Human Capital and Regional Development

Nicola Gennaioli, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer

March 2011[1]

Abstract

We investigate the determinants of regional development using a newly constructed database of 1569 sub-national regions from 110 countries covering 74 percent of the world's surface and 96 percent of its GDP. We combine the cross-regional analysis of geographic, institutional, cultural, and human capital determinants of regional development with an examination of labor productivity and wages in several thousand establishments located in these regions. To organize the discussion, we present a new model of regional development that introduces into a standard migration framework elements of both the Lucas (1978) model of the allocation of talent between entrepreneurship and work, and the Lucas (1988) model of human capital externalities. The evidence points to the paramount importance of human capital in accounting for regional differences in development, but also suggests from model estimation and calibration that both entrepreneurial inputs and human capital externalities are critical to understanding the data.

[1] The authors are from CREI-UPF, Dartmouth College, EDHEC, and Harvard University, respectively. We are grateful to Nicolas Ciarcia, Nicholas Coleman, Sonia Jaffe, Konstantin Kosenko, Francisco Queiro, and Nicolas Santoni for dedicated research assistance over the past 4 years. We thank Elhanan Helpman, Edward Glaeser, Josh Gottlieb, Matt Kahn, Pete Klenow, Elias Papaioannou, Jacopo Ponticelli, Giacomo Ponzetto, Jesse Shapiro and seminar participants at the UCLA Anderson School for extremely helpful comments.

I. Introduction

We investigate the determinants of regional development using a newly constructed database of 1569 sub-national regions from 110 countries covering 74 percent of the world's surface and 96 percent of its GDP. We consider a variety of fundamental determinants of economic development, such as geography, natural resource endowments, institutions, human capital, and culture, by looking within countries. We combine this analysis with an examination of labor productivity and wages in several thousand establishments covered by the World Bank Enterprise Survey, for which we have both establishment-specific and regional data. Throughout the analysis, human capital measured using education emerges as the most consistently important determinant of both regional income and productivity of regional establishments. The combination of regional and establishment-level data enables us to investigate some of the key channels through which human capital operates, including education of workers, education of entrepreneurs/managers, and externalities.

To organize this discussion, we present a new framework describing the channels through which human capital influences productivity, which combines three features. First, human capital of workers enters as an input into the neoclassical production function, but the entrepreneur's human capital independently influences firm-level productivity. The distinction between entrepreneurs and workers has been shown empirically to be critical in accounting for productivity and size of firms in developing countries (La Porta and Shleifer 2008, Bloom and von Reenen 2010). In the models of allocation of talent between work and entrepreneurship such as Lucas (1978), Baumol (1990), and Murphy, Shleifer, and Vishny (1991), returns to entrepreneurial schooling may appear as profits rather than wages. By modeling this allocation, we trace these two separate contributions of human capital to productivity.

Second, our approach allows for human capital externalities, emphasized in the regional context by Jacobs (1969), and in the growth context by Lucas (1988, 2008). These externalities result from

people in a given location spontaneously interacting with and learning from each other, so knowledge is transmitted across people without being paid for. Because our framework incorporates both the allocation of talent between entrepreneurship and work as in Lucas (1978), and human capital externalities as in Lucas (1988), we call it the Lucas-Lucas model[2]. Human capital externalities have been shown to be important in a variety of development and regional contexts (Rauch 1993, Glaeser, Scheinkman and Shleifer 1995, Angrist and Acemoglu 2000, Glaeser and Mare 2001, Moretti 2004, Iranzo and Peri 2009), although Ciccone and Peri (2006) and Caselli (2005) find them to be unimportant. By decomposing human capital effects into those of worker education, entrepreneurial/managerial education, and externalities using a unified framework, we try to disentangle different mechanisms.

Third, because we are looking at the regions, we need to consider the mobility of firms, workers, and entrepreneurs across regions, which is presumably less expensive than that across countries. To this end, our model follows the standard urban economics approach (e.g., Roback 1982, Glaeser and Gottlieb 2009) of labor mobility across regions with scarce resources, such as land and housing, limiting universal migration into the most productive regions. This aspect of the model allows us to consider jointly the education coefficients in regional and establishment level regressions. A key benefit of our model of the three channels of influence of education is to reconcile regional and firm-level evidence.

As a first step in the empirical analysis, we use regional data to examine the determinants of regional income in a specification with country fixed effects. The approach follows development accounting, as in Hall and Jones (1999), Caselli (2005), and Hsieh and Klenow (2010). Among the determinants of regional productivity, we consider geography, as measured by temperature (Dell, Jones, and Olken 2009), distance to the ocean (Bloom and Sachs 1998), and natural resources endowments.

[2] We do not consider the role of human capital in shaping the adoption of new technologies. Starting with Nelson and Phelps (1966), economists have argued that human capital accelerates the adoption of new technologies. Strictly speaking, this is an externality across rather than within countries. For recent models of these effects, see Benhabib and Spiegel (1994), Klenow and Rodriguez-Clare (2005), Caselli and Coleman (2006), the most persuasive supporting empirical evidence is Ciccone and Papaioannou (2009).

We also consider institutions, which have been found by King and Levine (1993), DeLong and Shleifer (1993), Hall and Jones (1999), and Acemoglu et al. (2001) to be significant determinants of development. We also examine some measures of culture, such as trust, for which we have data at the regional level, and which have been shown to matter at the national level (Landes 1998, Knack and Keefer 1997), as well as ethnic heterogeneity (Easterly and Levine 1997, Alesina et al. 2003). Last, we consider the effect of average education in the region on its level of development. A substantial cross-country literature points to a large role of education. Barro (1991) and Mankiw, Romer, and Weil (1992) are two early empirical studies; de La Fuente and Domenech (2006) and Cohen and Soto (2007) are recent confirmations. In the cross-country analysis, the effects of education and institutions have been difficult to disentangle empirically (Glaeser et al 2004).

We find that favorable geography, such as lower average temperature and proximity to the ocean, as well as higher natural resource endowments, are associated with higher per capita income in regions within countries. We do not find that culture, as measured by ethnic heterogeneity or trust, explains regional differences. Nor do we find that institutions as measured by survey assessments of the business environment in the World Bank's Enterprise Surveys help account for cross-regional differences within a country. Some institutions or culture are likely to matter only at the national level, but in this case finding large income differences within countries calls for explanations other than culture and institutions. In contrast, differences in educational attainment account for a large share of the differences in per capita income of regions within a country. The within country R-squared in the univariate regression of the log of per capita income on the log of education is about 25 percent; this R-squared is not higher than 8 percent for any other variable.

Acemoglu and Dell (2010) examine sub-national data from North and South America to disentangle the roles of education and institutions in accounting for development. The authors find that

about half of the within-country variation in levels of income is accounted for by education. This is very similar to the Mankiw et al. (1992) estimate that half of the differences in per capita incomes across countries is attributable to education. We confirm the large role of education, but try to go further in identifying the channels. Acemoglu and Dell also conjecture that institutions shape the remainder of the local income differences. We do have regional data on several aspects of institutional quality, but find that their ability to explain cross-regional differences is minimal[3].

We next combine establishment-level and regional data to estimate the determinants of firm productivity. As a first step, we merge our data with World Bank Enterprise Surveys, which provide establishment-level information on sales, labor force, educational level of management and employees, as well as energy and capital use for several thousand establishments in the regions for which we have data. The Surveys tell us about the location of establishments, so we can estimate firm-level productivity across regions as a function of establishment inputs but also regional education. Such micro results may be less vulnerable to the reverse causality concern that income drives education, since there is no reason why firm-level productivity should drive regional education.

Most importantly, the simultaneous use of regional and firm-level data enables us to explore the effects of human capital by combining estimation with calibration. Because education is endogenous in national and regional regressions, scholars have turned to calibration techniques, using Mincerian estimates of private returns to education, to compute the parameters of the production function (e.g., Caselli 2005). We rely on Mincerian estimates of returns to worker education (Psacharopoulos 1994) and on previous research regarding factor shares (e.g., Gollin 2002, Caselli and Feyer 2007, Valentyi and

[3] A recent literature looks at colonial history within countries, and argues that regions that were treated particularly badly by colonizers have poor institutions and lower income many years later (Banerjee and Iyer 2005, Dell 2010). It is surely possible, even likely, that severe institutional shocks have long run consequences because they influence human capital accumulation and institutions in the long run. But to the extent that we have adequate measures of institutional quality, the consequences of such shocks for modern institutions do not appear to add a great deal of explanatory power to understanding cross-regional evidence today.

Herrendorf 2008), but then combine them with coefficient estimates from productivity regressions to calculate the parameters of our production function. We find substantial consistency between regional and firm-level results, as well as plausible estimates of the parameters of the production function.

Specifically, the micro data show that establishments with employees and managers with higher education are more productive, holding capital and energy inputs constant and including country and industry fixed effects. When regional education is added to these regressions, it has a large and positive coefficient. Our calibrations show that worker education, entrepreneurial education, and externalities all contribute to productivity. Crucially, while we find the role of workers' human capital to be in line with standard calibration exercises (e.g., Caselli 2005), our results indicate that focusing on worker education alone enormously underestimate both private and social returns to education. Private returns are extremely high but to a substantial extent are earned by entrepreneurs, and hence might appear as profits rather than wages, consistent with Lucas (1978). Social returns are also high because of externalities, as in Lucas (1988, 2008). In sum, the evidence points to substantial influences of both entrepreneurial human capital and human capital externalities on productivity.

The model has a number of additional empirical implications, and we consider some of them using a third data source, namely establishment censuses for a large number of countries, which also provide geographic identifiers of firm location. We find that higher human capital regions have larger establishments, but also sharply higher rates of participation in the official labor force. These results are broadly consistent with the implications of the Lucas-Lucas model. To better understand some of this evidence, however, one needs to draw the distinction between official and unofficial sectors.

In the next section, we present our model of regional development that organizes the evidence. In section 3, we describe our data in some detail. Section 4 presents the evidence on the role of various factors in accounting for the differences in both national and regional development. Section 5 presents

firm-level evidence to disentangle the channels of through which human capital influences productivity. We combine the regional and the micro estimates to assess the parameters of the model. Section 6 presents some of the census evidence to test additional implications of the model. Section 7 concludes.

II.A Lucas-Lucas spatial model of regional and national income

A country consists of a measure 1 of regions, a share p of which has productivity A_G and a share $1-p$ of which has productivity $A_B < A_G$. We refer to the former regions as "productive" and to the latter regions as "unproductive" and denote them by $i = G, B$. A measure 2 of agents is uniformly distributed across regions. An agent j enjoys consumption and housing according to the utility function:

$$u(c,a) = c^{1-\theta_j} a^{\theta_j}, \tag{1}$$

where c and a denote consumption and housing, respectively. Half the agents are "rentiers," the remaining half are "labourers''. Each rentier is owns 1 unit of housing, T units of land and K units of physical capital (and no human capital). Each labourer is endowed with $h \in \mathrm{R}_{++}$ units of human capital, which is distributed according to a Pareto distribution with support $[\underline{h},+\infty)$ and mean $E(h) = \mu\underline{h}/(\mu-1)$, where $\mu, \underline{h} > 1$. One can view rentiers as the elderly, and labour as the young. Rentiers only maximize consumption income, i.e. they have $\theta_j = 0$; labourers enjoy housing also so they have $\theta_j = \theta \in (0,1)$. This assumption is not substantive but simplifies the characterization of housing market equilibrium.

A labourer can become either an entrepreneur or a worker. By operating in region i, an entrepreneur with human capital h who hires physical capital $K_{i,h}$, land $T_{i,h}$, and workers with total human capital $H_{i,h}$ produces an amount of the consumption good equal to:

$$y_{i,h} = A_i h^{1-\alpha-\beta-\delta} H_{i,h}^{\alpha} K_{i,h}^{\delta} T_{i,h}^{\beta}, \quad \alpha+\beta+\delta<1. \tag{2}$$

As in Lucas (1978), a firm's output increases, at a diminishing rate, in the entrepreneur's human capital *h* as well as $H_{i,h}$, $K_{i,h}$ and $T_{i,h}$. We first consider the Lucas-only model where A_i captures exogenous regional differences, such as institutions or geography. We then extend the analysis to the Lucas-Lucas model, where A_i depends also on regional human capital due to the presence of externalities. Either way, productivity A_i induces human and physical capital to sort across regions.

Rentiers earn their income by renting land and physical capital to firms, and housing to labour (both entrepreneurs and workers). In region *i*, a representative rentier earns $\lambda_i T$ and η_i by renting land and housing, where λ_i and η_i are rental rates. A rentier renting physical capital in region *i* earns $\rho_i K$. Each region's land and housing endowments T and 1 are immobile, while physical capital is fully mobile. Labour earns its income by using its human capital in work or in entrepreneurship. By operating in region *i*, a labourer with human capital h earns either profits $\pi_i(h)$ as an entrepreneur or wage income $w_i \cdot h$ as a worker, where w_i is the wage rate. All labourers, whether they become entrepreneurs or workers, are partially mobile: a labourer moving to region *i* loses φw_i units of income, where $\varphi < \underline{h}$.[4]

The timing of the model works as follows: At *t* = 0, a labourer with human capital h selects the location and occupation that maximize his income. The housing market clears, so houses are allocated to each region's labour. At *t* = 1, entrepreneurs hire land, human, and physical capital. Production is carried out and distributed in wages, land rental, capital rental, housing rental and profits. Consumption takes place.

A spatial equilibrium is a regional allocation $\left(H_i^E, H_i^W, K_i\right)_i$ of entrepreneurial human capital H_i^E, workers' human capital H_i^W, and physical capital K_i such that: a) entrepreneurs hire workers, physical capital, and land to maximize profits, b) labourers optimally choose location, occupation and the fraction

[4] For simplicity, we assume that moving costs are non-rival with the time spent working. This ensures that the total human capital employed by firms in a region, as well as the aggregate income of laborers, do not depend on moving costs.

of income devoted to consumption and housing, and c) capital, labour, land and housing markets clear. Because physical capital is fully mobile, there is a unique rental rate ρ. Since land and housing are immobile, their rental rates λ_i and η_i vary across regions depending on productivity and population. To determine the sorting of labourers into work and entrepreneurship, we must compute regional wages w_i and profits $\pi_i(h_j)$. We proceed in two steps. First, we determine regional output and factor returns at a given allocation $\left(H_i^E, H_i^W, K_i\right)_i$. Second, we solve for the equilibrium allocation. Throughout the analysis, the price of consumption is normalized to one.

The Lucas-only model: production and occupational choice

An entrepreneur with human capital h operating in region i maximizes his profit by solving:

$$\max_{H_{i,h},T_{i,h},K_{i,h}} A_i h^{1-\alpha-\beta-\delta} H_{i,h}^{\alpha} K_{i,h}^{\delta} T_{i,h}^{\beta} - w_i H_{i,h} - \rho K_{i,h} - \lambda_i T_{i,h}, \tag{3}$$

implying that in each region firms employ factors in the same proportion. Since at $\left(H_i^E, H_i^W, K_i\right)_i$ firm j employs a share of entrepreneurial capital h_j / H_i^E, it hires the others factors according to:

$$H_{i,j} = \frac{h_j}{H_i^E} \cdot H_i^W, \quad K_{i,j} = \frac{h_j}{H_i^E} \cdot K_i, \quad T_{i,j} = \frac{h_j}{H_i^E} \cdot T. \tag{4}$$

As in Lucas (1978), more skilled entrepreneurs run larger firms.

