors, allowing prices to be set endogenously would *lower* the gain from a reallocation of labor. To see this, consider the agricultural sector, where the wage-wedge is negative, labor is underpaid compared to the average, and so labor is over-used. When the wedge is removed, the cost of labor goes up for the agricultural sector, and it is now higher than the value marginal product of labor. Hence labor leaves the agricultural sector, which pushes up the value marginal product of labor. This will continue until the value marginal product of labor equals the wage. If the price of agricultural goods is endogenous, then as labor leaves the agricultural sector and output falls, the price of agricultural goods will rise relative to other goods. This raises the value marginal product of labor in agricultural as well. With endogenous prices, fewer workers need to leave agriculture for the value marginal product of labor to rise to equal the wage. There is a smaller reallocation of labor when prices are endogenous, and the implied gain from removing the wage-wedges is smaller. So by assuming that prices remain constant, I am keeping the implied gains from reallocation relatively large.

2.2. Empirical implementation

Given the set-up in the prior section, how do I utilize individual-level wage data to calculate the gains from reallocation, R? Rewrite Eq. (3) as

$$w\left(1+\tau_j^W\right) = (1-\alpha)\Omega_j H_j^{-\alpha}. \tag{8}$$

Individual-level earnings data will allow me to find values for both $1+\tau_j^W$ and H_j for each sector using data on observed earnings and human capital characteristics. Describe observed earnings for individual i working in sector j by

$$m_{ij} = w\left(1+\tau_j^W\right)h_{ij}, \tag{9}$$

where $w(1+\tau_j^W)$ is the wage earned in sector j, and h_{ij} is the human capital of individual i employed in sector j. Let the log of human capital for individual i in sector j be written as

$$\ln\ h_{ij} = X_i'\beta_X + \epsilon_{ij} \tag{10}$$

where X_i' is a vector of individual characteristics and β_X is a vector translating those characteristics into units of human capital. ϵ_{ij} is random variation in human capital unrelated to the other terms. Here I am assuming that there is no sector-specific human capital, an assumption that I will relax as a robustness check later in the paper.

Given this specification for human capital, I can write individual earnings as

$$\ln\ m_{ij} = \ln\ w + \ln\left(1+\tau_j^W\right) + X_i^{'}\beta + \epsilon_{ij} \tag{11}$$

where one can see that the wage-wedge is the only sector-specific term. This equation is amenable to estimation, taking the form of a typical Mincerian wage regression. The actual specification I take to the micro-level data is

$$\ln\ m_{ij} = \beta_0 + \delta_j + X_i^{'}\beta_X + \epsilon_{ij}. \tag{12}$$

δ_j is a sector dummy, and to implement this regression I exclude the J-th sector, so that implicitly the value of $\delta_J = 0$. Because of this, the constant term and the sector dummies have the following interpretations

$$\beta_0 = \ln\ w + \ln\left(1+\tau_J^W\right) \tag{13}$$

$$\delta_j = \ln\left(1+\tau_j^W\right) - \ln\left(1+\tau_J^W\right). \tag{14}$$

After performing the regression in Eq. (12), the estimated log wage in sector j is

$$\hat{w}\left(1+\hat{\tau}_j^W\right) = exp\left(\hat{\beta}_0 + \hat{\delta}_j\right). \tag{15}$$

The estimated value of log human capital for an individual i in sector j is then

$$\hat{h}_{ij} = X_i^{'}\hat{\beta}_X + \hat{\epsilon}_{ij}. \tag{16}$$

With these individual human capital values, and the assumption that human capital is perfectly substitutable within sectors, the estimated stock of human capital in sector j is

$$\hat{H}_j = \sum_{i \in j} \hat{h}_{ij}. \tag{17}$$

Similarly, the total human capital stock in the economy is

$$\hat{H} = \sum_{j=1}^{J} \hat{H}_j. \tag{18}$$

Re-arranging Eq. (8), I can get an estimate of the level of Ω_j for a sector as

$$\hat{\Omega}_j = \frac{\hat{w}\left(1+\hat{\tau}_j^W\right)}{1-\alpha}\hat{H}_j^{\alpha}. \tag{19}$$

Using the values of $\hat{\Omega}_j$ and $\hat{H}$, I can calculate an estimate for the hypothetical output after removing wage-wedges,

$$\hat{Y}^* = \left(\sum_j \hat{\Omega}_j^{1/\alpha}\right)^{\alpha}\hat{H}^{1-\alpha}. \tag{20}$$

Given this expression for Y^*, the estimated value of R can be found from

$$\hat{R} = \frac{\left(\sum_j \hat{\Omega}_j^{1/\alpha}\right)^{\alpha}}{\left(\sum_j \frac{\hat{\Omega}_j^{1/\alpha}}{\left(1+\hat{\tau}_j^W\right)^{1/\alpha}}\right)^{\alpha}}. \tag{21}$$

To find this estimate, I will need to run Mincerian regressions of the form in Eq. (12) to recover wage-wedges and human capital stocks, which will then allow me to find the productivity terms.

Before continuing, it is worth noting the possible effects of biased estimates. The value of R depends positively on the variation in $\hat{\delta}_j$ across sectors. If there is bias in the $\hat{\delta}_j$ estimates that increase the variation in these terms across sectors, then I will overestimate the productivity gain from reallocation. Similarly, biases that reduce the variation in the $\hat{\delta}_j$ terms across sectors will mean that I am underestimating productivity gains. While I cannot rule out the latter case entirely, it seems most likely that the former case holds, and I am overestimating the productivity gains. This would be the case if, for example, agricultural workers tended to have low levels of unmeasured human capital (as suggested by Lagakos and Waugh, 2013). I would improperly estimate that τ_{ag}^W was low compared to other sectors, implying a greater spread of wages across sectors than in fact exists, and overestimate the productivity gains of reallocation.

3. Estimating gains from reallocation

As discussed in the introduction, I will be using individual-level data from the RIGA database of Davis et al. (2010). This database combines separate Living Standards Measurement Surveys (LSMS) done under the guidance of the World Bank, providing comparable measures of earnings and hours worked across different countries. For the current purpose, the most important feature of the RIGA database is the availability of individual earnings information and human capital characteristics that are comparable across individuals. This highlights why a broader set of countries, as in Gollin et al. (2013), is not used in the analysis. To find the wage-wedges specifically, I require information on individual wages in addition to the typical human capital information. Census data is not sufficient because it typically does not include income data for individuals, and in the few situations where it does the income data contains both wage earnings as well as non-wage earnings.[6]

Here, I am specifically concerned with earnings in the form of wages, and the RIGA database separates labor income from other sources of income (rents, transfers, etc.). In addition, the database reports the wage, by job, for each individual if they are employed in more than one position. This ensures that the wages earned by an individual in different sectors can be distinguished from each other.[7] The wage data are for paid employment, and are reported net of taxes, so that I have a measure of the net return to labor for an individual.[8] RIGA does not, however, provide an equivalent measure of individual self-employment income. I will therefore only be able to test for significant sector-level differences in wage work, and will not be able to offer any direct evidence on the efficiency of labor allocations between wage work and self-employment. Following the main results, though, I consider various

[6] An additional country in the RIGA dataset, Madagascar, had individual-level data available. However, due to a large amount of missing data on individual characteristics, I chose to drop it from the analysis.

[7] While a very useful feature of the database, it turns out that there are very few individuals engaged in more than one wage-earning job. Less than 1% of workers in each survey report a second wage-earning job. For those that do, total earnings in the second job are generally less than 10% of earnings in the first. Because of the limited nature of these jobs, I simply drop them from the analysis.

[8] Gross pay would be relevant to a sector making the decision about how much labor to hire. Any differences in gross pay that exist beyond the differences in net pay that I measure will be implicitly picked up in the revenue-wedge.

assumptions regarding the earnings and human capital of self-employed workers and calculate how the value of R would change if they were included.

For each person engaged in wage work, the RIGA database provides a measure of daily wages. For many individuals, this is what they report directly, while for others this daily wage is imputed from data on monthly or weekly wage income as well as data on days worked in those same time periods. Daily wages are not adjusted for hours worked in a given day, as information on hours worked is not available for a large majority of the individuals. Hence there will remain some uncertainty in my results. In terms of the Mincerian regressions I am trying to estimate, this will create a bias in the sector-specific terms δ_j if hours worked per day differ systematically by sector. I may overestimate or underestimate the variation in δ_j across sectors depending on how correlated hours worked are with hourly wages. When I use U.S. data as a comparison, I find that the implied gains to reallocation are smaller when using hourly wages compared to weekly wages. This indicates that hours worked are positively related to hourly wages, and that I am likely overestimating the variation in δ_j across sectors.

Most importantly, the RIGA database harmonizes the various industry classifications used across the LSMS surveys into standard ISIC codes. This provides 10 broad sectors: (1) Agriculture, Forestry, and Fishing, (2) Mining, (3) Manufacturing, (4) Utilities, (5) Construction, (6) Commerce, (7) Transport, Communications, and Storage, (8) Finance and Real Estate, (9) Services, and (10) Miscellaneous.

In terms of measuring the characteristics influencing human capital, X_i, the RIGA database provides a wide range of useful information. I have information on years of education and age, as well as gender. These provide a minimal level of control for human capital. Beyond those, though, RIGA provides a classification of occupation for each job held. Hence I can distinguish individuals by their type of work, independent of sector. This provides a level of control for human capital that has not been available before in macro-level studies that have, at best, been able to account for differences in education.

The different occupations reported are based on the International Labor Organization's (ILO) International Standard Classification of Occupations (ISCO). There are ten major categories, in addition to an Other/Unknown category: (1) Legislators, Senior Officials, and Managers, (2) Professionals, (3) Technicians and Associate Professionals, (4) Clerks, (5) Service Workers and Shop and Market Sales Workers, (6) Skilled Agricultural and Fishery Workers, (7) Craft and Related Trade Workers, (8) Plant and Machine Operators and Assemblers, (9) Elementary Occupations, and (10) Armed Forces Occupations.

In Table 1, I've provided summary information on the 15 different surveys (from 14 different countries) used, as well as aggregate-level information about the countries. GDP per capita and the percent rural population are both from the World Development Indicators, with GDP measured at PPP in 2005 international dollars. Education is from Barro and Lee (2010), and represents years completed by those over 25. The surveys represent several ex-Communist countries (Albania, Bulgaria, and Tajikistan), as well as Asian developing nations (Bangladesh, Indonesia, Nepal, and Vietnam), Latin American (Ecuador, Guatemala, Nicaragua, and Panama), and Sub-Saharan Africa (Ghana, Malawi, and Nigeria). These countries range from middle income countries with per-capita GDP above \$5000 a year to below \$1000 as in Malawi and Nepal. Education and the percent rural correspond to these development levels in a manner that is consistent with broad development trends.

The next four columns show summary statistics from the individual country surveys. While there are as few as 1631 observations (Ghana), most countries have around 5000 observations to work with. The wage data is reported in local currency units, and so is not directly comparable across countries. However, I have calculated the coefficient of variation (CV) for daily wages in each sample. One can see relatively high variation in wages in Nigeria, Panama, and Albania. Compared to these countries, the Asian countries have relatively low variation, while Latin American variation is relatively large. One thing to note is that there doesn't appear to be a clear relationship between wage variation within the samples and overall development levels.

From the individual surveys, I've computed average years of education for workers, as well as the percent that report living in rural areas. Compared to the country level data, the RIGA surveys show relatively high education levels, and relatively low rural percentages. This is perhaps not surprising, as the surveys are restricted to wage workers, while the country level data covers the entire population. Regardless, there does not appear to be anything alarming about the comparison, and the individual level data from the RIGA database gives us a more accurate picture of those participating in the labor market.[9]

An important omission from the RIGA data is self-employment data at the individual level. To see how much of the labor force RIGA covers, I've separately pulled information from IPUMS International (Minnesota Population Minnesota Population Center, 2013) for as many of the 14countries as possible to find the percentage of all workers that report themselves as wage-workers. The results are in the final column of Table 1. For all the countries, only Nicaragua (2001) and Panama have more than 50% of their workers as wage-workers. The percentage averages roughly 30% for all countries. Hence RIGA is capturing only a portion of the workforce. After discussing my baseline results, I provide an extension that attempts to incorporate the self-employed into the calculation of R.

Regardless, Table 2 reports the average wage of each sector in a country relative to the overall average wage, giving an initial indication of the variation in sector-level wages. As can be seen, aside from Albania and Bulgaria, agricultural wages tend to be around 2/3 of the mean wage. This ratio is similar across Malawi, Nigeria, and Ghana, despite their very different fraction of workers employed in that sector. Compared to agriculture, nearly every other sector has an average wage higher than the overall mean. Thus the main variation in wages tends to be between agriculture on one hand, and the rest of the economy on the other. This is perhaps not surprising given the existing literature on dual economy effects related to the rural/urban or agriculture/non-agriculture difference.

3.1. Estimation of sector returns to labor

To proceed, recall that I will be using the following specification

$$\ln m_{ij} = \beta_0 + \delta_j + X_i^{'}\beta_X + u_{ij} \tag{22}$$

to obtain the estimates of $\hat{\delta}_j$ and $\hat{h}_{ij}$ that are used in the calculation of the gain from reallocation, R. As noted above, for now I am assuming that there is no difference in the level of unmeasured human capital across sectors, and so the estimates $\hat{\delta}_j$ represent the wage of a unit of human capital in sector j relative to the excluded sector J, which in all of my specifications will be the agricultural sector.

An important note is that the above specification does not include a country identifier. I am not estimating this equation for all 15 surveys at once, but rather for each country individually. This allows the returns paid by a sector to be unique to each country, and not reflect some common sector-level effect on wages. There may be differences in sector-level wages that are common to all countries, perhaps reflecting the unmeasured component of human capital required for work in those sectors. By ignoring those, I will be overstating the variation in δ_j within countries, and as mentioned this will inflate the ultimate calculation of R.

[9] The distribution of jobs across sectors in each country can be found in the online appendix.

Table 1
Summary statistics, by country.

Country	Year	Country data			From RIGA				From IPUMS
		GDP p.c.	Education	Perc. rural	Obs.	CV (wage)	Education	Perc. rural	Perc. wage-worker
Ex-communist									
Albania	2005	$6107	10.2	55.0	2396	3.05	11.6	26.9	.
Bulgaria	2001	$7664	9.4	31.0	3397	1.39	11.4	24.3	.
Tajikistan	2003	$1250	9.9	74.0	4593	1.24	11.1	72.4	.
Asia									
Bangladesh	2000	$1003	3.7	76.0	6898	0.80	3.3	64.6	37.5
Indonesia	2000	$2623	4.8	58.0	8824	1.71	10.7	42.6	34.1
Nepal	2003	$919	2.7	85.0	4397	0.90	3.8	72.5	24.3
Vietnam	1998	$1469	4.5	77.0	6212	0.58	7.4	63.2	27.1
Latin America									
Ecuador	1995	$5664	6.9	42.0	7833	1.09	8.6	39.4	40.8
Guatemala	2000	$3960	3.8	55.0	10,173	1.07	5.5	49.5	.
Nicaragua	1998	$1982	4.6	46.0	5069	1.47	5.7	40.6	42.4
Nicaragua	2001	$2169	4.6	45.0	5412	0.91	6.2	38.4	53.6
Panama	2003	$8240	9.0	31.0	7745	3.44	9.9	40.4	68.1
Sub-Saharan Africa									
Ghana	1998	$1033	6.3	58.0	1631	1.21	7.8	48.3	16.0
Malawi	2004	$646	3.4	83.0	13,030	1.57	4.8	88.1	20.4
Nigeria	2004	$1702	–	55.0	3756	11.80	5.8	54.4	.