Equation (4) implies that the aggregate regional output is given by:

$$Y_i = A_i \left(H_i^E\right)^{1-\alpha-\beta-\delta} \left(H_i^W\right)^{\alpha} K_i^{\delta} T^{\beta}. \tag{5}$$

Equation (5) allows us to determine wages, profits, and capital rental rates as a function of regional factor supplies via the usual marginal product pricing. That is:

$$
\begin{aligned}
w_i &= \frac{\partial Y_i}{\partial H_i^W} = \alpha \cdot A_i \left(H_i^E / H_i^W\right)^{1-\alpha-\beta-\delta} \left(K_i / H_i^W\right)^{\delta} \left(T / H_i^W\right)^{\beta}, \\
\pi_i &= \frac{\partial Y_i}{\partial H_i^E} = (1-\alpha-\beta-\delta) \cdot A_i \left(H_i^W / H_i^E\right)^{\alpha} \left(K_i / H_i^E\right)^{\delta} \left(T / H_i^E\right)^{\beta}, \\
\rho &= \frac{\partial Y_i}{\partial K_i} = \delta \cdot A_i \left(H_i^E / K_i\right)^{1-\alpha-\beta-\delta} \left(H_i^W / K_i\right)^{\alpha} \left(T / K_i\right)^{\beta}.
\end{aligned}
\tag{6}
$$

Thus, profit $\pi_i(h)$ is equal to π_i – the marginal product of the entrepreneur's human capital in region *i* times the entrepreneur's human capital h, namely $\pi_i(h) = \pi_i \cdot h$.

Using Equation (6) we can solve for a labourer's occupational choice. A labourer *j* with human capital h_j chooses to be an entrepreneur when $\pi_i \cdot h_j > w_i \cdot h_j$ and a worker if $\pi_i \cdot h_j < w_i \cdot h_j$. In equilibrium, labourers must be indifferent between the two occupations (i.e., $\pi_i = w_i$), which implies:

$$
H_i^E = \left(\frac{1-\alpha-\beta-\delta}{1-\beta-\delta}\right) \cdot H_i, \quad H_i^W = \left(\frac{\alpha}{1-\beta-\delta}\right) \cdot H_i, \tag{7}
$$

where $H_i = H_i^E + H_i^W$ is total human capital in region *i*. H_i^E increases with the share of the total *private* return to human capital earned by entrepreneurs [i.e. with $(1-\alpha-\beta-\delta)/(1-\beta-\delta)$]. Equation (7) yields the regional allocation of labour from the total quantities of human and physical capital (H_i, K_i).

Equation (7) does not say whether the total amount of entrepreneurial human capital is allocated to a few firms run by very skilled entrepreneurs or to many firms run by less skilled ones. Assuming a fixed income cost of setting up a firm eliminates this indeterminacy – thus determining the number of firms - because now the most skilled labourers become entrepreneurs, as in Lucas (1978). In this case, though, a wedge between the profit and wage rates would arise to compensate the marginal entrepreneur for bearing the fixed cost. Since this wedge complicates the regional allocation problem, we focus on the limiting case where the cost of setting up a firm tends to zero. In this case, there is no wedge between profits and wages, but it is still the case that the most skilled labourers become

entrepreneurs. With density $F_i(h)$ of human capital in region *i*, there is a threshold h_i^E defined as:

$$\int_{h_i^E}^{+\infty} h dF_i(h) = \left(\frac{1-\alpha-\beta-\delta}{1-\beta-\delta} \right) H_i \,, \tag{8}$$

so that in region *i* laborers become entrepreneurs if and only if $h \geq h_i^E$, and the number of firms is 1 – $F_i($ h_i^E). Crucially, the density $F_i(h)$ of human capital in region *i* and the total quantity H_i of human capital in that region are endogenously determined by labour mobility, which we study next.

The Lucas-only spatial equilibrium: consumption, housing and mobility

We consider symmetric spatial equilibria in which all productive regions share the same factor allocation (H_G,K_G), the same wage w_G and rental rates λ_G and η_G, and unproductive regions share the same factor allocation (H_B,K_B), wage rate w_B, and rental rates λ_B and η_B. The uniformity of the capital rental rate ρ across regions pins down the regional allocation of physical capital. By using Equation (7) and the expression for ρ in (6) one finds that the marginal return to capital is constant across productive and unproductive regions if and only if:

$$\frac{K_G}{K_B} = \left(\frac{A_G}{A_B} \right)^{\frac{1}{1-\delta}} \left(\frac{H_G}{H_B} \right)^{\frac{1-\beta-\delta}{1-\delta}} \,, \tag{9}$$

Intuitively, the physical capital stock allocated to productive regions increases in their productivity advantage A_G/A_B and in their relative human capital stock H_G/H_B.

Since the equilibrium allocation of human capital is shaped by labour mobility, we next need to find the utility that labourers obtain from operating in different regions. Labourers maximize their utility in (2) by devoting a share θ of their income to housing and the remaining share $(1-\theta)$ to consumption.

A labourer with human capital *h* living in region *i* earns income $w_i h$ and sets $c = (1-\theta)\cdot w_i h$ and $a = \theta \cdot w_i h / \eta_i$. With a fixed housing supply and only labourers (not rentiers) demanding housing, the housing market in region *i* is in equilibrium when:

$$\frac{\int_{\underline{h}}^{\infty} \theta w_i h dF_i(h)}{\eta_i} = \theta \frac{w_i H_i}{\eta_i} = 1 \tag{10}$$

The housing rental rate is equal to the share θ of labour (worker and entrepreneur) income in the region. As a consequence, the utility of a labourer from staying in his birth region *i* is equal to:

$$u_{w,i}(c,a) = \frac{w_i h}{\eta_i^{\theta}} = \frac{w_i^{1-\theta}}{\theta^{\theta}} \cdot \frac{h}{H_i^{\theta}} . \tag{11}$$

A labourer's utility rises with the wage w_i and falls with regional human capital H_i due to higher rents.

Equation (11) implies that since at the initial human capital allocation wages are higher in productive regions, labourers migrate there from unproductive regions. The resulting increase in human capital in productive regions raises housing costs until the incentive to migrate disappears. Formally, a labourer with human capital h migrates if $w_G^{1-\theta}(h-\varphi)/H_G^{\theta} \geq w_B^{1-\theta} h / H_B^{\theta}$, where φ captures migration costs. This identifies a human capital threshold h_m such that agent *j* migrates if and only if $h_j \geq h_m$. Intuitively, more skilled people have a greater incentive to pay the migration costs because the wage (or profit) gain they experience from doing so is higher. The threshold h_m can be implicitly expressed as:

$$h_m \cdot \left[1 - \left(\frac{w_B}{w_G} \right)^{1-\theta} \left(\frac{H_G}{H_B} \right)^{\theta} \right] = \varphi \ . \tag{12}$$

By exploiting the wage equation in (6) and the equilibrium condition (9), we can rewrite (12) as:

$$h_m \cdot \left[1 - \left(\frac{A_B}{A_G} \right)^{\frac{1-\theta}{1-\delta}} \left(\frac{H_G}{H_B} \right)^{\frac{\beta(1-\theta)+\theta(1-\delta)}{1-\delta}} \right] = \varphi \cdot \tag{13}$$

To pin down the equilibrium, note that the aggregate resource constraint is given by:

$$p \cdot H_G + (1-p) \cdot H_B = E(h), \tag{14}$$

and that the human capital stock employed in a productive region is equal to the initial endowment plus an amount 1/*p* of the total human capital of migrants, namely:

$$H_G = E(h) + \frac{1-p}{p} \int_{h_m}^{+\infty} h \cdot (\mu \underline{h}^{\mu} h^{-\mu-1}) \cdot dh = E(h) \left[1 + \frac{1-p}{p} \left(\frac{\underline{h}}{h_m} \right)^{\mu-1} \right]. \tag{15}$$

By replacing (14) and (15) into (13) we find that the equilibrium is reached when:

$$\frac{\varphi}{\underline{h}} \left(\frac{p}{1-p} \right)^{\frac{1}{\mu-1}} \left(\widetilde{H}_G - 1 \right)^{\frac{1}{\mu-1}} = 1 - \left(\frac{A_B}{A_G} \right)^{\frac{1-\theta}{1-\delta}} \left(\frac{\widetilde{H}_G (1-p)}{1-p\widetilde{H}_G} \right)^{\frac{\beta(1-\theta)+\theta(1-\delta)}{1-\delta}}, \tag{16}$$

where $\widetilde{H}_G = H_G / E(h) > 1$ is the ratio of employment to the initial human capital endowment in the region.

As Figure 1 shows, the left hand side of (16) increases in $\widetilde{H}_G$, the right hand side decreases in $\widetilde{H}_G$.

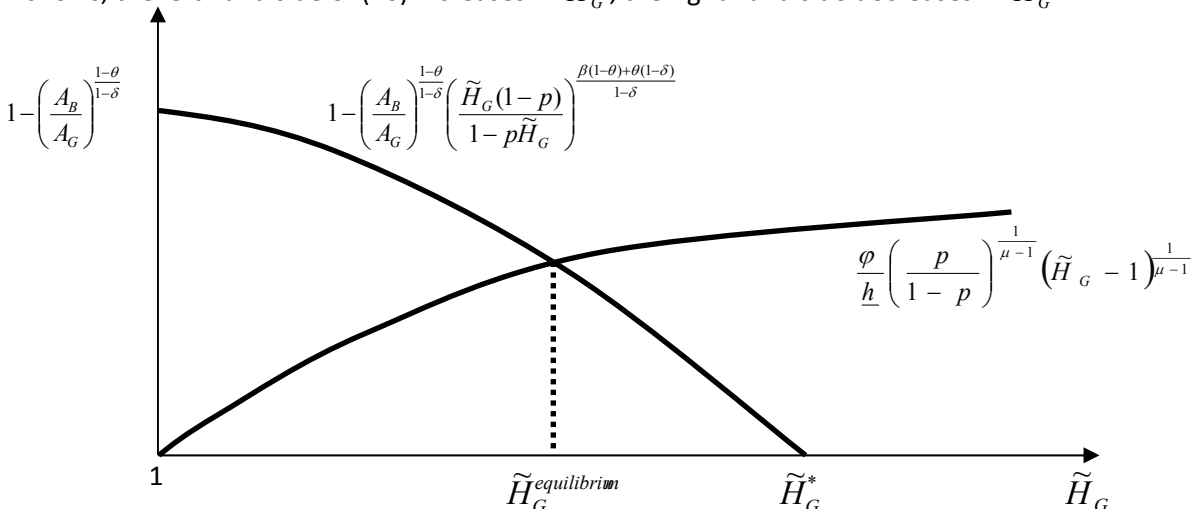


Figure 1: *Equilibrium with mobility costs (i.e. φ > 0)*

There is a unique equilibrium $\widetilde{H}_G^{equilibrium}$, in which productive regions import human capital ($\widetilde{H}_G > 1$) from unproductive ones. Note that the value $\widetilde{H}_G^*$ at which the right hand side of (16) is equal to zero identifies the equilibrium prevailing under perfect mobility, i.e. when $\varphi = 0$. In that case, we have:

$$\widetilde{H}_G^* = \frac{H_G}{E(h)} = \frac{A_G^{\frac{1-\theta}{\beta(1-\theta)+\theta(1-\delta)}}}{E\left[A^{\frac{1-\theta}{\beta(1-\theta)+\theta(1-\delta)}}\right]}, \tag{17}$$

where $E\left[A^{\frac{1-\theta}{\beta(1-\theta)+\theta(1-\delta)}}\right] = pA_G^{\frac{1-\theta}{\beta(1-\theta)+\theta(1-\delta)}} + (1-p)A_B^{\frac{1-\theta}{\beta(1-\theta)+\theta(1-\delta)}}$ is the average of $A^{\frac{1-\theta}{\beta(1-\theta)+\theta(1-\delta)}}$. Under full mobility, productive regions employ human capital in proportion to their relative productivity.

As Figure 1 shows, in the presence of mobility costs, when $\varphi > 0$, the equilibrium $\widetilde{H}_G^{equilibrium}$ settles between the no-mobility equilibrium $\widetilde{H}_G = 1$ and the full mobility equilibrium $\widetilde{H}_G^*$. In equilibrium, for any value $\varphi \geq 0$ wages are higher in more productive regions, $w_G > w_B$, but the housing rental rate is also higher there, $\eta_G > \eta_B$, so that the marginal worker has no incentive to migrate.

To simplify national aggregation, we study the case where mobility costs are negligible, i.e. $\varphi \to 0$, which might be a reasonable approximation in our within country analysis. In this case, we have:

$$H_i = \frac{A_i^{\frac{1-\theta}{\beta(1-\theta)+\theta(1-\delta)}}}{E\left[A^{\frac{1-\theta}{\beta(1-\theta)+\theta(1-\delta)}}\right]} \cdot E(h) \quad K_i = \frac{A_i^{\frac{\beta(1-\theta)+(1-\delta)}{(1-\delta)[\beta(1-\theta)+\theta(1-\delta)]}}}{E\left[A^{\frac{\beta(1-\theta)+(1-\delta)}{(1-\delta)[\beta(1-\theta)+\theta(1-\delta)]}}\right]} \cdot K \cdot \tag{18}$$

Productive regions attract both human and physical capital. Using Equation (18) one can find national output by decomposing regional human capital into its working and entrepreneurial components according to (7). National output is then equal to:

$$Y = \hat{A}\left(H^E\right)^{1-\alpha-\beta-\delta}\left(H^W\right)^{\alpha} K^{\delta} T^{\beta}. \tag{19}$$

Here $\hat{A}$ is a function $\hat{A}(\beta,\delta,\theta,A_G,A_B,p)$ of exogenous parameters. This model describes firm-level output with Equation (2), regional output with Equation (5), and national output with Equation (19).

The Lucas-Lucas spatial equilibrium

We model externalities by assuming that regional total factor productivity is equal to:

$$\widetilde{A}_i = A_i\left(E_i(h)^{\psi} L_i\right)^{\gamma}, \quad \gamma > 0, \psi \geq 1, \tag{20}$$

where $E_i(h)$ is the average level of human capital in region *i* and L_i is the measure of labour in that region. Productivity depends not only on exogenous regional conditions but also on the sorting of human capital. Parameter ψ captures the importance of the quality of human capital: when ψ = 1 only the total quantity of human capital $H_i = \mathrm{E}_i(h)L_\mathrm{i}$ matters for externalities; as ψ becomes larger the quality of human capital becomes relatively more important than quantity. Parameter γ captures the overall importance of externalities.