Notes: For country-level data, GDP per capita and percent rural are from the World Development Indicators. GDP per capita is PPP in 2005 international dollars. Education data is years attained by population 25 or older, from Barro and Lee (2010). The RIGA data is described in the text. IPUMS International data was used to calculate the percent of wage/salary workers out of all workers for countries where data was available. The census year in IPUMS are as follows: Bangladesh (2001), Indonesia (2000), Nepal (2001), Vietnam (2009), Ecuador (1990), Nicaragua (1995, 2005), Panama (2000), Ghana (2000) and Malawi (2008).

For each country I estimate Eq. (22) four times, varying the contents of the X_i vector of individual characteristics. The four different specifications are as follows:

1. X_i is empty. This provides a raw estimate of sector-level wage per worker. It is this specification that provides the information in Fig. 1.
2. X_i contains education (in years), age, age-squared, and gender. These are the primary controls for human capital.
3. X_i contains occupation dummies, as well as the human capital controls in specification 2. Occupation will capture differences in human capital that may exist between workers who have identical education, age, or gender. It is a crude measure of skill outside of years of education.
4. X_i contains occupation dummies, and occupation-specific returns to education, as well as the human capital controls in specification 2. Here, I am allowing for the possibility that some occupations (e.g. Tradesmen) may see a lower return to formal schooling than others (e.g. Professionals and Technicians). This specification is what underlies Fig. 2 in the introduction.

Each specification is estimated using ordinary least squares, with standard errors calculated using the sampling weight provided by the original LSMS survey. This is simply an explicit way of accounting for heteroskedasticity. Results without weights, but with robust standard errors, provide nearly identical results. Each specification also excludes sectors that have fewer than 10 observations, as there are too few to provide useful estimates of sector dummies.

In each specification I obtain estimates $\hat{\delta}_1$ through $\hat{\delta}_{J-1}$, where J is the total number of sectors in a given country. I report the values of $\hat{\delta}_j$ in the online appendix, as the specific values are not of primary interest. Recall that wages in a sector are estimated by $\hat{w}\left(1+\hat{\tau}_j^W\right) = exp\left(\hat{\beta}_0+\hat{\delta}_j\right)$, so that variation in the values of $\hat{\delta}_j$ implies variation in the wage paid to a unit of human capital across sectors, and under the assumptions I've made this would suggest a misallocation of human capital.

Table 2
Wages relative to overall mean wage, by sector and country.

Country (year)	Sector								
	Agriculture	Mining	Manufacturing	Utilities	Construction	Commerce	Transportation	Finance	Services
Albania (2005)	1.63	1.15	0.84	0.87	1.27	0.76	0.96	1.46	0.97
Bulgaria (2001)	1.09	–	1.04	–	1.11	0.83	1.05	0.84	1.03
Tajikistan (2003)	0.62	–	1.43	–	2.39	2.24	2.08	–	1.65
Bangladesh (2000)	0.81	–	1.00	1.71	1.26	0.84	1.40	1.18	1.15
Indonesia (2000)	0.62	1.23	0.94	1.35	0.90	0.94	1.19	1.75	1.42
Nepal (2003)	0.74	1.33	1.03	1.57	1.36	1.25	1.73	2.53	1.48
Vietnam (1998)	1.03	0.98	1.09	1.02	1.23	1.21	1.42	0.74	0.75
Ecuador (1995)	0.44	1.73	1.26	1.80	1.52	1.07	1.54	2.27	1.52
Guatemala (2000)	0.69	0.94	1.03	1.26	1.19	1.12	1.44	2.04	1.25
Nicaragua (1998)	0.66	1.46	1.25	1.83	1.03	1.06	1.47	1.89	1.24
Nicaragua (2001)	0.69	1.04	1.14	1.91	1.25	1.19	1.47	2.03	1.09
Panama (2003)	0.58	0.88	1.02	1.43	1.07	1.05	1.21	1.96	1.24
Ghana (1998)	0.68	2.25	0.73	1.03	0.88	0.56	1.09	1.45	1.31
Malawi (2004)	0.84	–	1.81	2.15	0.96	1.51	2.86	2.47	2.16
Nigeria (2004)	0.49	1.21	0.87	0.69	1.17	0.80	0.81	1.82	1.54

Notes: The table reports the average of the individual wages relative to the overall mean wage, by country and sector. The data are from the RIGA database, Davis et al. (2010). See text for more details. Blank cells indicate that no individuals reported themselves as working in that sector.

As a first step consider the test with the null hypothesis H_0:$\delta_1 = \delta_2 = \ldots = \delta_{J-1} = 0$. If the δ_j values are all identical this would imply an efficient allocation of human capital across sectors under the assumptions I've made regarding the substitution of human capital. If I reject the null hypothesis, then the sector-level wages (or at least some of them) are statistically different from one another. A rejection thus suggests that there is potentially some scope for finding effects of human capital misallocation. Note that a rejection implies misallocation only under the assumptions made: perfect substitutability of human capital and the absence of unmeasured human capital. If those assumptions are not true, then rejecting H_0 doesn't necessarily imply anything about labor market efficiency.

The test statistic is distributed as an $F(J-1, N-J-1)$ distribution, where N is the number of observations and J is the number of sectors. For all countries, and under all specifications of the Mincerian regression, the null hypothesis is rejected. Practically, the values of δ_j vary enough between sectors to reject that their differences could be due simply to sampling variation in the RIGA data. The exact values of the F-statistic are reported in the online appendix, as they are not themselves of any particular interest, and the p-values are all well below 0.001.

It's possible to add some additional insight into how sector returns to labor vary within countries, and to identify any patterns across countries in which sectors have particularly high or low returns. I convert the estimated dummies, $\hat{\delta}_j$, into comparable measures of the wage-wedges, $(1+\tau_j^W)$ following the work of Krueger and Summers (1988). Specifically, I first calculate a sector-level dummy, μ_j, that gives wages in sector j relative to the average wage (rather than relative to Agriculture).

$$\mu_j = \hat{\delta}_j - \sum_{i=1}^{J} \hat{\delta}_j s_j \tag{23}$$

where $\hat{\delta}_J = 0$ for the reference sector of Agriculture. The value s_j is the share of all workers in sector j. The estimated wage-wedge is then simply

$$\left(1+\hat{\tau}_j^W\right) = exp\left(\mu_j\right). \tag{24}$$

Fig. 1 from the introduction is a plot of the values of $\left(1+\hat{\tau}_j^W\right)$ for each country, starting from the estimates of δ_j from specification (1), which incorporates no individual controls.[10] It thus represents the widest possible variation in sector-level wages. One can see that Bulgaria and Vietnam have particularly small variation across sectors in wages, while Tajikistan and Malawi appear to have the most. There are sectors with wages roughly 2.5 to 3 times the average wage, while most cluster in the range from 1 to 2.

If we incorporate the full individual controls of specification (4), then this reduces the variation in wage-wedges. This can be seen in Fig. 2 from the introduction. Compared to Fig. 1, the values of $\left(1+\hat{\tau}_j^W\right)$ are clearly more compressed, clustering around the value 1, indicating sector wages equivalent to the average wage in the economy. There are a few outliers, such as the Mining sector in Ghana, but for the most part the outlying sectors in Fig. 1 have collapsed towards the average wage, meaning the wage-wedge is itself is close to zero. This reflects the fact that much of the "raw" difference in sector-level wages is driven by differences in human capital. Once we control for these, the implied sector-level variation declines.

As an alternative way of looking at the data, consider Fig. 3, which instead shows the different country values of $\left(1+\hat{\tau}_j^W\right)$, grouped by industry. The figure is a box and whiskers plot, allowing one to see more clearly the amount of variation in sector-wages across the different country samples. In Fig. 3, the estimates of $\left(1+\hat{\tau}_j^W\right)$ are based on specification (1), with no individual characteristics controlled for. One can see that agricultural wages tend to be below average in each country. No other sector has a great tendency to lie below one, meaning they have above-average wages. All the true outliers (denoted by the dots in the figure) lie well above one, indicating positive wage-wedges.

Compare this to Fig. 4, in which the various sector $\left(1+\hat{\tau}_j^W\right)$ values are obtained from specification (4), which includes the full human capital and occupation controls. As can be seen, there is a great deal of compression, so that there is less variation across countries in the sector wages. Agriculture remains below one, in general, but not as severely. The commerce sector also has some tendency to be below one, but most other sectors retain wages that are above the average wage.

3.2. Calculating R

Given the estimates of $\hat{\beta}_0$ and $\hat{\delta}_j$ from the Mincerian regressions in a country, I follow the procedure outlined in Section 2 to find $\hat{w}\left(1+\hat{\tau}_j^W\right)$, $\hat{H}_j$, and $\hat{\Omega}_j$. Using these values I calculate the implied gain from reallocation, R.

I make the initial calculations using an assumed value of $\alpha = 0.3$ for the elasticity of wages with respect to human capital. This would be the value that would hold if the share of human capital in output was 0.7. This value is consistent with the typical approach used in the existing literature based on macro-level data. In a subsequent section, I will show how R changes when the assumed value of α is smaller.

The results are contained in Table 3, in columns (1) through (4). Column (1) shows the calculated gain from reallocation, R, under the first specification of human capital, X_i. In this case, there are no human capital controls included in the regressions, which maximize the variation of the $\hat{\delta}_j$ sector dummies, and therefore maximizes the possible gain from reallocation. One can see that with this limited specification Albanian earnings would rise by 6.4.

Moving across columns, the more refined specifications for human capital are shown, and as noted before these limit the variation in $\hat{\delta}_j$, and narrow the gains possible from reallocation. Within each country, the gains from reallocation shrink. In column (2), with controls for education, age, and gender in the Mincerian equation, the gain drops noticeably for several countries. In Indonesia, Nepal, and most central American countries the gains all drop by substantial amounts, falling

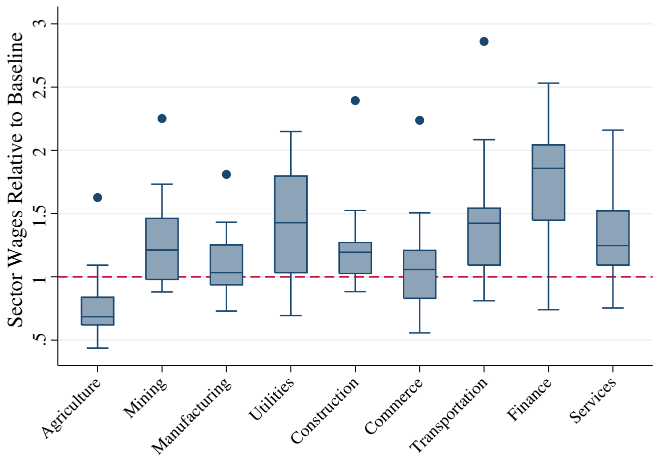

Fig. 3. Sector-level wages, no human capital controls. Note: The figure shows the sector-level wage relative to the average country wage $(1+\tau_j^W)$. The sector-level wages are calculated using the sector dummies from a Mincerian regression with no additional controls, as described in the text.

[10] Strictly speaking, calculating $\left(1+\hat{\tau}_j^W\right)$ from μ_j should include an accounting for the standard error of the δ_j estimates. In practice, this adjustment is so small I have not bothered to report it.

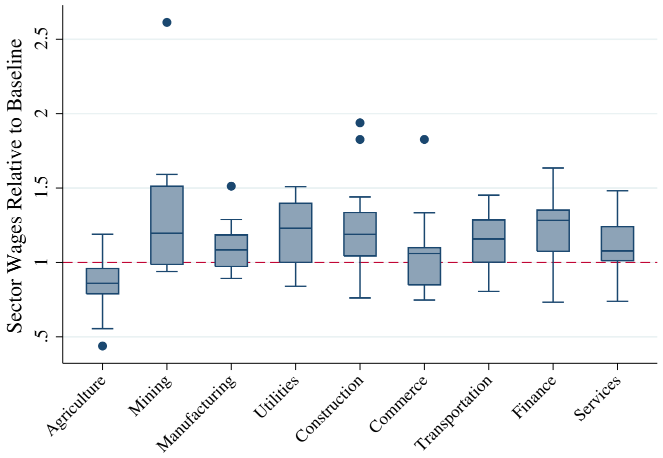

Fig. 4. Sector-level wages, full human capital controls. Note: The figure shows the sector-level wage relative to the average country wage $(1+\tau_j^W)$. The sector-level wages are calculated using the sector dummies from a Mincerian regression that includes controls for education, gender, age, age-squared, occupation dummies, and occupation-specific returns to education, as described in the text.

below 6% for each. In the African countries, the gains also decrease and fall below 10% for all; Malawi in particular sees a drop from 26% to 6%. Only in Tajikistan (25%) and Ecuador (14%) are the gains over 10%.

The gains of reallocation in column (4) have become negligible for nearly all the countries, where I include the full set of human capital controls. The gain in Tajikistan is only 15%, Ecuador's 9.6%, and Malawi's 2.6%. For Bangladesh, Indonesia, and Panama, the gains from reallocation are under 1% after controlling for occupations. Gains are under 5% for most of the countries shown. More importantly, as noted in Section 2, these gains are almost certainly upper bounds on reallocation gains. I am making the severe assumption that human capital is perfectly substitutable between sectors and that unmeasured human capital does not vary systematically across sectors. These assumptions allow for the greatest possible gain from reallocation across sectors. However, even with these assumptions in place the gains are generally only on the order of 5%, rising only to 15% for Tajikistan. Practically speaking, these gains are over-stated because units of human capital are not perfect substitutes and the underlying estimates of δ_j almost certainly capture unmeasured human capital as opposed to differences in wages between sectors.

These relatively small gains of reallocation are not due to assumptions limiting the movement of labor between sectors. As noted, I am allowing human capital to perfectly substitute across sectors, and this leads to rather large movements of human capital between sectors after reallocation. In Tajikistan, nearly 50% of the total human capital of the workforce must be removed from agriculture and reallocated to construction, commerce, and services. Across the Latin American countries, between 5 and 26% of total human capital must be shifted out of agriculture, generally to move into services, construction or utilities. For Albania, Bulgaria, Vietnam, and Ghana, the service sector would lose up to 19% of total human capital. In Ghana, every sector loses human capital save mining, where fully 35% of the human capital is implied to move in a reallocation. For most countries, about 20% of the entire stock of human capital must move sectors. Despite these large shifts, the implied earnings gain, as seen above, is not necessarily that large.[11]

One item to note is that columns (3) and (4) involve controls for occupation (and occupation-specific education returns). Is it appropriate to control for occupation as a component of human capital, when in fact many workers who switch sectors could well switch occupations as well? That is, occupation is not necessarily a fixed trait. However, it is not apparent that individuals could move seamlessly between the occupation categories controlled for here. The Professional category, for example, is not one that someone who currently has Clerk or Elementary Occupation can presumably just switch into easily without further training, education, or certification. Skilled Agricultural Workers are a separate occupation, and are those workers analogous to Craft and Trade Workers? In columns (3) and (4) I have gone to the extreme of assuming that occupations are fixed, while in column (2) the assumption is that workers are perfectly fluid between occupations. The truth would appear to lie somewhere in between, so estimates or R lying in between the results in these columns are most likely accurate. Without taking a stand on precisely what the transition possibilities are between occupations, more accuracy is not possible.