In our regional and firm level regressions, we employ the flexible specification of Equation (20). To illustrate the determination of national income in the model, we however consider the case ψ = 1, which allows for closed form solutions. We show in the appendix how to characterize the equilibrium for $\psi > 1$. When ψ = 1, we can show that in equilibrium equation (18) becomes:

$$H_i = \frac{A_i^{\frac{1-\theta}{(\beta-\gamma)(1-\theta)+\theta(1-\delta)}}}{E\left[A^{\frac{1-\theta}{(\beta-\gamma)(1-\theta)+\theta(1-\delta)}}\right]} \cdot E(h) \quad K_i = \frac{A_i^{\frac{\beta(1-\theta)+(1-\delta)}{(1-\delta)[(\beta-\gamma)(1-\theta)+\theta(1-\delta)]}}}{E\left[A^{\frac{\beta(1-\theta)+(1-\delta)}{(1-\delta)[(\beta-\gamma)(1-\theta)+\theta(1-\delta)]}}\right]} \cdot K \tag{21}$$

The sorting of human capital across regions depends on the difference $\beta - \gamma$ between the diseconomies of scale due to the presence of a fixed factor (land) and the positive human capital spillovers. The smaller is $\beta - \gamma$, the stronger are the spillovers and the greater is the migration of human (and physical) capital into the more productive regions. To obtain a stable equilibrium, Equation (21) requires that:

$$(\beta - \gamma)(1 - \theta) + \theta(1 - \delta) > 0. \tag{22}$$

Equation (22) says that if spillovers are sufficiently strong to more than offset the presence of a fixed factor in production, namely if $\gamma > \beta$, then the share θ of income spent by labourers on housing must be sufficiently high. When (22) holds, then even though migration increases productivity by creating spillovers, it also increases the housing (and land) rental rates so that after some point migration stops and an equilibrium is reached. When (22) does not hold, then in the only stable equilibrium the whole population migrates into the most productive region. In equilibrium, national output is equal to:

$$Y = \widehat{A} H^{\gamma} \left(H^{E}\right)^{1-\alpha-\beta-\delta} \left(H^{W}\right)^{\alpha} K^{\delta} T^{\beta} , \tag{23}$$

where $\widehat{A}$ is a function $\widehat{A}(\beta, \delta, \theta, A_G, A_B, p)$ of exogenous parameters. More generally, as we show in the Appendix, a sufficient condition for a unique stable equilibrium for any $\psi \geq 1$ is that $(\beta - \psi\gamma)(1 - \theta) + \theta(1 - \delta) > 0$. If this condition is met, then even if we cannot compute national output in closed form, the counterpart of Equation (2) for firm level output becomes:

$$y_{i,j} = A_i E_i (h)^{\psi\gamma} L_i^{\gamma} h_j^{1-\alpha-\beta-\delta} H_{i,j}^{\alpha} K_{i,j}^{\delta} T_{i,j}^{\beta} , \tag{24}$$

while the counterpart of Equation (5) for regional output becomes:

$$Y_i = A_i E_i (h)^{\psi\gamma} L_i^{\gamma} (H_i^{E})^{1-\alpha-\beta-\delta} (H_i^{W})^{\alpha} K_i^{\delta} T^{\beta} . \tag{25}$$

Empirical Predictions of the Model

To obtain from equations (24) and (25) predictions on the role of schooling, we need to specify a link between human capital (which we do not observe) and schooling (which we do observe). We follow the Mincerian approach in which for an individual *j* the link between human capital and schooling is:

$$h_j = \exp\left(\mu_j S_j\right), \tag{26}$$

where $S_j \geq 0$ and $\mu_j \geq 0$ are two random variables (distributed to ensure that the distribution of human capital h_j is Pareto. Relative to standard Mincerian regressions, in Equation (26) the return to schooling μ_i varies across individuals, potentially due to talent. Card (1999) offers some evidence of such heterogeneity. Human capital in region *i* is then equal to:

$$\int_{\underline{h}}^{+\infty} h dF_i(h) = \iint_{S\geq 0,\mu\geq 0} e^{\mu S} g_i(S,\mu) dS d\mu, \tag{27}$$

where $dF_i(h)$ is the density of region *i* labourers with human capital *h* and $g_i(S,\mu)$ is accordingly the density of region *i* labourers having $h = e^{\mu S}$, so that $\iint_{S\geq 0,\mu\geq 0} g_i(S,\mu) dS d\mu = L_i$. Denoting by $\overline{S}_i$ the average observed level of schooling in region *i*, define $\overline{\mu}_i$ as the average Mincerian return in the region's population as the $\overline{\mu}_i$ satisfying the condition $e^{\overline{\mu}_i \overline{S}_i} L_i = \iint_{S,\mu\geq 0} e^{\mu S} g_i(S,\mu) dS d\mu$. This analysis can be exploited to obtain from the regional income Equation (25) an implementable empirical test.

Regional Income Differences

To test Equation (25) we need to specify a regression in terms of observables, which entails regressing regional per capita income on human capital and population (we do not have regional data

on physical capital). From the Equation for ρ in (6) we obtain the condition $K_i = B\, A_i^{\frac{1}{1-\delta}} H_i^{\frac{1-\beta-\delta}{1-\beta}}$ where B > 0 is a constant. By substituting this condition into Equation (25) we find that:

$$\ln(Y_i/L_i) = C + [1/(1-\delta)]\ln A_i + [1+\gamma\psi - \beta/(1-\delta)]\ln E_i(h) + [\gamma - \beta/(1-\delta)]\ln L_i, \quad (28)$$

where C is a constant absorbed by the country fixed effect. Using Equation (27), the observable counterpart of Equation (28) becomes:

$$\ln(Y_i/L_i) = C + [1/(1-\delta)]\ln A_i + [1+\gamma\psi - \beta/(1-\delta)]\,\overline{\mu}_i\,\overline{S}_i + [\gamma - \beta/(1-\delta)]\ln L_i, \quad (29)$$

where $\overline{\mu}_i$ is the previously defined average return to schooling in region *i*. Estimating (29) using OLS implies that the coefficient on average regional schooling should be interpreted as the product of the technological parameter $(1+\gamma\psi - \beta)$ and the nation-wide average Mincerian return $\overline{\mu} = \int_i \overline{\mu}_i di$.

The technological parameter $[1+\gamma\psi - \beta/(1-\delta)]$ subtracts from externalities $\gamma\psi$ an amount increasing in the rent accruing to the rentiers β because, as H_i rises, the presence of fixed land supply progressively reduces the productivity of human capital. Population L_i on the right hand side captures the productive return to increasing regional population (or workforce), which is equal to $[\gamma - \beta/(1-\delta)]$: when population rises, on the one hand the total quality of human capital rises also, but on the other hand productivity falls due to fixed land supply. Our model does not yield specific predictions as to whether γ is above or below β because the stabilizing force comes from the housing rental parameter θ. When $\gamma > \beta/(1-\delta)$ there is a productivity (and thus wage) benefit of increasing migration to productive regions. However, mobility is hindered by the increase in housing rents, consistent with Moretti (2004). We later use the regression results along with calibrations from the literature to shed light on $[\gamma - \beta/(1-\delta)]$. In section 4, we test these implications using regional data, but also control for the determinants of A_i using geographic and other variables.

Firm-Level Productivity

If one regresses firm-level output per worker $y_{i,j}/l_{i,j}$ on all inputs expressed in per worker terms, where $l_{i,j}$ measures the firm's workforce, Equation (24) implies:

$$\ln(y_{i,j}/l_{i,j}) = \ln A_i + (1-\alpha-\beta-\delta)\ln[E_{i,j}(h^E)(l^E_{i,j}/l_{i,j})] + \alpha\ln[E_{i,j}(h^W)(l^W_{i,j}/l_{i,j})] + \delta\ln k_{i,j} +$$

$$+ \beta\ln t_{i,j} + \gamma\ln L_i + \gamma\psi\ln E_i(h), \qquad (30)$$

where $x_{i,j} = X_{i,j}/l_{i,j}$ denote per-worker values, while $l^E_{i,j}/l_{i,j}$ and $l^W_{i,j}/l_{i,j}$ capture the share of a firm's total employment on managerial and non-managerial jobs, respectively.

To test (30), we use Equation (26) to tease out the return to workers' and entrepreneurs' schooling. In region *i*, an entrepreneur with schooling *S* is someone whose return satisfies:

$$e^{\mu S} \geq h_{E,i}, \qquad (31)$$

where $h_{E,i}$ is the human capital threshold for becoming an entrepreneur in region *i*. At a schooling level *S*, the entrepreneurial class includes labourers whose return satisfies $\mu \geq \mu_{E,i}(S) \equiv \ln h_{E,i}/S$. More talented people become entrepreneurs, less talented people with $\mu < \mu_{E,i}(S)$ become workers.

Denote by $\overline{\mu}_{E,ij}$ the Mincerian return of the entrepreneur of firm *j* in region *i*, and by $\overline{\mu}_{W,ij}$ the average Mincerian return of workers total human capital at the average level of schooling $S_{W,i,j}$. Obviously, $\overline{\mu}_{E,ij} > \overline{\mu}_{W,ij}$. The empirical counterpart of Equation (30) then becomes:

$$\ln(y_{i,j}/l_{i,j}) = \ln A_i + (1-\alpha-\beta-\delta)\,\overline{\mu}_{E,ij}\,S_{E,ij} + \alpha\,\overline{\mu}_{W,ij}\,S_{W,ij} +$$

$$(1-\alpha-\beta-\delta)\ln(l^E_{i,j}/l_{i,j}) + \alpha\ln(l^W_{i,j}/l_{i,j}) + \delta\ln k_{i,j} + \beta\ln t_{i,j} + \gamma\ln L_i + \gamma\psi\,\overline{\mu}_i\,S_i, \qquad (32)$$

where $S_{E,ij}$ and $S_{W,ij}$ stand for the average schooling of the firm's entrepreneur and workers, respectively. When implementing (32), the coefficient on the entrepreneur's schooling should be interpreted as the product of entrepreneurs' rents ($1-\alpha-\beta-\delta$) and the average nation-wide Mincerian entrepreneurs' return to education $\overline{\mu}_E$. The coefficient on worker's average schooling should be interpreted as the labour share α times the average Mincerian returns to workers $\overline{\mu}_W$. The coefficient on regional schooling should be interpreted as the product of the externality parameter $\gamma\psi$ and the population-wide average Mincerian return $\overline{\mu}$.

Equation (32) highlights one key property of our analysis, which is to separate the remuneration α of "low human capital" labour, and the remuneration ($1-\alpha-\beta-\delta$) of "high human capital" managerial or entrepreneurial labour. The total remuneration of human capital ($1-\beta-\delta$) is not captured by wage remuneration, for a substantial portion of managerial compensation ($1-\alpha-\beta-\delta$) is likely to take the form of entrepreneurial rents, eschewing the standard labour income categorization.

Concerning externalities, note that insofar as we suitably control for fixed and mobile capital (obtaining coefficients which are a mixture of β and δ) and for the exogenous productivity A_i in Equation (32), regional population obtains a pure externality coefficient of γ while regional schooling obtains the product between the quality adjusted externality $\gamma\psi$ and the average mincerian return $\overline{\mu}$. The firm level regression (32) yields a larger "externality" than would be estimated by subtracting 1 from the (return adjusted) coefficient on schooling and on population in the regional level regression (29). Indeed, while at the level of the firm a marginal increase in H_i or L_i benefits a firm just by the "pure" spillovers effect,

at the regional level an increase in these variables increases output less than the "pure externality," because fixed factors create a region-wide countervailing force.[5]

The Size of Firms and Regional employment

Not only do productive regions employ more human capital (i.e., $H_G > H_B$), but such human capital is of better quality, namely $E_G(h) > E_B(h)$. This is because in our model migrants are more skilled than average.[6] To show this point, Figure 2 below plots the resulting skill distribution in these regions.

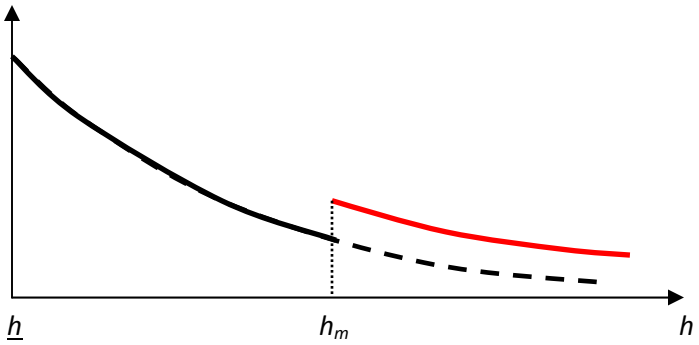


Figure 2: *regional distributions of human capital*

The black bold line is the skill distribution in the unproductive region, which is truncated at h_m. The skill distribution in the productive region coincides with the black bold line until h_m and jumps to the red line for $h > h_m$. Since in productive regions the right tail of the skill distribution is fatter, we have $E_G(h) > E_B(h)$. This fact has important repercussions on the size of firms. In fact, the appendix proves:

[5] Both the regional level Equation (29) and the firm level Equation (32) imply that the average return to human capital should vary across regions. One way to empirically account for such possibility is to run random coefficient regressions. We have performed this analysis and the results change very little (the results on human capital become slightly stronger). We do not report the results to save space.

[6] This should not be viewed as literally saying that all or even the majority of migrants are very skilled labourers, but rather that among the skilled people those who have the highest skill have the greatest incentive to migrate. This is the most important ingredient needed to obtain our main result. One could add to the model a completely unskilled part of the population that provides unskilled labour services, which constitute a different input in production. In this case, unskilled workers may have an incentive to migrate even if middle skilled people do not.

Proposition 1 If p is sufficiently large, there are two thresholds z_1 and z_2 where $z_1 < z_2$ such that for $\left(A_G / A_B\right)^{\frac{1-\theta}{(\beta-\gamma)(1-\theta)+\theta(1-\delta)}} \in (z_1, z_2)$ we have $l_G/f_G > l_B/f_B$. That is, productive regions have: i) a larger average firm, ii) a larger share of workers in the population.

Being rich in very skilled labour, productive regions will have – relative to the unproductive ones – a larger workforce, concentrated into fewer, larger, more productive firms. If there is a sizeable supply of very skilled entrepreneurs, most labourers choose to work for the latter rather than to set up a firm.

III. Data.

Our analysis is based on measures of income, geography, institutions, infrastructure, and culture in up to 110 (out of 193 recognized sovereign) countries for which we found regional data on either income or education. Almost all countries in the world have administrative divisions.[7] In turn, administrative divisions may have different levels. For instance a country may be divided into states or provinces, which are further subdivided into counties or municipalities. For each variable, we collect data at the highest administrative division available (i.e., states and provinces rather than counties or municipalities) or, when such data does not exist, at the statistical division (e.g. the Eurostat NUTS in Europe) that is closest to it. Because we focus on regions, and typically run regressions with country fixed effects, we do not include countries with no administrative divisions in the sample.

The reporting level for data on income, geography, institutions, infrastructure, and culture differs across variables. GDP and education are typically available at the first-level administrative division (i.e., states and provinces). In contrast, GIS geo-spatial data on geography, climate, and

[7] The exceptions are Cook Islands, Hong Kong, Isle of Man, Macau, Malta, Monaco, Niue, Puerto Rico, Vatican City, Singapore, and Tuvalu.

infrastructure is typically available for areas as small as 10 km^2. Finally, survey data on institutions and culture are typically available at the municipal level. In our empirical analysis, we aggregate all variables for each country to a *region* from the most disaggregated level of reporting available.[8] To illustrate, we have GDP data for 27 first-level administrative regions in Brazil, corresponding to its 26 states plus the Federal District, but survey data on institutions for 248 municipalities. For our empirical analysis, we aggregate the data on institutions by taking the simple average of all observations for establishments located in the same first-level administrative division. Similarly, we aggregate the GIS geo-spatial data on geography, climate, and infrastructure at the first-administrative level using the Collins-Bartholomew World Digital Map.

The final data set has 1,569 regions in 110 countries: (1) 79 countries have regions at the first-level administrative division; and (2) 31 countries have regions at a more aggregated level than the first-administrative level because one or several variables (often education) are unavailable at the first-administrative level. For example, Ireland has 34 first-level divisions (i.e. 29 counties and 5 cities), but publishes GDP per capita data for 8 regions and education for 2 regions. Thus, we aggregate all the Irish data to match the 2 regions for which education statistics are available. Appendix A identifies the reporting level for the regions in our dataset. As noted earlier, all countries have administrative divisions (although 31 countries in our sample report statistics for statistical regions). The principal

[8] We used a variety of aggregation procedures. Specifically, we computed population-weighted averages for GDP per capita and years of schooling. We computed regional averages for temperature, precipitation, distance to coast, travel time, and soil characteristics by first summing the (average) values of the relevant variable for all grid cells lying within a region and then dividing by the number of cells lying within a region. We computed regional averages for the density (e.g., power lines) and natural resources variables (oil and gas) by first summing the relevant variable for all grid cells within a region and then dividing by the region's population. We averaged the responses within a region for all the variables from the Enterprise and World Value Surveys. We sum up the number of unique ethnic groups and computed the probability that people within a region speak the same language based on the total number of people in each "language" area.

constraint on the sample is the availability of human capital data. Of course, all countries have periodic censuses and thus have sub-national data on human capital, but these data are hard to find.

Figure 3 presents the 1,569 regions in our sample. It shows that coverage is extensive outside of North and sub-Saharan Africa. Sample coverage is strongly related to a country's surface area, presumably because very small countries do not report regional data. For example, the smallest country in our dataset is Lebanon (10,400 km^2), leaving out of our sample some very prosperous countries such as Luxembourg (2,590 km^2) and Singapore (699 km^2). Among countries ranked by their surface area, we only have data for 14% of the first (smallest) 50 countries, 44% of the first 100 countries, and 53% of the first 150 countries. Similarly, sample coverage rises with the absolute level of GDP but not with GDP per capita. For example, we have data for 18% the first 50 countries ranked in terms of GDP in 2005, 38% of the first 100 countries, and 49% of the first 150 countries. The comparable figures for countries ranked on the basis of GDP per capita are 52%, 57% and 57%, respectively. Since sample coverage rises with GDP, it turns out that the countries in our sample account for 97% of world GDP in 2005.