How do these gains compare to those found in other studies? Comparisons are not easy because of differences not only in methodology but in time frame and data quality. Vollrath (2009a) finds that in some sub-Saharan African countries the gains from reallocation could be as large as 150%, and for much of Central America gains are on the order of 50–60%. As noted before, that paper eliminates both the revenue and wage-wedges, so it is not surprising that the gains are larger than those found here. However, that paper is using data from 1985, so perhaps some of the difference is that labor markets have become more efficient in the roughly twenty years that elapsed before the surveys I use here.

An additional issue is that measurement of human capital was crude in Vollrath (2009a). The more accurate data I have in this paper on human capital could explain why the estimated gains are so much smaller. A better comparison is probably the recent work of Gollin, Lagakos, and Waugh (2013), who look at the "agricultural productivity gap", defined as the ratio of labor productivity in non-agriculture to that of agriculture. They use detailed census and survey information to better measure human capital, but as in Vollrath (2009a) use aggregate data at the sector level to measure output.

Gollin et al. provide (their Table 4) sufficient information to calculate the total gain (i.e. eliminating both revenue and wage-wedges) from reallocation between agriculture and non-agriculture for 10 developing countries.[12] In terms of comparison, their data indicates a gain in Ghana of about 20%, about double the gain that I find for Ghana using only the wage wedge. In Guatemala, their work suggests a gain of 14%, while here I've estimated the wage-wedge to imply a gain of about 3%. For Panama, their data indicate a gain of 3.7%, and I've found a gain of between 1 and 2.4% depending on how I control for human capital. The most striking comparison is Bulgaria, where their data indicates a gain of 81% while my analysis shows one of only about 1%. It is clear that eliminating both revenue and wage-wedges would raise the gain from reallocation, and in at least one case dramatically. What my analysis suggests is that observed wage-wedges by themselves do not indicate large aggregate losses to productivity.[13]

3.3. Cross-country comparisons

Leaving aside the absolute size of the gains from reallocation, for inefficiency of human capital allocations to provide a useful explanation for cross-country differences in income per capita, it must be that poorer countries tend to have more inefficient markets. If one goes through Table 3, it becomes apparent that there is no distinct relationship between GDP per capita and the estimated size of R. The poorest countries I study, Malawi and Nepal, have gains from reallocation of 2.6% and 5.1%, respectively. The Latin American countries, generally richer, have gains around 2.5% for most, and as large as 9.6% for Ecuador.

[11] The online appendix contains Tables A.6 and A.7 showing the exact shifts in human capital between sectors necessary to reach an efficient allocation in the case where the full set of human capital controls are included.

[12] To find this, consider re-arranging Eq. (3) to be $(1-\alpha)Y_j/H_j = w(1+\tau_j^W)/(1-\tau_j^R)$. Given data on output, Y_j, and the human capital stock H_j in each sector, one can find a value for the combined wedge term on the right-handside. Then it is straightforward to calculate R after eliminating both wedges.

[13] For Tajikistan, the Gollin et al. (2013) data suggest a gain of only 3.4%, while I calculate a gain on the order of 15%. This is due to the fact that I am using more sectors than they are, and so there are more avenues to find gains.

Table 3
Estimated efficiency gain from re-allocation, R, by specification and country.

	Baseline				Including self-employed			Wage elasticity	
Country (year)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Albania (2005)	1.064	1.071	1.027	1.027	–	–	–	1.277	1.128
Bulgaria (2001)	1.008	1.010	1.015	1.014	–	–	–	1.033	1.016
Tajikistan (2003)	1.342	1.249	1.153	1.155	–	–	–	1.640	1.389
Bangladesh (2000)	1.045	1.027	1.010	1.009	1.025	1.036	1.025	1.103	1.047
Indonesia (2000)	1.102	1.019	1.005	1.006	1.024	1.019	1.027	1.062	1.031
Nepal (2003)	1.129	1.060	1.050	1.051	1.056	1.057	1.063	1.202	1.102
Vietnam (1998)	1.036	1.036	1.037	1.035	1.018	1.033	1.016	1.096	1.056
Ecuador (1995)	1.219	1.141	1.100	1.096	1.191	1.137	1.253	1.308	1.199
Guatemala (2000)	1.097	1.030	1.023	1.025	–	–	–	1.105	1.049
Nicaragua (1998)	1.075	1.034	1.028	1.028	1.046	1.038	1.054	1.109	1.055
Nicaragua (2001)	1.067	1.031	1.025	1.025	1.040	1.035	1.045	1.117	1.052
Panama (2003)	1.065	1.024	1.009	1.009	1.036	1.028	1.041	1.089	1.039
Ghana (1998)	1.140	1.089	1.108	1.114	1.102	1.108	1.074	1.529	1.192
Malawi (2004)	1.261	1.059	1.028	1.026	1.052	1.052	1.056	1.229	1.106
Nigeria (2004)	1.119	1.086	1.100	1.111	–	–	–	1.182	1.123
United States (2000, weekly)	1.051	1.021	1.018	1.018	–	–	–	1.065	1.030
United States (2000, hourly)	1.023	1.009	1.009	1.009	–	–	–	1.034	1.015
Controls included in specification									
Education and demographic	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation	No	No	Yes	Yes	No	No	No	No	No
Occupation-specific returns	No	No	No	Yes	No	No	No	No	No
Rel. self-emp. ag. wage	–	–	–	–	1.0	0.1	1.0	–	–
Rel. self-emp. non-ag. wage	–	–	–	–	1.0	0.1	0.1	–	–
Wage elasticity (α)	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.1	0.2

Notes: The value of R is the ratio of wages under the optimal allocation of human capital to the observed level of wages, see text for details. The estimates of R depend upon estimated Mincer equations, using the human capital controls indicated, again see text for details regarding the exact specification. The self-employed columns are calculated using distributions of workers from IPUMS that include self-employed workers, and with their earnings relative to wage-workers in their sector set as noted in the table. The wage elasticity columns refer to the value of α, which dictates the response of wages to the amount of human capital in a sector. The different rows for the U.S. refer to whether weekly earnings or hourly earnings are used in the calculations.

To make a more concrete comparison, I have made a similar calculation of R for the United States. I use earnings data from the March Current Population Survey from the year 2000. This dataset contains over 161,000 observations on individual-level earnings, and includes not only those working for wages but self-employed workers as well, which I will return to in the following section. The CPS data has similar human capital controls available at the individual level, and I have recoded the reported industry and occupation classifications to match those used in RIGA. See the online appendix for full details.

The CPS reports weekly earnings as well as hours worked per week. As I do not have hours worked in the RIGA database, nor do I have days worked per month (or week) in the CPS, I cannot create a strictly comparable measure of wages between the U.S. and the other countries. I report results in Table 3 in the final two rows using both weekly earnings and hourly earnings for the U.S.

As can be seen, for weekly earnings the gain from reallocation across sectors is on the order of 5% when no human capital controls are included, but fall to 1.8% with the full controls in column (4). This gain from reallocation is smaller than that found in most countries, but is notable for still being larger than that found in Bulgaria, Bangladesh, Indonesia, and Panama. The implied gains in Albania, Guatemala, Nicaragua, and Malawi are not much larger than 1.8%, suggesting that their human capital is allocated as efficiently as in the United States.

When using hourly earnings, the implied gains from reallocation in the U.S. fall by about half, so that in column (4) with full human capital controls, the gain is roughly 1%. This makes the gain smaller than in nearly every country save Indonesia, suggesting that the in general the U.S. has the most efficient allocation of human capital among the countries being compared.

Note that the adjustment to hourly earnings lowered the gains from reallocation substantially in the U.S., implying that weekly earnings differences between sectors may reflect differences in hours worked as opposed to differences in the wage rate paid to human capital. For the developing countries using the RIGA data, recall that I have only daily wages, and not hourly wages. The U.S. results suggest that there could be a meaningful reduction in R if I could examine hourly earnings, as differences in hours worked per day may vary across sectors.

Regardless, even if one simply compares these developing countries with the U.S., the implied reallocation gains pale in comparison to the actual income gaps. U.S. has GDP per capita on the order of 20–50 times higher than the countries in this paper. Gains from reallocation of even 15% are simply not capable of explaining a significant fraction of this gap.

3.4. Unmeasured human capital and reallocation

To this point I have been assuming that any observed sector-level differences in earnings (after controlling for observed human capital characteristics), δ_j, were the result of differences in wages. However, it is almost certainly true that there are in fact significant differences in unmeasured human capital between sectors. Abowd, Kramarz, and Margolis (1999) find, using French data, that most of the observed inter-industry wage differentials they find are due to unmeasured individual heterogeneity, and Goux and Martin (1999) find a similar result. Using matched worker-firm data from the state of Washington, Abowd, Finer, and Kramarz (1999) attribute half of the inter-industry wage differentials they find to individual heterogeneity.

Practically, this would mean that a more appropriate description of earnings is

$$\ln m_{ij} = \ln\ w_j + u_j + X_i'\beta + \epsilon_{ij} \tag{25}$$

where u_j is unobserved human capital for individuals working in sector j. The sector dummies that I estimate in my Mincerian regressions are thus picking up $\delta_j = \ln w_j + u_j$, and not simply a wage difference.

While by definition I do not have the means to measure u_j, I can provide a crude means of understanding how ignoring it inflates my estimates of the gains from reallocation, R. Assume that some fraction $\gamma \in (0,1)$ of the estimated value δ_j is actually capturing the average unmeasured human capital in sector j. That is, let $u_j = \gamma\delta_j$. As I do not have any better information, I assume that this fraction is the same in each sector. Setting $\gamma = 0$ is implicitly what I have already been doing. As γ goes towards one, less of the estimated δ_j values represent differences in the wage paid to human capital, and more represent differences in unmeasured human capital. So as γ goes to one, the potential gains of reallocation shrink, as the observed differences in δ_j simply represent my inability to accurately measure the human capital employed in each sector.

I recalculate values of R using varying values for γ to provide an idea of how important unmeasured human capital might be for the gains for reallocation. Specifically, for each individual I now estimate human capital as:

$$\ln \hat{h}_{ij} = \gamma\hat{\delta}_j + X_i'\hat{\beta} + \hat{u}_{ij} \tag{26}$$

which adds in the fraction γ of the estimated sector dummy $\hat{\delta}_j$ to each individual in sector j. H_j is again the sum of individual human capital employed in sector j, only now with individual human capital estimated using Eq. (26). The estimate of $w(1+\tau_j^W)$ from Eq. (15) is modified to be

$$\hat{w}\left(1+\hat{\tau}_j^W\right) = exp\left(\beta_0 + (1-\gamma)\hat{\delta}_j\right). \tag{27}$$

Given this estimate of the wage wedge and the values of H_j, I can again calculate productivity terms Ω_j and then R, the gain from reallocation.

In Fig. 5 I've plotted the values of R for the four countries that had the largest gains from reallocation in my original estimation: Ecuador, Ghana, Nigeria, and Tajikistan. As can be seen, as γ rises, the value of R drops rather quickly for each. If γ is equal to 0.25, indicating that one-quarter of the sector-specific return $\hat{\delta}_j$ is actually unmeasured human capital, then the gain for Tajikistan is down to 10%, and for the others it is only around 6–7%. By the time $\gamma = 0.5$ all four countries have gains from reallocation of less than 5%. The work of Abowd, Finer, and Kramarz (1999) attributed half of inter-industry wage differentials to unmeasured individual human capital, suggesting that $\gamma = 0.5$ is not an unreasonable possibility, and work using French workers and firms suggests γ may be even higher. Regardless, Fig. 5 shows that even a mild degree of unmeasured human capital makes the gain from reallocation even smaller than originally estimated.

Recall that the other countries all began with gains smaller than the four shown in Fig. 5, and these shrink demonstrably towards zero as γ rises. The evidence suggests that unmeasured human capital is likely an important source of the apparent sector-level differences in wages, and this means reallocation would have little effect on wages. It seems that wage labor is very close to efficiently allocated across sectors.

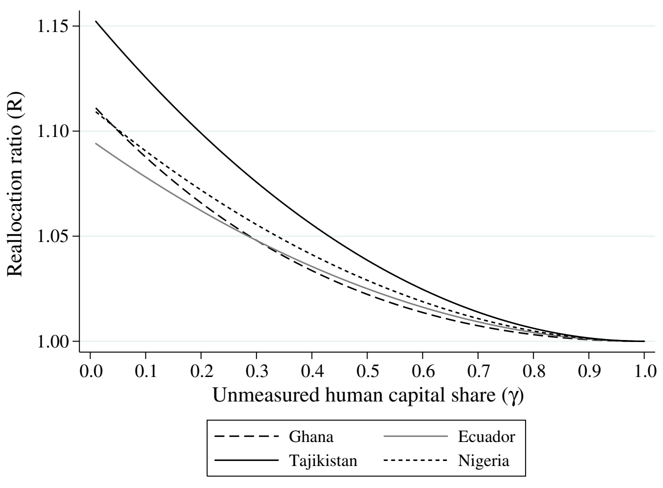

Fig. 5. Reallocation Ratio and Unmeasured Human Capital. Note: The figure shows the gains from reallocation (R) plotted against the fraction (γ) of the sector-level return (δ_j) that is attributed to unmeasured human capital. The four countries in the figure are those with the highest estimated gains from reallocation based on the original analysis with $\gamma = 0$.

3.5. Accounting for self-employed labor

One of the shortcomings of the RIGA data I am using is that it only includes wage workers, and excludes the self-employed. In many developing countries, a large majority are working for their own account. One possibility is that while the gains from reallocation among those employed for wages are small, there could well be large gains to reallocation between the self-employed and those earning wages. Given the large gaps observed between agriculture (with a large proportion of self-employed farmers) and non-agriculture in labor productivity in the aggregate data, this distinction is worth exploring further.

To do this, however, I require data on the distribution of all workers (wage-workers and self-employed) across sectors. To obtain this information, I use IPUMS census data for years as close to the RIGA survey dates as possible. IPUMS does not perfectly overlap with RIGA, so I only am able to obtain data for 10 of the 15 surveys I consider. In each census, there is a standardized code for work status. For all individuals that report a work status, the fraction of "wage/salary" workers is reported in Table 1 in the final column.[14]

IPUMS also reports the industry of employment for each worker, regardless of work status. For several countries the classification is not standard, but mapping the idiosyncratic codes from those countries to standard ISIC codes is relatively straightforward. The online appendix shows the re-mapping I did to standardize the industry codes in the census data. With standard industry classifications, I now have the distribution of all workers – regardless of self-employed or wage-work – across all sectors. Online Appendix Table A.8 shows this data for the 10 countries with data.