Our final dataset has regional income data for 107 countries in 2005, drawn from sources including National Statistics Offices and other government agencies (42 countries), Human Development Reports (36 countries), OECDStats (26 countries), the World Bank Living Standards Measurement Survey (Ghana and Kazhakstan), and IPUMS (Israel). When regional income data for 2005 is missing, we use log-linear interpolation based on as much data as it is available for the period 1990-2008 or, when interpolation is not possible, the closest available year.[9] Our measure of regional income per capita is typically based on value added but we use data on income (6 countries), expenditure (8 countries), wages (3 countries), gross value added (2 countries), and consumption, investment and government expenditure (1 country) to fill-in missing values. We measure regional GDP in current purchasing-

[9] We are missing regional GDP per capita for Bangladesh and Costa Rica and national GDP per capita in PPP terms for Cuba.

power-parity dollars as we lack data on regional price indexes. To ensure consistency with the national GDP figures reported by World Development Indicators, we adjust regional income values so that -- when weighted by population-- they total GDP at the country level. Not surprisingly, adjustments exceeding 20% were necessary in 19 out of the 20 countries for which use GDP proxies rather than actual GDP. Adjustments exceeding 20% were also necessary in 13 countries (Democratic Republic of Congo, Lebanon, Lesotho, Madagascar, Malaysia, Nepal, Niger, Philippines, Senegal, Swaziland, Syria, Uganda, and Venezuela) where our GDP data are in real terms.

We compute regional income per capita using population data from *Thomas Brinkhoff: City Population*, which collects official census data as well as population estimates for regions where official census data are unavailable.[10] We adjust these regional population values so that their sum matches the country's population in the World Development Indicators database. This adjustment exceeds 10% for 6 countries: Bangladesh (+13%), Benin (+11%), Democratic Republic of Congo (-10.5%), Gabon (-25%), Swaziland (+16%) and Uzbekistan (-22%).

In addition to GDP per capita, we also examine two other dimensions of regional economic development. First, we gather data on the number of manufacturing and service establishments as well as on their employment for up to 1,068 regions in 69 countries from economic censuses (62 countries) and official business directories (7, mostly OECD, countries). Note that both censuses and directories track establishments rather than firms. This distinction is relevant for large firms as we wouldn't want to allocate, for example, Wal-Mart's 2.2 million world-wide employees to Arkansas. Economic censuses are carried out periodically (e.g. every 10 years) while business directories are continuously updated.

[10] We also used data from OECDStats (for Denmark, Greece, Ireland, Italy, and the UK) and the National Statistics Office of Macedonia.

Critically, they both cover establishments that are registered with the tax authorities, and largely miss the informal economy. Appendix B provides further details regarding our census data.

Second, we examine productivity and its determinants using establishment-level data from the Enterprise Survey for as many as 53,957 establishments in 82 countries and 539 of the regions in our sample.[11] We collect operating data on sales, cost of raw materials, cost of labor, cost of electricity, and the cost of communications. We also collect data on the book value of property, plant, and equipment. Critically, the Enterprise Survey keeps track of the highest educational attainment of the establishment's top manager as well as of its workers. Finally, we collect the two-digit ISIC code (e.g., food, textiles, chemicals, etc) of the establishments in our sample. Like the economic census and business registry data, the Enterprise Survey data only covers registered establishments. A limitation of the Enterprise Survey data is that it largely excludes OECD countries (Ireland and Mexico are the exceptions).

We relate regional economic development to five sets of potential determinants: (1) geography, (2) education, (3) institutions, (4) infrastructure, and (5) culture. To narrow down the list of candidate variables, we restrict attention to variables that are available at the regional level for at least 40 countries and 200 regions.

We use three measures of geography and natural resources obtained from the WorldClim database, which are available for all regions of the world. They include the average temperature during the period 1950-2000, the (inverse) average distance between the cells in a region and the nearest coastline, and the estimated volume of oil production and reserves in the year 2000.[12]

[11] The Enterprise Survey data was collected between 2002 and 2009. When data from the Enterprise Survey for one of the countries in our sample are available for multiple years, we use the most recent one.

[12] The results in the paper are robust to controlling for the standard deviation of temperature, the average annual precipitation during the period 1950-2000, the average output for multiple cropping of rain-fed and irrigated

We gather data on the educational attainment of the population 15 years and older for 106 countries and 1491 regions from EPDC Data Center (55 countries), Eurostat (17 countries), National Statistics Offices (27 countries) and IPUMS (7 countries).[13] We collect data on school attainment during the period 1990-2006 and use data for the most recently available period. We compute years of schooling following Barro and Lee (1993). Specifically, we use UNESCO data on the duration of primary and secondary school in each country and assume: (a) zero years of school for the pre-primary level, (b) 4 additional years of school for tertiary education, and (c) zero additional years of school for post-graduate degrees. We do not use data on incomplete levels because it is only available for about half of the countries in the sample. For example, we assume zero years of additional school for the lower secondary level. For each region, we compute average years of schooling as the weighted sum of the years of school required to achieve each educational level, where the weights are the fraction of the population aged 15 an older that has completed each level of education.

To illustrate these calculations consider the Mexican state of Chihuahua. The EPDC data on the highest educational attainment of the population 15 years and older in Chihuahua in 2005 shows that 4.99% of the that population had no schooling, 13.76% had incomplete primary school, 22.12% had complete primary school, 5.10% had incomplete lower secondary school, 23.04% had complete lower secondary school, 17.94% had complete upper secondary school, and 13.05% had complete tertiary school. Next, based on UNESCO's mapping of the national educational system of Mexico, we assign six years of schooling to people who have completed primary school and 12 years of schooling to those that have completed secondary school. Finally, we calculate the average years of schooling in 2005 in Chihuahua as the sum of: (1) six years times the fraction of people whose highest educational

cereals during the period 1960-1996, the estimated volume of natural gas production and reserves in year 2000, and dummies for the presence of various minerals in the year 2005.

[13] Appendix C provides further details regarding data sources for educational attainment data.

attainment level is complete primary school (22.12%), incomplete lower secondary (5.1%), or complete lower secondary school (23.04%); (2) 12 years times the fraction of people whose highest attainment level is complete upper secondary school (17.94%); and (3) 16 years times the fraction of people whose highest attainment level is complete tertiary school. Accordingly, we estimate that the average years of schooling of the population 15 and older in Chihuahua in 2005 is 7.26 years (=6*0.5026+12*0.1794+16*0.1305).

We compute years of schooling at the country-level by weighting the average years of schooling for each region by the fraction of the country's population 15 and older in that region. The correlation between this measure and the number of years of schooling for the population 15 years and older in Barro and Lee (2010) is 0.9. For the average (median) country in our sample, the number of years of schooling in Barro and Lee (2010) is 8.18 vs. 6.88 in ours (8.56 vs. 6.92 years). Two factors could potentially explain why the Barro-Lee dataset yields a higher level of educational attainment than ours: (1) Barro-Lee captures incomplete degrees while we do not; and (2) education levels have increased rapidly over time but some of our educational attainment data is stale (e.g. for 14 countries our educational attainment data is for the year 2000 or earlier). To make the Barro and Lee (2010) measure of educational attainment more comparable to ours, we make two adjustments to their data. First, we apply our methodology to the Barro-Lee dataset and compute the level of educational attainment in 2005. After this first adjustment, the level of educational attainment computed with the Barro-Lee dataset for the average (median) country in our sample drops to 7.07 (7.23). Second, we apply our methodology to the Barro-Lee dataset but –rather than use data for 2005 -- use figures for the year that best matches the year in our dataset. After this second adjustment, the level of educational attainment using the Barro-Lee dataset for the average (median) country in our dataset drop further to 6.95

(7.22).[14] Since most of our results are run with country-fixed effects, country-level biases in our measure of human capital do not affect our results.[15]

We gather data on seven measures of the quality of institutions from the Enterprise Survey and the Sub-national Doing Business Reports. The Enterprise Survey covers as many as 80 of the countries and 410 of the regions in our sample.[16] The Enterprise Survey asked business managers to quantify: (1) informal payments in the past year (percent of sales spent in informal payments by a typical firm in the respondent's industry), (2) the number of days spent in meeting with tax authorities in the past year, (3) the number of days without electricity in the previous year, and (4) security costs (cost of security equipment, personnel, or professional security service as a percentage of sales). The Enterprise Survey also asks managers to rate a variety of obstacles to doing business, including: (1) access to land, and (2) access to finance.[17] For each of these obstacles to doing business, we keep track of the percentage of the respondents that rate the item as a moderate, a major, or a very severe obstacle to business. The final Enterprise Survey variable that we examine is the perception of government predictability

[14] After the second adjustment, there are 5 countries (i.e., Great Britain, Poland, Switzerland, Syria, and Uruguay) for which our educational attainment numbers remain 25% or more above the adjusted Barro-Lee numbers, and 12 countries (i.e., Armenia, Bangladesh, Benin, Bolivia, Cambodia, Honduras, Laos, Morocco, Niger, Pakistan, Senegal, and Sri Lanka) for which our numbers remain 25% or more below the adjusted Barro-Lee numbers. In all but two of these 17 cases (Great Britain and Poland are the two exceptions), data sources differ (our data for these two countries comes from household or individual surveys and theirs from national censuses). For Great Britain we have 12.14 years of schooling, as does the OECD, while Barro-Lee has 9.21. For Poland, we have 11.15 years of schooling while Barro-Lee has 9.65 and the OECD has 10.55.

[15] Results for our cross-country regressions are qualitatively similar if we use educational attainment from Barro-Lee (2010) rather than the population-weighted average of regional values.

[16] The main reason why we have fewer regions with measures of institutions than regions with productivity data is because we imposed a filter of a minimum of 10 establishments answering the particular institutions question. The rest of the discrepancy in the number of regions is because some questions about institutions were not included in the survey for some countries.

[17] From the Enterprise Survey, we also assembled data on the number of days in the past year with telephone outages, the percentage of sales reported to the tax authorities, and the confidence that the judicial system would enforce contracts and property rights in business. Results for these variables are available upon request.

(measured as the percentage of respondents who tend to agree, agree in most cases, or fully agree that government officials' interpretations of regulations are consistent and predictable).

To make sure that our results on the importance of institutions are not driven by measurement error, we also gather objective measures of the quality of institutions from the Sub-national Doing Business Reports, which are available for 19 countries and 180 regions in our sample. We focus on the number of procedures and their cost in four areas: starting a new business, enforcing contracts, registering property, and dealing with licenses. Interestingly, variation in the cost of regulation swamps the variation in the number of procedures. For example, there is no variation in the number of steps required to enforce a contract in the 30 Chinese cities tracked by the Sub-National Doing Business Report. However, the estimated time to enforce a contract ranges from 112 days in the city of Nanjing (Jiangsu) to 540 days in the city of Changchun (Jilin). As it turns out, results using objective measures of institutions are qualitatively similar to the results using subjective measures that we have described.

We use two measures of infrastructure. The first is the density of power lines in 1997 from the US Geological Survey Global GIS database.[18] The second measure is the average estimated travel time between cells in a region and the nearest city of 50,000 people or more in the year 2000 from the Global Environment Monitoring Unit. Both measures of infrastructure are available for all regions of the world.

Cultural variables are the last set of potential determinants of regional income that we examine. We gather two proxies for cultural values and attitudes from the World Value Survey for as many as 75 of our sample countries and 745 of our regions.[19,20] The first survey measure is the percentage of

[18] Results using other density measures of infrastructure (e.g. air fields, highways, and roads) also available on the US Geological Survey Global GIS database are qualitatively similar.

[19] We set to missing World Value Survey data for five countries (France, Japan, Philippines, Russia, and the United States) because they are only available at a very coarse level.

respondents in each region that answer that “most people can be trusted” when asked whether "Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?" The second measure is a proxy for civic values based on whether each of the following behaviors "can always be justified, never be justified or something in between.": (a) "claiming government benefits which you are not entitled to"; (b) "avoiding a fare on public transport"; (c) "cheating on taxes if you have the chance"; and (d) "someone taking a bribe" (Knack and Keefer, 1997).[21] In addition, we gather two measures of fractionalization for up to 1,568 regions and 110 of our sample countries. The first one is simply the number of ethnic groups that inhabited each region in 1964. The second one is the probability that a randomly chosen person in a region shares the same mother language with a randomly chosen people from the rest of the country in 2004.

Finally, in addition to running regressions using regional data, we examine GDP per capita at the country level, which come from World Development Indicators. All the other country-level variables in the paper are computed based on our regional data rather than drawn from primary sources. Specifically, the country-level analogs of our regional measures of education, geography, institutions, public goods, and culture are the area- and population-weighted averages of the relevant regional variables, as appropriate.

---

[20] The World Value Survey was collected between 1981 and 2005. When data from the World Value Survey for one of the countries in our sample are available for multiple years, we use the most recent one.

[21] We also examined proxies for confidence in various institutions (government, parliament, armed forces, education, civil service, police, and justice), for what is important in people’s lives (family, friends, leisure, politics, work, and religion) as well as for characteristics valued in children (determination, faith, hard work, imagination, independence, obedience, responsibility, thrift, and unselfishness). Moreover, we also examined proxies for broad cultural attitudes with regards to authority (percent who think that one must always love and respect one’s parents regardless of their qualities and faults), tolerance for other people (percent who select tolerance and respect for other people as an important quality for children to learn), and family (percent who think that parents have a duty to do their best for their children even at the expense of their own well-being). Finally, we examined the percentage of respondents that participate in professional and civic associations. The results for these variables are qualitatively similar to the results for the WVS variables that we discuss in the text.

Table 1 summarizes our data. For each variable we examine in the regional regressions, it shows the number of regions for which we have the information, the number of countries these regions are in, the median and the average number of regions per country, and the median range and standard deviation within a country. The data show substantial income inequality among regions within a country. On average, the ratio of the income in the richest region to that in the poorest region is 4.41. This ratio is 3.74 in both Africa and Europe, 4.60 in North America, 5.61 in South America, and 5.63 in Asia. The country with the highest ratio of incomes in the richest to that in the poorest region is Russia (43.3); the country with the lowest ratio is Pakistan (1.32). Interestingly, this ratio is 5.16 for the United States, 2.59 for Germany, 1.93 for France, and 2.03 for Italy. These facts are noteworthy in part because Italy has attracted enormous attention because of differences in income between its North and its South, usually attributed to culture. As it turns out, Italian regional inequality is not particularly high. It is also interesting to note that regional inequality of incomes within a country, as measured by the standard deviation of the logarithm of per capita incomes, declines with income, perhaps because richer countries have more equalizing policies (Figure 4).

There is likewise substantial inequality in education among regions within a country. On average, the ratio of educational attainment in the richest region to that in the poorest region is 1.80. This ratio is 2.71 for Africa, 1.68 for Asia, 1.16 for Europe, 1.33 for North America, and 1.81 for South America. The highest ratio is in Burkina Faso (14.66), where education is 0.22 in the Sahel region and 3.20 in the Centre region. The lowest ratio is 1.05 in Ireland. One striking fact in this data is the much great regional equality in the distribution of education in the richer than in the poorer countries. Figure 5 presents the evidence for the relationship between standard deviation of education levels in a country and its per capita income. Such tendency to equality might follow from the more uniform educational policies in richer countries, and may account for greater regional income equality in the richer countries.

The patterns of inequality among regions within countries is interesting for some of the other variables as well. Table 1 shows that differences in endowments, such as temperature and distance to coast, are small, which suggests that these variables will have difficulty in explaining regional differences in per capita income. Density of power lines and travel time to the next big city varies a great deal across regions, suggesting that urban theories of development might be helpful in explaining regional inequality. There is also considerable variation across regions in the estimates of the quality of institutions, which suggests that, at least in principle, there is a regional aspect to institutional quality that could relate to differences in economic development.

IV. Accounting for National and Regional Productivity.

In this section, we present cross-country and cross-region evidence on the determinants of productivity. We present national regressions only for comparison. These regressions are difficult to interpret in our model because it is not possible to express national output in closed form. More importantly, the problem of endogeneity of education is particularly severe in the national context, which of course turned some scholars to calibration. In Section V, we interpret the coefficients in the context of regional and firm level regressions. With respect to regional income, our benchmark is Equation (29). As already mentioned, we have measures of average education at the regional level, but we do not have either national or regional data on physical capital (except for public infrastructure) or other inputs, so these variables only appear in the firm-level regressions in Section V.

Table 3 presents our basic regional results in perhaps the most transparent way. The table reports the results of univariate regressions of regional GDP per capita on its possible determinants, all with country fixed effects. Such specifications are loaded in favor of each variable seeming important since it does not need to compete with any other variable. We report both the within country and

between countries R-squareds of these regressions. The first row presents the main result: education explains 58% of between country variation of per capita income, and 38% of within country variation of per capita income. Although several other variables explain a significant share of between country variation, none comes close to education in explaining within country variation in income per capita.

Starting with geographical variables, temperature and inverse distance to coast – taken individually – explain 27 and 13 percent of between country income variation, but 1 and 4 percent respectively of within country variation. Oil explains a trivial amount of variation at either level. Turning to institutions, some of the variables, such as access to finance or the number of days it takes to file a tax return, explain a considerable share of cross-country variation, consistent with the empirical findings at the cross-country level such as King and Levine (1993) or Acemoglu et al. (2001), but none explains more than 2 percent of within country variation of per capita incomes. Indicators of infrastructure or other public good provision do slightly better: on their own many explain a large share of between country variation, while density of power lines and travel time account for up to 7% of within country variation. These variables are obviously highly endogenous, and still do much worse than education. Some of the cultural variables account for a substantial share of between country variation, none account for much of within country variation. Of course, culture might operate at the national rather than the sub-national level, although we note that much of the research on trust focuses on regional rather than national differences (e.g., Putnam 1993). After presenting the regression results, we try to explain why some of these variables do so poorly.

Tables 4 through 6 show the multivariate regression results at the national and regional level. Table 4 presents the baseline regressions of national (Panel A) and regional (Panel B) per capita income on geography and education, controlling in some instances for population or employment, as suggested by our model. At the country level, temperature, inverse distance to coast, and oil endowment are all

highly statistically significant in explaining cross-country variation in incomes, and explain an impressive 50% of the variance. Education is also statistically significant, with a coefficient of .25, raising the R-squared to 63%. Note that oil comes in positive and highly statistically significant.

As Panel B shows, the coefficients on geography and education continue to be significant at the regional level, although, critically, the within country R-squared is much higher for education than for the geographic variables. The coefficient on regional labour force is now positive and statistically significant, and ranges from .01 for population to .07 for employment. The coefficient on education is around .27; we return to the interpretation of this coefficient in Section V.

Table 5 presents country-level regressions with measures of institutions (Panel A) and of infrastructure and culture (Panel B) added to the specification in Table 4. Education remains highly statistically significant in each specification, and its coefficient does not fall much. At the country level, only the logarithm of tax days is statistically significant. The last two rows of Table 5 show the adjusted R-squared of each regression if we omit the institutional (or infrastructure or cultural) variable, as well as the adjusted R-squared if we omit education. Dropping education sharply reduces explanatory power, while the only institutional variable that adds explanatory power is the logarithm of tax days.

Table 6 presents the corresponding results at the regional level. The education coefficient is slightly higher than in Table 4, and is highly significant, as illustrated in Figure 6. Institutional variables are almost never significant, and their incremental explanatory power is tiny. We find a small adverse effect of ethnic heterogeneity on income at the regional level, although the incremental explanatory power of all the institutional and cultural variables is small[22].

[22] We have tested the robustness of these results using data on regional luminosity instead of per capita income (see Henderson, Storeygard, and Weil 2009). The results are highly consistent with the evidence we have described, both with respect to the importance of human capital, and the evidence of relative unimportance of other factors, in accounting for cross-regional differences.

What are some of the possible explanations of the low explanatory power of institutions, keeping in mind that endogeneity of institutions should if anything raise the coefficients? It is possible that we have inapprpropriate measures of institutions, although the measures we have are commonly considered to be relevant to economic outcomes. It is also possible that the measures from Enterprise Surveys are particularly noisy, although one should remember that these are surveys of managers who should be particularly focused on institutional constraints. In general, such subjective assessments correlate much better with measures of development than objective measures of institutions (Glaeser et al. 2004). It is also likely that at least some institutions only matter at the national level, if, for example, the critical-to-development business activity is concentrated in the capital.

To shed further light on these issues, Table 7 presents at the national level (we have no regional data) regressions in the same format as Table 5, but using standard measures of institutions, including autocracy, constraints on the executive, expropriation risk, proportional representation, and corruption. Except for proportional representation, all of these variables are highly statistically significant in these specifications. However, with the exception of expropriation risk and corruption, both of which are highly endogenous, the incremental explanatory power of institutional variables is minimal, and in most cases much smaller than the incremental explanatory power of education. Perhaps the more important point is that Enterprise Surveys do not cover rich countries. If we run the regressions in Table 7 for the 72 countries with data on informal payments, we find that proportional representation is insignificant, autocracy and executive constraints are significant at only 10% level, expropriation risk is significant at the 5% level, and corruption is significant at the 1% level. Critically, the value of estimated coefficients falls, rather than standard errors rising. Our bottom line is that the weakness of institutional variables results in part from different (and possibly but not definitely inferior) data, and in part from the focus on poorer countries, for which institutional variables indeed matter less.

We have previously indicated that, because of endogeneity problems, we cannot give a causal interpretation to the large education coefficients, which appear to come through with a similar magnitude across a range of specifications. We next present our micro evidence and combine it with calibration results to interpret the regional and firm-level coefficients in a unified framework.

V. Firm-Level Evidence.

In Table 8, we turn to the micro evidence and estimate essentially Equation (32). We use the Enterprise Survey data described in Section III[23]. In our first two specifications, we measure the log of per capita output $y_{i,j}/l_{i,j}$ by using the log of establishment sales per worker. We also run parallel regressions with the log of average wages paid by the establishment (which in our Cobb-Douglas production function correspond to a constant fraction of output) as a dependent variable in alternative specifications. We measure capital (which include both land $t_{i,j}$ and physical capital $k_{i,j}$) by the log of property, plant and equipment per employee. As an alternative, we proxy for capital by the log of expenditure on energy per employee. We also use the log of the number of employees, which is a proxy of $l_{i,j}$, to control for the share of the entrepreneur's labour $l^E_{i,j}/l_{i,j}$ and of the workers' labour $l^W_{i,j}/l_{i,j}$. Indeed, assuming that each firm has only one entrepreneur we have $l^E_{i,j}/l_{i,j} = 1/l_{i,j}$ and $l^W_{i,j}/l_{i,j} = (l_{i,j}-1)/l_{i,j}$. Unfortunately, the regression coefficient of the log of employees is not susceptible of a clean interpretation in terms of technological parameters.

Most important, to trace out the effects of human capital, we include the years of schooling of the manager S_E, the years of schooling of workers S_W, and the average years of schooling in the region S_i. As we explained in Section 2, the Mincer model of the relationship between education and human

[23] The Enterprise Survey questionnaire varies from country to country. Data on the cost of raw materials used in production is available for roughly 23,000 establishments, i.e. roughly half the number of observations than wages per employee. As an alternative productivity measure, we computed value added as sales net of wages and raw material. The correlation between value added per employee and sales per employee is 0.93. The correlation between value added per employee and wages per employee is 0.95.

capital implies that schooling should enter the specification in levels, rather than in logs. Accordingly, the regression coefficients of the schooling variables should respectively capture parameters $(1-\alpha-\beta-\delta)$ $\overline{\mu}_{E,i}$, $\alpha\overline{\mu}_{W,i}$ and $\gamma\psi\overline{\mu}_i$ in Equation (32). To capture scale effects in regional externalities, we control for the log of the region's population L_i. The regression coefficient on this variable should capture γ in Equation (32). In Table 8, we do not control for proxies of regional productivity A_i, but we do so in the robustness table 8A by including geography, and some indicators of regional institutions and infrastructure as independent variables. Finally, every specification includes country and industry fixed effects, using dummies for 16 industries. All standard errors are clustered at the regional level.

Beginning with Table 8, in the sales per employee specification, the coefficient on energy per employee is .36, while that on capital per employee is .33, giving us a relatively precise estimate of capital share of about .35. These coefficients, however, are closer to .2 in the average wage specifications. The coefficient on worker schooling is about .035 in the productivity specification, but closer to .015 in the wage specifications. The coefficient on management schooling is also about .035 in the productivity specification, but falls to about .02 in the wage specification. The coefficient on the log employment (firm size) is about .1 or .15 in the specifications with energy, but zero in specifications with capital. We map the various coefficients to statements about the production function below. Even controlling for all these inputs, we find consistent evidence of large effects on productivity from regional factors. The coefficient on regional schooling is between .06 and .1. The coefficient on regional population is also fairly consistently positive at around .1 in productivity specifications, but closer to 0 in wage specifications.

So what do these coefficients mean in light of our model, and how do they fit with the work in development accounting? To answer this question, we mainly focus on the firm level regression of table 8, which is the specification closest to Equation (32), but also exploit the regional regressions of Table 6.

Columns (1)-(4) of Table 8 suggest that the coefficient on workers' average schooling is about 0.035, which in our model implies $\alpha \overline{\mu}_W$ = 0.035. If we take the standard calibration for the U.S. labour share α = 0.6, we back out an average Mincerian return of workers equal to $\overline{\mu}_W$ = 0.058, which is just slightly lower than the ballpark 0.06-0.1 of micro evidence on workers' Mincerian returns (Psacharopoulos 1994). If we calibrate α = 0.55 to capture the fact that in developing countries the labour share tends to be lower than in the U.S. (also because part of labour income goes to self employment, Gollin 2002), we obtain $\overline{\mu}_W$ = 0.064, which is the value we stick to. This finding indicates that our firm level productivity regressions help reduce identification problems, at least as far as firm-level variables are concerned.

The regressions also point to an overall capital share (considering energy or equipment) roughly equal to 0.35. In our model, this captures the income share $\delta + \beta$ going to K and T which leaves – under constant returns – a share of 0.1 going to entrepreneurial rents. That is, $(1-\alpha-\beta-\delta)$ = (1 – 0.55 – 0.35) = 0.1. Since the estimated coefficient on managerial education roughly implies $(1-\alpha-\beta-\delta)\overline{\mu}_E$ = 0.035, our results are most consistent with a Mincerian return $\overline{\mu}_E$ equal to 35% for entrepreneurs. This preliminary assessment suggests that a neglected but critical channel through which schooling and human capital affect productivity is via entrepreneurial inputs. Individuals selected into entrepreneurship appear to be vastly more talented than workers, driving up productivity. Of course, entrepreneurial talent may be more important than entrepreneurial schooling in explaining this finding. Our analysis cannot adequately address this issue (which would require better data and an endogenous determination of the connection between schooling and talent). Our analysis is, nevertheless, sufficient to identify a critical role of management and entrepreneurship in determining productivity.

The enormous returns to entrepreneurial education, compared to the modest returns to worker education, might explain the problem that the previous literature on development accounting has

experienced with the Mincer regressions (Caselli 2005, Hsieh and Klenow 2010): the returns to the education of labor are indeed low unless a worker becomes an entrepreneur. Entrepreneurial returns might not be measured in surveys seeking to capture returns to education.

Next, consider the role of externalities. The coefficient on population in Table 8, which is roughly equal to 0.09, suggests that γ is also about 0.09. This assessment is quite consistent with the regional regressions in Table 4, where the coefficient on population is positive and roughly equal to 0.01. In the context of the regional regression (29), this implies that $\gamma - \beta/(1-\delta) = 0.01$. Given that γ = 0.09 and $\beta + \delta = 0.35$, this condition yields β to be roughly 0.05, which is in the ballpark of the land share estimated from income accounts (Valentiny and Herrendorf 2008). In sum, β = 0.05 and γ = 0.09 are roughly consistent with both firm level and regional regressions.

The coefficient on regional schooling in the firm-level regressions of Table 8, roughly equal to 0.065, implies (given γ = 0.09) that $\psi\overline{\mu}$ = 0.72. This leaves it indeterminate whether schooling externalities are mainly due to the population-wise Mincerian return $\overline{\mu}$ or to the multiplier ψ that captures the importance of average schooling for social interactions and thus for externalities. One way to resolve this indeterminacy is once again to exploit the regional regressions. According to Equation (29), in these regressions the coefficient on schooling is equal to $[1+ \gamma\psi - \beta/(1-\delta)]\overline{\mu}$. In Table 4, this coefficient is about 0.27. Since we have already established that γ = 0.09 and $\beta/(1-\delta)$ = 0.08 are reasonable estimates, we are left with two equations with two unknowns, namely $(1+ 0.09\psi - 0.08)\overline{\mu}$ = 0.27 and $\psi\overline{\mu}$ = 0.72. These equations indicate an average population-wide Mincerian return $\overline{\mu}$ of about 0.22 (which is in between our estimates of workers' and entrepreneurs' values) and that the social return to schooling ψ of about 3.27. These estimates point to a large effect of schooling for productivity via social interactions, consistent with Lucas (1985, 2008) as well as with the literature in

urban economics. Finally, note that at the above parameter values and at a reasonable share of housing consumption of θ = 0.4, the spatial equilibrium is stable, since $(\beta - \psi\gamma)(1-\theta) + \theta(1-\delta) = -(0.21)(0.6) + (0.4)(0.73) > 0$.

We can trace the magnitude of externalities implied by our parameter estimates. Using the parameters of the paper, raising the educational level from the sample mean of 6.58 years by one year can be calculated to increase regional TFP by about 6.7%. Rauch (1993) estimates a comparable magnitude of 3-5%. Acemoglu and Angrist (2000) estimate that a one year increase in average schooling is associated with a 7% increase in average wages. Moretti (2004) examines the impact of spillovers associated with the share of college graduates living in a city. If we run the regressions in Table 8 using the fraction of the population with college degrees instead of our measure of years of schooling, our estimates imply that a one percentage point increase in the share of region's population with a college degree increases output per capita by 7.9%. Iranzo and Peri (2009) estimate that one extra year of college per worker increase the state's TFP by a very significant 6-9%, whereas the effect of an extra year of high school is closer to 0-1%. The range of agreement among the various estimates is actually quite striking.

Table 8A presents some robustness checks of these findings by controlling for the geographic factors that we have found to influence regional productivity, as well as log power line density in the region and for access to finance. The coefficients on geography variables are quite unstable in these specifications with inverse distance to coast exerting a large positive influence on productivity in two specifications (but not the other six), and oil endowment exerting now a large negative effect in four specifications (those using energy expenditure as a proxy for capita). Access to finance is never significant at the regional level, and log power line density is significant and positive in two specifications. Critically, the coefficients on years of education of managers and years of education of

workers do not fall very much relative to the specifications in Table 8, indicating that returns to education of entrepreneurs remain very high even with controls. The coefficient on years of education in the region stays at around 0.06 and significant in six out of 8 specifications, but declines to zero in specifications with the huge effect of distance to the coast. This evidence needs to be explored further, but in our view most of the specifications confirms both the general findings, and parameter estimates, presented in Table 8.

In sum, our firm level and regional regressions suggest that: i) in line with the development accounting literature, workers' human capital is an important but not a large contributor to productivity differences, but also that ii) entrepreneurial input and externalities represent two fundamental and relatively neglected channels for understanding the role of schooling in shaping productivity differences. Our parameter estimates point to very large returns to entrepreneurial schooling (perhaps due to entrepreneurs' general talent) and to large social returns at the regional level arising from education.