I then re-calculate R using the distribution of workers from IPUMS, including the self-employed, as a comparison to the gains calculated using only the wage workers. To do this I use the wage-wedges estimated already, and then vary the amount the self-employed workers earn relative to wage-workers in their sector. This allows me to calculate total earnings of all workers in a sector, and from that I can back out the total human capital (which includes both wage-workers and the self-employed) used in a sector, H_j. With the wage-wedges, $(1+\tau_j^W)$, and human capital stocks, H_j, I use Eq. (19) to back out the productivity term in each sector, Ω_j. With this I can calculate R.

In Table 3, columns (5)–(7) show the results of these calculations. In all three columns, the wage-wedges I use are calculated using specification 2, which includes only basic human capital controls. So columns (5)–(7) are best compared to column (2). To begin, in column (5) I assume that self-employed workers earn exactly the same amount as their wage-working peers.[15] As can be seen, using the distribution of all workers does not materially alter the results compared to column (2). The gain in Ecuador rises to 19% from 14%, and that in Ghana to 10% from 9%, but for the most part the gains remain relatively small. In several cases – Bangladesh, Nepal, Vietnam, and Malawi – the gains actually go *down* when using the IPUMS distribution of workers.

[14] The remainder of the workers are either "self-employed", "unpaid workers" (typically doing household work), or "other".

[15] Without individual-level data on self-employed earnings (which IPUMS does not have) I cannot say anything with certainty about self-employed earnings. However, RIGA reports total household labor earnings, broken down by source (non-agricultural wages, agricultural wages, non-agricultural self-employment, agricultural self-employment). There is also information on the number of working adults in each household. Using this information, I calculate earnings per worker for each household. Across the countries in RIGA, households in agriculture with only self-employed income earn about 80–120% of what households in agriculture with only wage income earn. There are exceptions like Panama where the self-employed earn only 33% as much. In non-agriculture, households with only self-employment earnings earn about 70–100% (depending on the country) of households with only wage income.

Assuming that self-employed workers earn the same as wage-workers in a sector may not be accurate, in particular for the non-agricultural sectors. In column (6) I instead assume that self-employed workers in a sector earn only 10% as much as wage-workers in the same sector. As can be seen, the effect is not monotonic. Some countries see a larger gain, some smaller gains. However, in no case are the gains materially different than in column (2).

Finally, in column (7) I show the results when I assume that agricultural workers earn 100% as much as agricultural wage-workers, while all other non-agricultural self-employed workers earn only 10% as much as wage-workers in their sectors. Here again the results do not deviate strongly from the previous ones, with the possible exception of Ecuador, where the gain reaches 25%.

In many of these countries the distribution of all workers across sectors differs markedly from the distribution of wage workers. In a country like Ghana, while only 11% of wage-workers are in agriculture, 55% of all workers are in agriculture. Why does the inclusion of the self-employed workers not have a larger effect? Recall from Section 2.1 that what drives the size of R is the variance in productivity across sectors, not the variance in wage-wedges. The wage-wedges in agriculture and commerce are low because of the very large stocks of human capital employed there. So relative to the rest of the economy, productivity in these sectors is not necessarily that low, and that means the gains from reallocation remain small. Accurate data on self-employment earnings, so that one could estimate the gains directly, could well overturn the results in columns (5)–(7). However, simply using the distribution of workers found in IPUMS does not, by itself, appear to materially change the results.

3.6. The elasticity of wages

A final element influencing the size of R will be α, the elasticity of wages with respect to human capital. This dictates how fast the marginal product falls in the high-wage sector as human capital is increased, and how fast the marginal product rises in the low-wage sector as human capital decreases. In one extreme $\alpha = 0$ and the optimal allocation would be to move every unit of human capital into the highest-wage sector. In the example of Ghana, this would imply that all units of human capital go to work in the mining sector. This seems to be unrealistic, as even economies that we consider to have very efficient labor markets (the United States, for example) are not completely specialized. In the empirical calculations that follow I will consider several values for α in an attempt to see how relevant it is to the implied gains from reallocation.

As noted previously, the elasticity of wages with respect to human capital by sector is assumed to be $\alpha = 0.3$ in all the prior calculations. The smaller the elasticity, though, the greater will be the implied gains from reallocation. The reason is that as we add more and more human capital to a high-wage sector, the implied wage does not fall as much, and more human capital can be loaded into the high-wage sector before wages converge across sectors.

Hamermesh (1993) discusses estimates of the own-price elasticity of labor demand, finding a range of [0.15, 0.75] plausible and naming 0.30 as the "best" estimate of this elasticity. With an elasticity of substitution between capital and labor of one (i.e. Cobb–Douglas) α will be equal to this own-price elasticity. Hence the chosen value of 0.30 matches the existing literature, but values as low as 0.15 may be plausible.[16]

In a static setting very small values are likely to be unrealistic, but I recalculate R for both $\alpha = 0.10$ and $\alpha = 0.20$. In both cases, I do so using the estimates produced using only the education and demographic controls (specification 2) and ignoring unmeasured human capital ($\gamma = 0$). Table 3 reports the results of these alternative calculations in columns (8) and (9).

As can be seen, in column (8), with $\alpha = 0.1$, the value of R is distinctly larger for every country, as expected. However, many countries the gains are still around 10% or less. Albania (27.7%), Tajikistan (64.0%), Nepal (20.2%), Ecuador (30.8%), Ghana (52.9%), Malawi (22.9%) and Nigeria (18.2%) now all have gains above 15%. As these results do not account for unmeasured human capital, the actual gains are smaller than these upper bounds.

For the most part, similar conclusions follow from before. Gains of less than 10% could well be practically important, as certainly wage workers in those countries would appreciate an increase in their wages. However, these gains are again very small relative to the income gaps with the developed world, and they therefore would have trouble explaining much of the existing income gaps. More importantly, there still remains no meaningful relationship between the size of gains and the existing income level. In other words, the gains are larger, but are still not systematically related to development.

For those countries that do have large gains, it is also useful to look back at Fig. 2. Consider Ghana, a country with one of the largest potential gains. As can be seen, in Ghana the Mining sector has a sector-level wage that lies well outside the group of other sectors, with wages per unit of human capital about 2.5 times the country average. With a very low value for α the reallocation is loading nearly everyone into the Mining sector. Thus the average wage is drawn upwards significantly. Whether this is a realistic reallocation, though, is open to question. In the case where $\alpha = 0.1$ the implied efficient allocation is to have 97.3% of all the human capital in Ghana work in the Mining sector. It seems unlikely that it is possible to employ that many people (and their human capital) in mining. So the gain of 52.9% in wages after reallocation seems unreasonable as well.

The spread of wages in Tajikistan, Ecuador, Nepal, and Nigeria also show similar wide gaps in Fig. 2. These generate the large values for R, but the question remains whether the elasticity of wages is actually so low that reallocation could generate the gains in columns (8) and (9) of Table 3. Most importantly, the reallocation is loading much of the human capital stock into the Finance and Commerce sectors for these countries. The question is whether these sectors could absorb large amounts of labor without decreasing wages by very much, or if they could even absorb large amounts of labor at all.

This comes back to a point made earlier regarding the flexibility of human capital. In these calculations I am assuming that the human capital stock is perfectly substitutable between different sectors and all units of human capital are perfectly substitutable with one another. Values of α close to 0.10 generate larger gains by exploiting this flexibility fully. However, the perfect substitutability of different units of human capital seems unlikely to be true and so the gains are surely smaller than those indicated by Table 3 for $\alpha = 0.10$. Unmeasured human capital would also play a role here, decreasing the implied gains as shown in Fig. 5. In the end, the values in columns (8) and (9) are at the extreme end of any plausible range for the gains from labor reallocation. Even raising the elasticity to $\alpha = 0.20$ in column (9) generates substantially lower gains from reallocation.