VI. Additional Implications.

The model has a number of additional implications. Specifically, it predicts that higher human capital regions within a country should have larger establishments, as well as a higher share of employment in population. As described in Section III, we have collected data from official censuses of establishments and population for 1, 068 regions from 69 countries in our sample. Before looking at the data, note that in our model these additional predictions are influenced by regional variation in the average firm size l_i/f_i. Since by Proposition 1 productive regions have larger firms on average and since they have a larger number of workers (i.e., $l_G > l_B$), they should also have a higher share of the workforce in total population $l_i/(f_i + l_i + r)$, where r is the measure of rentiers in productive and unproductive regions (recall we assumed $r = 1$, but here for clarity we keep it general). Accordingly, the number of

establishments relative to the population would also be higher in productive regions provided these regions have a sufficiently higher number of firms than unproductive ones, i.e., if $(f_G - f_B) > (f_G f_B / r)(l_G/f_G - l_B/f_B)$. This latter condition however is not satisfied in the parameter region of Proposition 1.

Table 9 presents the results on the effects of human capital in the region (holding country fixed effect constant) on the number of official establishments relative to the population, two indicators of establishment size, and the number of formal employees relative to the population. It shows that higher human capital in the region is strongly associated with larger establishments, higher labor force participation rate, a larger share of employees working in large establishments, but also a higher number of establishments per person. Figure 7 presents the graphs of these relationships.

All of these facts are consistent with Proposition 1 except for the last one. Although there might be parameter constellations where our model reproduces all of the facts of Table 8, we believe that the most likely reason why in the data productive regions have a larger number of firms per capita is the presence of an informal sector, which is not currently included in our model. In fact, the share of unofficial firms and workers is probably lower in more productive regions because the larger firms of productive regions are less likely to be informal, an observation that is true empirically (La Porta and Shleifer 2008). In this respect, adding to our model the notion that larger firms are less likely to be informal would naturally yield the conclusion that productive regions, having a greater number of large firms, also have fewer firms in the informal sector, featuring a larger "official" firms/population ratio notwithstanding the fact that these regions have a larger average firm l_i/f_i.

More broadly, these results suggest the possibility that one channel through which higher human capital raises regional income is by drawing more workers out of the unofficial sector (or agriculture) into the productive formal sector. The externality from social interactions might be larger in higher human capital regions because more of the firms and workers in such regions benefit from

informational spillovers associated with human capital. These results are very consistent with the predictions of the model.

VII. Conclusion.

We have presented evidence from more than 1500 sub-national regions of the world on the determinants of regional income and labor productivity. The evidence suggests that regional education is the critical determinant of regional development, and the only such determinant that explains a substantial amount of regional variation. Using data on several thousand firms located in these regions, we have also found that regional education influences regional development through education of workers, education of entrepreneurs, and substantial regional externalities. Moreover, the externalities come primarily from education (the quality of human capital), and not from its total quantity (the number of people with some education). Finally, we found that better educated regions have larger, more productive firms, and higher labor force participation.

A simple Cobb-Douglas production function specification used in development accounting would have difficulty accounting for all this evidence. Instead, we presented what we called a Lucas-Lucas model of an economy, which combines the allocation of talent between work and entrepreneurship, human capital externalities, and migration of labour across regions within a country. Although many issues remain to be resolved, the empirical findings we presented are both consistent with the general predictions of this model, and provide plausible values for the model's parameters.

The central message of the estimation/calibration exercise is that, while private returns to worker education are modest and close to previous estimates, private returns to entrepreneurial education (in the form of profits) as well as social returns to education through external spillovers are

large. This evidence suggests that earlier estimates of return to education have perhaps underestimated one of its important benefits – the externalities, and largely missed the other – entrepreneurship. This final observation has significant implications for economic development.

Our data points most directly to the role of the supply of educated entrepreneurs for the creation and productivity of firms. From the point of view of development accounting, having such entrepreneurs seems much more important than having educated workers. Consistent with earlier observations of Banerjee and Duflo (2005) and LaPorta and Shleifer (2008), economic development occurs in educated regions that concentrate entrepreneurs, who run large productive firms. These entrepreneurs, as well, appear to contribute to the exchange of ideas, leading so significant regional externalities. The observed large benefits of education through the creation of a supply of entrepreneurs and through externalities offer an optimistic assessment of the possibilities of economic development through raising educational attainment.

Bibliography

Acemoglu, Daron and Joshua Angrist (2000). "How Large are Human-Capital Externalities? Evidence from Compulsory Schooling Laws." *NBER Macroeconomics Annual*, 15:9-59.

Acemoglu, Daron, Simon Johnson, James A. Robinson (2001). "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review*, 91(5): 1369-1401.

Acemoglu, Daron and Melissa Dell (2010). "Productivity Differences Between and Within Countries." *American Economic Journal: Macroeconomics*, 2(1):169-188.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat and Romain Wacziarg (2003). "Fractionalization." *Journal of Economic Growth,* 8:155-94.

Alvarez, Michael, Jose Cheibub, Fernando Limongi, and Adam Przeworski (2000). *Democracy and Development: Political Institutions and Material Well-Being in the World 1950-1990*. Cambridge, UK: Cambridge University Press.

Banerjee, Abhijit, and Esther Duflo (2005). "Growth Theory through the Lens of Development Economics." In Philippe Aghion and Steven Durlauf, eds. *Handbook of Economic Growth v. 1a*, Amsterdam: Elsevier.

Banerjee, Abhijit, and Lakshmi Iyer (2005). "History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure System in India." *American Economic Review* 95: 1190-1213.

Barro, Robert J (1991). "Economic Growth in a Cross-Section of Countries." *Quarterly Journal of Economics*, 106(2): 407-443.

Baumol, Wiliam (1990), "Entrepreneurship: Productive, Unproductive, and Destructive." *Journal of Political Economy* 98 (5): 893-921.

Beck, Thorsten, Gerge Clarke, Alberto Groff, Philip Keefer, and Patrick Walsh (2001). “New Tools in Comparative Political Economy: The Database of Political Institutions.” *World Bank Economic Review* 15 (1): 165-176.

Benhabib, Jess and Spiegel, Mark M. "The Role of Human Capital in Economic Development." *Journal of Monetary Economics*, October 1994, 34(2), pp. 143-174.

Bils, Mark and Peter J. Klenow (2000). “Does Schooling Cause Growth?” *American Economics Review*, 90(5): 1160-1183.

Bloom, David E. and Jeffrey D. Sachs (1998). “Geography, Demography, and Economic Growth in Africa.” *Brookings Papers on Economic Activity*, 29(1998-2: 207-296.

Bloom, Nicholas and John Van Reenen (2010). “Why Do Management Practices Differ across Firms and Countries?” *Journal of Economic Perspectives*, 24(1): 203-224.

Card, David (1999). “The Causal Effect of Education on Earnings.” In Orley Ashenfelter and David Card, eds., *Hanbook of Labor Economics*, vol 3A, Amsterdam: North Holland.

Caselli, Francesco (2005). “Accounting for Cross-Country Income Differences.” In Philippe Aghion and Steven Durlauf (eds.), *Handbook of Economic Growth*, vol.1, ch. 9: 679-741 Amsterdam: Elsevier.

Caselli, Francesco and Coleman, John Wilbur II (2005). “The World Technology Frontier.” *American Economic Review*, 96(3): 499-522.

Caselli, Francesco and James Feyrer (2007). “The Marginal Product of Capital.” *Quarterly Journal of Economics* 122 (2): 535-568.

Caselli, Francesco and Nicola Gennaioli (2009). “Dynastic Management.” Mimeo: LSE and CREI.

Ciccone, Antonio and Robert E. Hall (1996). “Productivity and the Density of Economic Activity.” *American Economic Review*, 86(1):54-70.

Ciccone, Antonio and Giovanni Peri (2006). “Identifying Human-Capital Externalities: Theory with Applications.” *Review of Economic Studies*, 73(2):381-412.

Ciccone, Antonio and Elias Papaioannou (2009). “Human Capital, the Structure of Production, and Growth.” *Review of Economics and Statistics*, 91(1): 66-82.

Cohen, Daniel and Marcelo Soto (2007). “Growth and Human Capital: Good Data, Good Results.” *Journal of Economics Growth*, 12:51-76.

de La Fuente, Angel and Rafael Domenech, (2006). Human Capital in Growth Regressions: How Much Difference Does Data Quality Make?” *Journal of the European Economics Association*, 4(1):1-36.

Dell, Melissa (2010). “The Persistent Effects of Peru’s Mining *Mita.*” *Econometrica*, forthcoming.

Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken (2009). “Temperature and Income: Reconciling New Cross-Sectional and Panel Estimates.” NBER Working Paper No. 14680.

De Long, J. Bradford, and Andrei Shleifer (1993). "Princes or Merchants? City Growth before the Industrial Revolution" *Journal of Law and Economics*, 36:671-702.

Easterly, William and Levine, Ross (1997). “Africa's Growth Tragedy: Policies and Ethnic Divisions.” *Quarterly Journal of Economics*, 112(4):1203-50.

Glaeser, Edward L., Hedi D. Kallal, Jose A. Scheinkman, and Andrei Shleifer (1992). “Growth in Cities.” *Journal of Political Economy*, 100(6):1126-1152.

Glaeser, Edward L. and Joshua D. Gottlieb (2009). "The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States." *Journal of Economic Literature*, 47(4):983-1028.

Glaeser, Edward L., Jose A. Scheinkman and Andrei Shleifer (1995). "Economic Growth in a Cross-Section of Cities." *Journal of Monetary Economics*, 36:117-143.

Glaeser, Edward L. and David C. Mare (2001). "Cities and Skills." *Journal of Labor Economics*, 19(2):316-342.

Glaeser, Edward L., Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer (2004). "Do Institutions Cause Growth?" *Journal of Economic Growth*, 9:271-303.

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

Hall, Robert E. and Charles I. Jones (1999). "Why Do Some Countries Produce So Much More Output per Worker than Others?" *Quarterly Journal of Economics*, 114(1)**:** 83-116.

Henderson, Vernon, Adam Storeygard, and David Weil (2009). "Measuring Economic Growth From Outer Space." NBER Working Paper No. 15199.

Hsieh, Chang-Tai and Peter J. Klenow (2009). "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics*, 124(4): 1403-1448.

Hsieh, Chang-Tai and Peter J. Klenow (2010). "Development Accounting." *American Economic Journal: Macroeconomics*, 2:1, 207–223.

Iranzo, Susana and Giovanni Peri (2009). "Schooling Externalities, Technology, and Productivity: Theory and Evidence from U.S. States." *Review of Economics and Statistics*, 91(2):420-431.

Jaggers, Keith, and Monty G. Marshall (2000). *Polity IV Project*. University of Maryland.

Jacobs, Jane (1969). *The Economy of Cities*. New York: Vintage.

King, Robert G. and Levine, Ross, 1993. "Finance and Growth : Schumpeter Might be Right." *Quarterly Journal of Economics*, 108(3): 717-737.

Klenow, Peter J. and Andres Rodriguez-Clare (2005). "Externalities and Growth." In Philippe Aghion and Steven Durlauf (eds.), *The Handbook of Economic Growth*, Volume 1A, pp. 818-861, Elsevier.

Knack, Stephen and Philip Keefer (1997). "Does Social Capital Have An Economic Payoff? A Cross-Country Investigation." *Quarterly Journal of Economics*, 112(4):1251-1288.

Krueger, Alan B. and Mikael Lindahl (2001). "Education for Growth: Why and For Whom?" *Journal of Economic Literature*, 39(4):1101-1136.

Landes, David S. (1998). *The Wealth and Poverty of Nations: Why Some Are So Rich and Some So Poor*. New York: W. W. Norton & Company.

La Porta, Rafael, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2008. "The Economic Consequences of Legal Origins." *Journal of Economic Literature*, 46(2): 285–332.

La Porta, Rafael and Andrei Shleifer (2008). "The Unofficial Economy and Economic Development." *Brookings Papers on Economic Activity*, 2008: 275-352.

Lucas, Jr., Robert E. (1978). "On the Size Distribution of Business Firms." *Bell Journal of Economics* 9(2): 508–23.

Lucas, Jr., Robert E. (1988). "On the Mechanics of Economic Development." *Journal of Monetary Economics*, 22(1):3-42.

Lucas, Jr., Robert E. (2009). "Ideas and Growth." *Economica*, 76(301): 1–19.

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

Moretti, Enrico (2004). "Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-Sectional Data." *Journal of Econometrics*, 121: 175-212.

Moretti, Enrico (2004). "Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions." *American Economic Review*, 94(3):656-690.

Murphy Kevin M., Andrei Shleifer, and Robert W. Vishny (1991). "The Allocation of Talent: Implications for Growth." *Quarterly Journal of Economics*, 106(2):503-530.

Nelson, Richard R. and Edmund S. Phelps (1966). "Investment in Humans, Technological Diffusion and Economic Growth." *American Economic Review, 56(2):69-75.*

Psacharopoulos, George (1994). "Returns to Investment in Education: A Global Update." *World Development* 22(9): 1325-1343.

Putnam, Robert (1993). *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton Univ. Press

Rauch, James E. (1993). "Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities." *Journal of Urban Economics*, 34:380-400.

Roback, Jennifer (1982). "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90(6): 1257-1278.

Romer, Paul (1990). "Endogenous Technical Change." *Journal of Political Economy*, 98(5): S71-S102.

Valentyi, Akos and Berthold Herrendorf (2008). "Measuring Factor Income Shares at the Sector Level." *Review of Economic Dynamics* 11: 820-838.

Appendix 1.

Spatial Equilibrium when $\psi > 1$. Consider a spatial equilibrium where each unproductive region retains only a fraction $x_B < 1$ of their initial human capital. By Equation (14), this implies $x_G = [1 - (1 - p)\, x_B]/p$. Even in this case, the spatial equilibrium is identified by Equation (13), which when $\psi \geq 1$ becomes:

$$h_m \cdot \left[1 - \left(\frac{A_B}{A_G} \right)^{\frac{1-\theta}{1-\delta}} \left(\frac{L_B}{L_G} \right)^{\gamma(1-\psi)\frac{1-\theta}{1-\delta}} \left(\frac{H_G}{H_B} \right)^{\frac{(\beta-\gamma\psi)(1-\theta)+(1-\delta)\theta}{1-\delta}} \right] = \varphi, \tag{1.A}$$

Which reduces to (13) for $\psi = 1$. To pin down the equilibrium, we must express all endogenous variables as a function of x_B. It is easy to see that $h_m = \underline{h}(1 - x_B)^{-\frac{1}{\mu-1}}$ and that $H_G/H_B = [1 - (1 - p)\, x_B]/p x_B$. We thus only need to determine L_G/L_B. Given the above expression for h_m, we have that:

$$\frac{L_B}{L_G} = p \frac{1 - (1 - x_B)^{\frac{\mu}{\mu-1}}}{p + (1-p)(1 - x_B)^{\frac{\mu}{\mu-1}}}. \tag{2.A}$$

The equilibrium is found by plugging (2.A) into (1.A). A full mobility (i.e. $\varphi = 0$) equilibrium arises when:

$$\left(\frac{A_G}{A_B} \right)^{\frac{1-\theta}{1-\delta}} = p^{\frac{(\beta-\gamma)(1-\theta)+(1-\delta)\theta}{1-\delta}} \left(\frac{p + (1-p)(1 - x_B)^{\frac{\mu}{\mu-1}}}{1 - (1 - x_B)^{\frac{\mu}{\mu-1}}} \right)^{\gamma(\psi-1)\frac{1-\theta}{1-\delta}} \left(\frac{1 - (1-p)x_B}{x_B} \right)^{\frac{(\beta-\gamma\psi)(1-\theta)+(1-\delta)\theta}{1-\delta}}. \tag{3.A}$$

If $(\beta - \psi\gamma)(1 - \theta) + \theta(1 - \delta) > 0$, the right hand side - which captures the cost of migrating to productive regions, decreases in x_B. At $x_B = 0$ the right hand side of (3.A) tends to infinity, so full sorting cannot be an equilibrium. At $x_B = 1$ the right hand side is equal to one, which is below the left hand side. As a result, no sorting is not an equilibrium either. In sum, if $(\beta - \psi\gamma)(1 - \theta) + \theta(1 - \delta) > 0$ there is a unique stable equilibrium with intermediate sorting.

Proof of Proposition 1 The threshold h_m above which agents migrate is implicitly defined by:

$$\int_{\underline{h}}^{h_m} \mu \underline{h}^{\mu} h^{-\mu} dh = E(h)\left[1-\left(\frac{\underline{h}}{h_m}\right)^{\mu-1}\right] = H_B \text{ ,} \tag{4.A}$$

and H_B is identified by Equation (21). Define $\hat{A}_i = \frac{A_i^{\frac{1-\theta}{(\beta-\gamma)(1-\theta)+\theta(1-\delta)}}}{E\left[A^{\frac{1-\theta}{(\beta-\gamma)(1-\theta)+\theta(1-\delta)}}\right]}$. It is then easy to find that:

$$h_m = \underline{h}\left[1-\hat{A}_B\right]^{-\frac{1}{\mu-1}}, \tag{5.A}$$

The higher is A_B, the more skilled must a labourer be in order for him to find migration profitable (i.e. the higher is h_m). Suppose than that:

$$\left(\frac{1-\alpha-\beta-\delta}{1-\beta-\delta}\right) < 1-\frac{\hat{A}_B}{\hat{A}_G}, \tag{6.A}$$

which ensures that the least productive migrants become workers. Then the human capital thresholds above which agents become entrepreneurs in the productive and unproductive regions are equal to:

$$h_G^E = \underline{h}\cdot\left[p\left(\frac{1-\alpha-\beta-\delta}{1-\beta-\delta}\right)\cdot\hat{A}_G\right]^{-\frac{1}{\mu-1}}, h_B^E = \underline{h}\cdot\left[1-\left(\frac{\alpha}{1-\beta-\delta}\right)\cdot\hat{A}_B\right]^{-\frac{1}{\mu-1}}, \tag{7.A}$$

where $h_G^E > h_B^E$, so that entrepreneurs are more skilled in productive regions. Defining f_G and f_B the number of firms in the productive and unproductive regions, respectively, and by l_G and l_B the respective workforces, one finds that in equilibrium we have that:

$$f_G = p^{\frac{1}{\mu-1}}\cdot\left[\left(\frac{1-\alpha-\beta-\delta}{1-\beta-\delta}\right)\hat{A}_G\right]^{\frac{\mu}{\mu-1}}, f_B = \left[1-\left(\frac{\alpha}{1-\beta-\delta}\right)\hat{A}_B\right]^{\frac{\mu}{\mu-1}} - \left(1-\hat{A}_B\right)^{\frac{\mu}{\mu-1}}, \tag{7.A}$$

where f_i is the integral of the density of skills above h_i^E according to the distributions of Figure 2. The size of the workforce in the two regions is equal to:

$$l_G = \int_{\underline{h}}^{h_m} \mu \underline{h}^{\mu} h^{-\mu-1} dh + \frac{1}{p}\int_{h_m}^{h_G^E} \mu \underline{h}^{\mu} h^{-\mu-1} dh = 1 + \frac{1-p}{p}(1-\hat{A}_B)^{\frac{\mu}{\mu-1}} - p^{\frac{1}{\mu-1}}\left(\frac{1-\alpha-\beta-\delta}{1-\beta-\delta}\hat{A}_B\right)^{\frac{\mu}{\mu-1}}.$$

$$l_B = \int_{\underline{h}}^{h_B^E} \mu \underline{h}^{\mu} h^{-\mu-1} dh = 1 - \left[1 - \left(\frac{\alpha}{1-\beta-\delta}\right)\hat{A}_B\right]^{\frac{\mu}{\mu-1}}.$$

As a result, we have that $l_G/f_G > l_B/f_B$ when the following condition holds:

$$\frac{p + (1-p)(1-\hat{A}_B)^{\frac{\mu}{\mu-1}}}{\left(\frac{1-\alpha-\beta-\delta}{1-\beta-\delta}\right)^{\frac{\mu}{\mu-1}}\left[1-(1-p)(1-\hat{A}_B)\right]^{\frac{\mu}{\mu-1}}} \geq \frac{1-(1-\hat{A}_B)^{\frac{\mu}{\mu-1}}}{\left(1-\frac{\alpha}{1-\beta-\delta}\hat{A}_B\right)^{\frac{\mu}{\mu-1}} - (1-\hat{A}_B)^{\frac{\mu}{\mu-1}}}$$

The numerator/denominator of the left hand side is larger/smaller than the numerator and denominator of the right hand side if and only if:

$$\begin{aligned} &\frac{1}{p(\hat{A}_G/\hat{A}_B)+(1-p)} \leq 1-\left(\frac{1-p}{2-p}\right)^{\frac{\mu-1}{\mu}} \\ &\frac{1}{p(\hat{A}_G/\hat{A}_B)+(1-p)} \geq \frac{1-\left(\frac{\alpha}{1-\beta-\delta}\right)^{\mu-1}}{1-\left(\frac{\alpha}{1-\beta-\delta}\right)^{\mu}} \end{aligned}, \qquad (8.A)$$

The first condition is met if $\hat{A}_G/\hat{A}_B$ is sufficiently large (i.e. larger than a certain value z_1). The second condition is met if $\hat{A}_G/\hat{A}_B$ is sufficiently low (i.e. smaller than a certain value z_2). Finally, (8.A) can be met (i.e. $z_1 < z_2$) when p is large.

Table 1: Descriptive Statistics

The table reports descriptive statistics for the variables in the paper. We report the total number of observations, the number of countries and medians for: (1) the number of regions with non-missing data, (2) the country average, (3) the within-country range, (4) the within-country standard deviation, and (5) the coefficient of variation for the variable in levels (rather than in logs). We also report the adjusted R squared from an univariate regression of each variable in the table on country dummies. All variables are described in Table 2.

Panel A: Regional GDP, Education, Geography, Institutions, Infrastructure, and Culture

			Medians for:						
	Number of Regions	Number of Countries	Regions per country	Mean	Minimum	Maximum	Within-country Range	Within-country std deviation	Coefficient of Variation for Variable in Levels
Ln(GDP per capita)	1,537	107	11	8.69	8.07	9.54	1.03	0.30	0.33
Years of Education	1,489	106	12	6.58	5.34	8.70	2.34	0.73	0.92
Temperature	1,568	110	12	16.84	10.23	21.13	4.47	1.45	0.09
Inverse Distance to Coast	1,569	110	12	0.90	0.80	0.99	0.13	0.05	0.05
Ln(Oil)	1,569	110	12	0.00	0.00	0.00	0.00	0.00	0.00
Informal Payments	361	76	4	1.02	0.40	1.60	0.94	0.45	0.59
ln(Tax Days)	270	58	5	1.29	1.06	1.51	0.36	0.19	0.18
Ln(Days without electricity)	222	75	2	3.03	2.73	3.37	0.54	0.36	0.32
Security costs	373	79	4	0.91	0.39	1.34	0.72	0.34	0.42
Access to land	519	81	5	0.15	0.04	0.27	0.21	0.09	0.40
Access to finance	536	82	5	0.28	0.14	0.47	0.29	0.12	0.24
Government Predictability	386	75	4	0.46	0.34	0.61	0.24	0.10	0.20
Doing Business Percentile Rank	180	19	6	0.40	0.21	0.49	0.22	0.11	0.31
Ln(Power line density)	1,569	110	12	1.34	0.00	2.53	1.87	0.61	0.61
Ln(Travel time)	1,569	110	12	5.28	4.21	6.00	1.82	0.54	0.46
Trust in others	745	69	9	0.23	0.12	0.38	0.22	0.07	0.35
Civic Values	683	75	8	2.23	1.71	3.12	1.08	0.48	0.19
Ln(Number of ethnic groups)	1,568	110	12	0.98	0.00	1.79	1.39	0.50	0.46
Probability of same language	1,545	109	12	0.67	0.28	0.79	0.26	0.09	0.21

Table 1: Descriptive Statistics (continued)

Panel B: Enterprise Survey and Census Data

			Medians for:						
	Number of Regions	Number of Countries	Regions per country	Mean	Minimum	Maximum	Within-country Range	Within-country std deviation	Coefficient of Variation for Variable in Levels
Ln(Establishments / Population)	984	65	12	-4.89	-5.45	-4.06	1.17	0.37	0.37
Ln(Employees / Establishments)	1,068	69	12	2.07	1.69	2.39	0.80	0.20	0.19
Ln(Employees / Population)	1,056	69	12	-2.66	-3.38	-1.80	1.58	0.43	0.41
Ln(Employees Big Firms / Employees)	540	31	13	-1.45	-2.17	-0.78	1.13	0.33	0.27
Ln(Sales / Employee)	550	82	5	10.21	9.79	10.59	0.79	0.35	1.22
Ln(Wages / Employee)	516	77	5	8.28	8.00	8.66	0.62	0.25	1.79
Ln(Employees)	550	82	5	3.25	2.72	3.71	0.82	0.35	1.46
Ln(Expenditure on energy / Employee)	326	66	4	6.10	5.52	6.36	0.60	0.30	1.22
Ln(Property, plant and equipment / Employee)	205	41	4	8.72	8.37	9.37	0.99	0.47	1.26
Years of Education of Workers	507	74	5	9.97	8.66	10.80	2.25	0.93	3.06
Years of Education of Managers	195	38	4	14.90	14.24	15.36	1.34	0.62	0.89

Table 2 – Definitions and sources for the variables used in the paper

This table provides the names, definitions and sources of all the variables used in the tables of the paper.

Variable	Description	Sources and links
	I. GDP per capita, population, employment and human capital	
Ln(GDP per capita)	The logarithm of Gross Domestic Product per capita in PPP constant 2005 international dollars in the region in 2005. Data on regional GDP is available for all countries except 20. For those 20 countries, we approximate GDP using data on income (6 countries), expenditure (8 countries), wages (3 countries), gross value added (2 countries), and consumption, investment and government expenditure (1 country). For each country, we scale regional GDP per capita values so that their population-weighted sum equals the World Development Indicators (WDI) value of Gross Domestic Product in PPP constant 2005 international dollars. Similarly, for each country, we adjust the regional population values so that their sum equals the country-level analog in WDI. For years with missing regional GDP per capita data, we interpolate using all available data for the period 1990-2008. When interpolating GDP values is not possible, we use the regional distribution of the closest year with regional GDP data. Population data for years without census data is interpolated and extrapolated from the available census data for the period 1990-2008. At the country level, we calculate this variable as the population-weighted average of regional GDP.	*Regional GDP*: See online appendix "Appendix GDP Sources". *Regional population*: Thomas Brinkhoff: City Population, http://www.citypopulation.de/ *Country-level GDP per capita and PPP exchange rates*: World Bank, (2010). Data retrieved on March 2, 2010, from World Development Indicators Online (WDI) database, http://go.worldbank.org/6HAYAHG8H0
Ln(Population)	The logarithm of the number of inhabitants in the region in 2005. Population data for years without census data is interpolated and extrapolated from the available census data for the period 1990-2008. For each country, we adjust the regional populations so that the sum of regional populations equals the country-level analog in the World Development Indicators (WDI). At the country level, we calculate this variable following the same methodology but using country boundaries.	*Regional population*: Thomas Brinkhoff: City Population, http://www.citypopulation.de/ *Regional spherical*: Collins-Bartholomew World Digital Map, http://www.bartholomewmaps.com/data.asp?pid=5.
Ln(Employment)	The number of manufacturing and service employees working in the establishments in the region. The data is for the year 2005 or the closest available. At the country level, we calculate this variable as the product of the total population and the employment ratio for the population 15 years and older.	See online appendix "Appendix on Economic Census Sources". Development Indicators Online (WDI) database, http://go.worldbank.org/6HAYAHG8H0
Years of education	The average years of schooling from primary school onwards for the population aged 15 years or older. Data for China and Georgia is for the population 6 years and older. We use the most recent information available for the period 1990-2006. To make levels of educational attainment comparable across countries, we translate educational statistics into the International Standard Classification of Education (ISCED) standard and use UNESCO data on the duration of school levels in each country for the year for which we have educational attainment data. Eurostat aggregates data for ISCED levels 0-2 and we assign such observations an ISCED level 1. Following Barro and Lee (1993): (1) we assign zero years of schooling to ISCED level 0 (i.e., pre-primary); (2) we assign zero years of *additional* schooling to (a) ISCED level 4 (i.e., vocational), and (b) ISCED level 6 (i.e. post-graduate); and (3) we assign 4 years of additional schooling to ISCED level 5 (i.e. graduate). Since regional data is not available for all countries, unlike Barro and Lee (1993), we assign zero years of *additional* schooling: (a) to all incomplete levels; and (b) to ISCED level 2 (i.e. lower secondary). Thus, the average years of schooling in a region is calculated as: (1) the product of the fraction of people whose highest attainment level is ISCED 1 or 2 and the duration of ISCED 1; plus (2) the product of the fraction of people whose highest attainment level is ISCED 3 or 4 and the cumulative duration of ISCED 3; plus (3) the product of the fraction of people whose highest attainment level is ISCED 5 or 6 and the sum of the cumulative duration of ISCED 3 plus 4 years. At the country level, we calculate this variable as the population-weighted average of the regional values.	See online appendix "Appendix on Education Sources". Links to online data: http://epdc.org/ http://epp.eurostat.ec.europa.eu/portal/page/portal/region_cities/introduction https://international.ipums.org/international/index.html http://stats.uis.unesco.org/unesco/TableViewer/document.aspx?ReportId=143&IF_Language=eng.
	II. Climate, geography and natural resources	
Temperature	Average temperature during the period 1950-2000 in degrees Celsius. To produce the regional and national numbers, we create equal area projections using the Collins-Bartholomew World Digital Map and the temperature raster in ArcGIS. For each region, we sum the temperatures of all cells in that region and divide by the number of cells in that region. At the country level, we calculate this variable following the same methodology but using country boundaries.	Climate: Hijmans, R. et al. (2005) , http://www.worldclim.org/ Collins-Bartholomew World Digital Map, http://www.bartholomewmaps.com/data.asp?pid=5