These are all arguments for small *static* gains from reallocation. As noted in the introduction, though, dynamic gains from reallocation may still be large. One simple way of looking at dynamic gains from reallocation is to imagine what would happen to physical capital in response to a reallocation of human capital. In sectors that add human capital, the marginal product of physical capital would rise, and we would expect either a flow of existing capital into that sector, or the accumulation of new capital. In either case, the additional physical capital would offset the declining marginal product of human capital. In a dynamic setting, this implies that the value of α approaches zero. The dynamic gains from reallocation would then be much larger, over 100% for most of the countries studied here.

[16] If labor and capital are complements in production, which in the static setting under consideration here may be an appropriate assumption, then α will be larger than the own-price elasticity, and values above 0.30 would be prudent. Only if one considers labor and capital to be highly substitutable, with an elasticity of substitution over four or five, would it make sense to consider values of α below 0.10.

4. Conclusion

Large differences in output per worker across sectors suggest that developing countries may be operating well below their maximum efficiency. One possible explanation for this is that human capital is misallocated across different sectors within these countries. This paper calculates the hypothetical aggregate impact of this misallocation. In contrast to the existing macro-level literature that uses aggregate data, I use micro-level wage data from 14 developing countries. With this data I use Mincerian regressions to find gaps between sector-level wages and the average wage, which I term "wage-wedges". The wage-wedges imply some type of misallocation of human capital across sectors. I am then able to calculate the counter-factual level of aggregate productivity that a country would achieve if I eliminated all the wage-wedges in 10 broad sectors.

I find that the static gains from eliminating wage-wedges tend to be small. In only a few cases are the gain to productivity over 10%, and for most countries the gain is less than 5%. As a comparison I do a similar calculation for the United States. While the U.S. has smaller implied gains (i.e. it has a more efficient allocation of human capital), the differences are miniscule compared to the overall differences in output per worker.

The main results assume that there is no unmeasured human capital. If I relax that assumption, the implied gains are even smaller in each country. The main results are also based on wage-workers only, ignoring self-employed workers. Using census data to establish the distribution of all workers across the 10 sectors, I recalculate the gains from reallocation. In general, the results conform to the main results. Gains from reallocation remain below 10% for most countries.

These findings do not imply that labor markets operate perfectly within these developing countries. A limit of my approach is that I cannot eliminate all wedges between the value marginal product of human capital and the wage. So there may be large gains to reallocation if these other wedges are removed. Additionally, I only consider *static* gains from reallocation, ignoring any follow-on effects on capital accumulation or total factor productivity. In a dynamic setting where these factors can adapt to the changing distribution of human capital, the gains of reallocation could be very large. My results suggest that if human capital misallocation is an important part of understanding under-development, then it is because of inefficiencies within sectors themselves or because of strong dynamic spillovers, but not necessarily because of static wedges between sectors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jdeveco.2014.01.009.

References

Abowd, John, Kramarz, Francis, Margolis, D., 1999. High wage workers and high wage firms. Econometrica 67 (2), 251–334.

Abowd, John, Finer, H., Kramarz, Francis, 1999. Individual and firm heterogeneity in compensation: an analysis of matched longitudinal employer–employee data for the state of Washington. In: Haltiwanger, J., Lane, J., Theeuwes, P. (Eds.), The Creation and Analysis of Employer–Employee Matched Data. North Holland, Amsterdam.

Banarjee, Abhijit, Duflo, Esther, 2005. Growth theory through the lens of development economics. In: Aghion, P., Durlauf, S. (Eds.), Handbook of Economic Growth. North Holland, New York.

Barro, Robert, Lee, Jong-Wha, 2010. A new data set of educational attainment in the world, 1950–2010. NBER Working Paper No. 15902.

Caselli, Francesco, 2005. Accounting for cross-country income differences. In: Aghion, P., Durlauf, S. (Eds.), Handbook of Economic Growth. North Holland, New York.

Chanda, Areendam, Dalgaard, Carl-Johan, 2008. Dual economies and international total factor productivity differences: channelling the impact from institutions, trade, and geography. Economica 75, 629–661.

Cordoba, Juan-Carlos, Ripoll, Marla, 2009. Agriculture and aggregation. Econ. Lett. 105 (1), 110–112.

Davis, Benjamin, Winters, Paul, Carletto, Gero, Covarrubias, Katia, Quinones, Estaban, Zezza, Alberto, Stamoulis, Kostas, Azzarri, Carlo, Digiuseppe, Stefania, 2010. A cross-country comparison of rural income generating activities. World Dev. 38 (1), 48.63.

Foster, Lucia, Haltiwanger, John, Krizan, C.J., 2001. Aggregate productivity growth: lessons from microeconomic evidence. In: Hulten, C., Dean, E., Harper, M. (Eds.), New Developments in Productivity Analysis. University of Chicago Press, Chicago and London.

Gollin, Douglas, Parente, Stephen, Rogerson, Richard, 2002. The role of agriculture in development. Am. Econ. Rev. Pap. Proc. 92 (2), 160–164.

Gollin, Douglas, Lagakos, David, Waugh, Michael, 2013. The agricultural productivity gap in developing countries. NBER Working Paper 19628.

Goux, D., Martin, E., 1999. Persistence of inter-industry wage differentials: a reexamination on matched worker-firm panel data. J. Labor Econ. 17, 492–533.

Graham, Bryan, Temple, Jonathan, 2006. Rich nations, poor nations: how much can multiple equilibria explain? J. Econ. Growth 11 (1), 5–41.

Hamermesh, Daniel, 1993. Labor Demand. Princeton University Press, Princeton, NJ.

Herrendorf, Berthold, Schoellman, Todd, 2011. Why is measured productivity so low in agriculture? Working paper.

Hsieh, Chang-Tai, Klenow, Peter J., 2009. Mis-allocation and manufacturing TFP in China and India. Q. J. Econ. 124 (4), 1403–1448.

Jones, Charles I., 2011. Misallocation, economic growth, and input–output economics. NBER working paper 16742.

Krueger, Alan, Summers, Lawrence, 1988. Efficiency wages and the inter-industry wage structure. Econometrica 56 (2), 259–293.

Kuznetz, Simon, 1971. Economic Growth of Nations: Total Output and Production Structure. Harvard University Press.

Lagakos, David, Waugh, Michael, 2013. Selection, agriculture, and cross-country productivity differences. Am. Econ. Rev. 103 (2), 948–980.

Lewis, W. Arthur, 1954. Economic development with unlimited supplies of labor. Manch. Sch. 22, 139–191.

McMillan, Margaret, Rodrik, Dani, 2011. Globalization, structural change, and productivity growth. NBER working paper w17143.

Minnesota Population Center, 2013. Integrated Public Use Microdata Series, International: Version 6.2 [Machine-readable database]. University of Minnesota, Minneapolis.

Restuccia, Diego, Rogerson, Richard, 2008. Policy distortions and aggregate productivity with heterogeneous establishments. Rev. Econ. Dyn. 11 (4), 707–720.

Restuccia, Diego, Yang, Dennis, Zhu, Xiaodong, 2008. Agriculture and aggregate productivity: a quantitative cross-country analysis. J. Monet. Econ. 55, 234–250.

Syverson, Chad, 2011. What determines productivity? J. Econ. Lit. 49 (2), 326–365.

Vollrath, Dietrich, 2009a. How important are dual economy effects for aggregate productivity? J. Dev. Econ. 11 (3), 187–228.

Vollrath, Dietrich, 2009b. The dual economy in long-run development. J. Econ. Growth 14, 287–312.