Variable	Description	Sources and links
Inverse distance to coast	The ratio of one over one plus the region's average distance to the nearest coastline in thousands of kilometers. To calculate each region's average distance to the nearest coastline we create an equal distance projection of the Collins-Bartholomew World Digital Map and a map of the coastlines. Using these two maps we create a raster with the distance to the nearest coastline of each cell in a given region. Finally, to get the average distance to the nearest coastline, we sum up the distance to the nearest coastline of all cells within each region and divide that sum by the number of cells in the region. At the country level, we calculate this variable following the same methodology but using country boundaries.	Collins-Bartholomew World Digital Map, http://www.bartholomewmaps.com/data.asp?pid=5
Ln(Oil)	Logarithm of one plus the estimated per capita volume of cumulative oil production and reserves by region, in millions of barrels of oil. To produce the regional measure, we load the oil map of the World Petroleum Assessment and the Collins-Bartholomew World Digital map onto ArcGIS. On-shore estimated oil in each assessment unit was allocated to the regions based on the fraction of assessment unit area covered by each region. Off-shore assessment units are not included. The World Petroleum Assessment map includes all oil fields in the world except those in the United States of America. Data for the United States is calculated using the national-level information on cumulative production and estimated reserves, available from the World Petroleum Assessment 2000 (USGS), and the United States' regional production and estimated reserves for the year 2000 from the U.S. Energy Information Administration (USEIA). The national level data for this variable is calculated following the same methodology outlined but using the data on national boundaries. The national level numbers for the U.S. are those available from the World Petroleum Assessment.	http://energy.cr.usgs.gov/oilgas/wep/products/dds60/export.htm. http://tonto.eia.doe.gov/dnav/pet/pet_crd_crpdn_adc_mbbl_a.htm. http://www.bartholomewmaps.com/data.asp?pid=5
	III. Institutions	
Informal payments	The average percentage of sales spent on informal payments made to public officials to "get things done" with regard to customs, taxes, licenses, regulations, services, etc, as reported by the respondents in the region. The country-level analog of this variable is the arithmetic average of the regions in the country. Data is from the most recent year available, ranging from 2002 through 2009.	World Bank's Enterprise Surveys. https://www.enterprisesurveys.org/
Ln(Tax days)	The logarithm of one plus the average number of days spent in mandatory meetings and inspections with tax authority officials in the past year as reported by respondents in the region. The country-level analog of this variable is the arithmetic average of the regions in the country. Data is for the most recent year available, ranging from 2002 through 2009.	World Bank's Enterprise Surveys. https://www.enterprisesurveys.org/
Ln(Days without electricity)	The logarithm of one plus the average number of days without electricity in the past year as reported by the respondents in the region. The country-level analog of this variable is the arithmetic average of the regions in the country. Data is for the most recent year available, ranging from 2002 through 2009.	World Bank's Enterprise Surveys. https://www.enterprisesurveys.org/
Security costs	The average costs of security (i.e., equipment, personnel, or professional security services) as a percentage of sales as reported by the respondents in the region. The country-level analog of this variable is the arithmetic average of the regions in the country. Data is for the most recent year available, ranging from 2002 through 2009.	World Bank's Enterprise Surveys. https://www.enterprisesurveys.org/
Access to land	The percentage of respondents in the region who think that access to land is a moderate, major, or very severe obstacle to business. The country-level analog of this variable is the arithmetic average of the regions in the country. Data is for the most recent year available, ranging from 2002 through 2009.	World Bank's Enterprise Surveys. https://www.enterprisesurveys.org/
Access to finance	The percentage of respondents in the region who think that access to financing is a moderate, major, or very severe obstacle to business. The country-level analog of this variable is the arithmetic average of the regions in each respective country. Data is for the most recent year available, ranging from 2002 through 2009.	World Bank's Enterprise Surveys. https://www.enterprisesurveys.org/
Government predictability	The percentage of respondents in the region who tend to agree, agree in most cases, or fully agree that their government officials' interpretation of regulations are consistent and predictable. The country-level analog of this variable is the arithmetic average of the regions in the country. Data is for the most recent year available, ranging from 2002 through 2009.	World Bank's Enterprise Surveys. https://www.enterprisesurveys.org/
Doing Business Percentile Rank	The average of the percentile ranks in each of the following five areas: (1) starting a business; (2) dealing with construction permits; (3) registering property; (4) enforcing contracts; and (5) paying taxes. Data is for the most recent year available, ranging from 2007 through 2010.	Word Bank's Doing Business Subnational Reports. http://doingbusiness.org/Reports/Subnational-Reports/
Autocracy	This variable classifies regimes based on their degree of autocracy. Democracies are coded as 0, bureaucracies (dictatorships with a legislature) are coded as 1 and autocracies (dictatorship without a legislature) are coded as 2. Transition years are coded as the regime that emerges afterwards. This variable ranges from zero to two where higher values equal a higher degree of autocracy. This variable is measured as the average from 1960 through 1990.	Alvarez et al. (2000).

Variable	Description	Sources and links
Executive Constraints	A measure of the extent of institutionalized constraints on the decision making powers of chief executives. The variable takes seven different values: (1) Unlimited authority (there are no regular limitations on the executive's actions, as distinct from irregular limitations such as the threat or actuality of coups and assassinations); (2) Intermediate category; (3) Slight to moderate limitation on executive authority (there are some real but limited restraints on the executive); (4) Intermediate category; (5) Substantial limitations on executive authority (the executive has more effective authority than any accountability group but is subject to substantial constraints by them); (6) Intermediate category; (7) Executive parity or subordination (accountability groups have effective authority equal to or greater than the executive in most areas of activity). This variable ranges from one to seven where higher values equal a greater extent of institutionalized constraints on the power of chief executives. This variable is calculated as the average from 1960 through 2000.	Jaggers and Marshall (2000).
Expropriation Risk	Risk of "outright confiscation and forced nationalization" of property. This variable ranges from zero to ten where higher values are equals a lower probability of expropriation. This variable is calculated as the average from 1982 through 1997.	International Country Risk Guide at http://www.countrydata.com/datasets/.
Proportional Representation	This variable is equal to one for each year in which candidates were elected using a proportional representation system; equals zero otherwise. Proportional representation means that candidates are elected based on the percentage of votes received by their party. This variable is measured as the average from 1975 through 2000.	Beck et al. (2001).
Corruption	The average score of the Transparency International index of corruption perception in 2005. The index provides a measure of the extent to which corruption is perceived to exist in the public and political sectors. The index focuses on corruption in the public sector and defines corruption as the abuse of public office for private gain. It is based on assessments by experts and opinion surveys. The index ranges between 0 (highly corrupt) and 10 (highly clean).	www.transparency.org
	IV. Infrastructure	
Ln(Power line density)	The logarithm of one plus the length in kilometers of power lines per $10km^2$ in the year 1997. To produce the regional numbers, we load the power line map from the US Geological Survey and the Collins-Bartholomew World Digital Map onto ArcGIS. We take the ratio of total length of the power lines in the region to the spherical area of that region. At the country level, we calculate this variable following the same methodology but using country boundaries.	US Geological Survey Global GIS database, accessed through Harvard University's Geospatial Library. Collins-Bartholomew World Digital Map, http://www.bartholomewmaps.com/data.asp?pid=5
Ln(Travel time)	The logarithm of the average estimated travel time in minutes from each cell in a region to the nearest city of 50,000 or more people in the year 2000. We use the raster from the Global Environmental Monitoring Unit and the Collins-Bartholomew World Digital Map. For each region, we sum the travel time from all its cells and divide by the number of cells in that region. At the country level, we calculate this variable following the same methodology but using country boundaries.	Global Environment Monitoring Unit, http://bioval.jrc.ec.europa.eu/products/gam/index.htm Collins-Bartholomew World Digital Map, http://www.bartholomewmaps.com/data.asp?pid=5
	V. Culture	
Trust in others	The percentage of respondents in the region who believe that most people can generally be trusted. The country-level analog of this variable is the arithmetic average of the regions in the country. Data is for the most recent available year, ranging from 1980 through 2005.	World Values Survey, http://www.worldvaluessurvey.org/
Civic values	The average of the value of the answers of respondents in the region about the degree of justifiability of the following four behaviors: (1) Claiming government benefits to which you are not entitled; (2) Avoiding a fare on public transport; (3) Cheating on taxes if you have a chance; and (4) Someone accepting a bribe in the course of their duties. For each question, possible answers range from 1 (never justifiable) to 10 (always justifiable). We only include observations with non-missing data for at least two of the four questions. The country-level analog of this variable is the arithmetic average of the regions in the country. Data is for the most recent available year, ranging from 1980 through 2005.	World Values Survey, http://www.worldvaluessurvey.org/
Ln(Number of ethnic groups)	The logarithm of the number of ethnic groups that inhabited the region in the year 1964. The country-level analog of this variable is constructed using country boundaries.	Weidmann et al., 2010, http://www.icr.ethz.ch/research/greg

Variable	Description	Sources and links
Probability of same language	The probability that two randomly chosen people, one from the corresponding region and one from the rest of the country, share the same mother tongue in the year 2004. Where language areas do not overlap with our regions, we compute the number of people speaking a language in a region by weighing the total number of people in a language area by the fraction of the region's surface covered by that language area. We compute the probability of same language separately for each language in a region and then calculate the surface-weighted average of the different languages in a region. The country-level analog of this variable is calculated as the population-weighted average of the regional values.	World Language Mapping System, http://www.gmi.org/wlms/
	VI. Enterprise Survey Data	
Ln(Sales / Employee)	The logarithm of the quotient of total annual revenue (in current USD) over the total number of employees in each establishment. Data is for the most recent available year, ranging from 2002 through 2009.	World Bank's Enterprise Surveys. https://www.enterprisesurveys.org/
Ln(Wages / Employee)	The logarithm of the quotient of total cost of labor (in current USD) the total number of employees in each establishment. Data is for the most recent available year, ranging from 2002 through 2009.	World Bank's Enterprise Surveys. https://www.enterprisesurveys.org/
Ln(Expenditure on energy / Employee)	The logarithm of the quotient of total energy and fuel costs over (in current USD) the total number of employees in each establishment. Data is for the most recent available year, ranging from 2002 through 2009.	World Bank's Enterprise Surveys. https://www.enterprisesurveys.org/
Years of Education of workers	The number of years of schooling from primary school onwards of the average non-management employee in each establishment. To compute this variable, we use the same assumptions and follow the same procedure as used for the previously described years of schooling variable at the regional level. Data is for the most recent available year, ranging from 2002 through 2009.	World Bank's Enterprise Surveys. https://www.enterprisesurveys.org/
Years of Education of manager	The number of years of schooling from primary school onwards of the manager of the establishment. To compute this variable, we use the same assumptions and follow the same procedure as used for the previously described years of schooling variable at the regional level. Data is for the most recent available year, ranging from 2002 through 2009.	World Bank's Enterprise Surveys. https://www.enterprisesurveys.org/
Ln(Property, plant, and equipment / Employee)	The logarithm of the quotient of the book value of property, plant and equipment (in current USD) over the total number of employees in the establishment. Data is for the most recent available year, ranging from 2002 through 2009.	World Bank's Enterprise Surveys. https://www.enterprisesurveys.org/
	VII. Economic Census Data	
Ln(Establishments / Population)	The logarithm of the quotient of the number of manufacturing and service establishments in the region and the region's population. The data is for the year 2005 or the closest available. An establishment is defined as a single physical location at which business is conducted or where services or industrial operations are performed.	See online appendix "Appendix on Economic Census Sources".
Ln(Employees / Population)	The logarithm of the quotient of the number of manufacturing and service employees working in the establishments in the region over the region's population. The data is for the year 2005 or the closest available.	See online appendix "Appendix on Economic Census Sources".
Ln(Employees / Establishments)	The logarithm of the quotient of the number of manufacturing and service employees working in the establishments in the region over the number of establishments in the region. The data is for the year 2005 or the closest available.	See online appendix "Appendix on Economic Census Sources".
Ln(Employees big firms / Employees)	The logarithm of the quotient of the number of manufacturing and service employees working in "big firms" over the total number of formal sector employees working in all the establishments in the region. We define "big firms" as establishments employing over 100 employees. However, when data for the 100 employee threshold is unavailable, we use the closest available threshold to 100 employees. The data is for the year 2005 or the closest available.	See online appendix "Appendix on Economic Census Sources".

Table 3: Univariate Fixed Effects Regressions

Fixed effects regressions of the log of GDP per capita at the regional level in the year 2005. The independent variables are proxies for: (1) geography, (2) Institutions, and (3) Infrastructure and Culture. All variables are explained in Table 2. The table reports the number of observations, the number of countries, the R^2 within, the R^2 between, and the fraction of the variance due to countries. All variables are described in Table 2.

	Observations	Countries	R^2 Within	R^2 Between
Independent Variables:				
Years of Education	1,470	104	38%	58%
Temperature	1,536	107	1%	27%
Inverse Distance to Coast	1,537	107	4%	13%
Ln(Oil)	1,537	107	2%	4%
Informal Payments	350	74	0%	21%
ln(Tax Days)	263	56	0%	20%
Ln(Days without electricity)	219	73	2%	6%
Security costs	362	77	0%	7%
Access to land	507	79	0%	15%
Access to finance	524	80	1%	8%
Government Predictability	380	73	1%	0%
Doing Business Percentile Rank	176	18	2%	13%
Ln(Power line density)	1,537	107	5%	36%
Ln(Travel time)	1,537	107	7%	15%
Trust in others	739	68	0%	18%
Ln(Number of ethnic groups)	1,536	107	5%	17%
Probability of same language	1,518	106	1%	26%

Table 4: GDP per capita and Geography

Ordinary least squares and fixed effects regressions of the log of GDP per capita. The dependent variable is the logarithm of the 2005 level of GDP per capita at the country level in Panel A and at the logarithm of regional GDP per capita in Panel B. The independent variables are (1) temperature, (2) inverse distance to coast, (3) the logarithm of per capita oil production and reserves, (4) the average years of education, (5) the logarithm of population, and (6) the logorithm of the number of employees. Robust standard errors are shown in parentheses. All variables are described in Table 2.

Panel A: Dependent Variable is Logarithm National GDP per capita

	(1)	(2)	(3)
Temperature	-0.0914[a]	-0.0189[c]	-0.0190[c]
	(0.0100)	(0.0106)	(0.0106)
Inverse Distance to Coast	4.4768[a]	2.9647[a]	2.9499[a]
	(0.5266)	(0.5736)	(0.5782)
Ln(Oil)	1.2192[a]	0.9503[a]	0.9473[a]
	(0.1985)	(0.1371)	(0.1375)
Years of Education		0.2566[a]	0.2574[a]
		(0.0308)	(0.0311)
Ln(Population)		0.0684[c]	
		(0.0408)	
Ln(Employment)			0.0576
			(0.0398)
Constant	6.3251[a]	3.5761[a]	3.7959[a]
	(0.4598)	(0.9372)	(0.8977)
Observations	107	104	103
Adjusted R^2	50%	63%	63%

Note: a = significant at the 1% level, b = significant at the 5% level, and c = significant at the 10% level.

Table 4: GDP per capita and Geography (continued)

Panel B: Dependent Variable is Logarithm Regional GDP per capita

	(1)	(2)	(3)
Temperature	-0.0156[c]	-0.0140[c]	-0.0206[c]
	(0.0082)	(0.0084)	(0.0105)
Inverse Distance to Coast	1.0318[a]	0.4979[a]	0.5096[a]
	(0.2078)	(0.1438)	(0.1745)
Ln(Oil)	0.1651[a]	0.1752[a]	0.1941[a]
	(0.0477)	(0.0578)	(0.0440)
Years of Education		0.2755[a]	0.2751[a]
		(0.0171)	(0.0271)
Ln(Population)		0.0125	
		(0.0168)	
Ln(Employment)			0.0661[a]
			(0.0244)
Constant	8.0947[a]	6.3886[a]	5.9154[a]
	(0.2282)	(0.1944)	(0.2516)
Observations	1,545	1,478	833
Number of countries	107	104	49
R^2 Within	8%	42%	50%
R^2 Between	47%	60%	70%
R^2 Overall	34%	62%	70%
Fixed Effects	Yes	Yes	Yes

Note: a = significant at the 1% level, b = significant at the 5% level, and c = significant at the 10% level.

Table 5: National GDP per capita, Institutions, Infrastructure, and Culture

Ordinary least square regressions of the log of GDP per capita at the country level. All regressions include the years of education, logarithm of population, temperature, inverse distance to coast, and the logarithm of per capita oil production and reserves. In addition, regressions include measures of: (1) institutions (Panel A) and (2) infrastructure and culture (Panel B). Robust standard errors are shown in parenthesis. For comparison, the bottom panel shows the adjusted R^2 of two alternative specifications: (1) a regression with all regressors except the measure of institutions or culture; and (2) a regression with all regressors except education. All variables are described in Table 2.

Panel A: Institutions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Years of Education	0.2566a	0.2310a	0.1890a	0.2339a	0.2291a	0.2301a	0.2264a	0.2355a	0.1749b
	(0.0308)	(0.0344)	(0.0310)	(0.0316)	(0.0336)	(0.0350)	(0.0344)	(0.0332)	(0.0703)
Ln(Population)	0.0684[c]	-0.0022	0.0887	0.0428	0.0320	0.0067	0.0299	0.0611	-0.0782
	(0.0408)	(0.0494)	(0.0582)	(0.0488)	(0.0481)	(0.0519)	(0.0473)	(0.0457)	(0.1074)
Temperature	-0.0189[c]	-0.0105	-0.0276[b]	-0.0083	-0.0094	-0.0066	-0.0082	-0.0129	-0.0147
	(0.0106)	(0.0128)	(0.0128)	(0.0119)	(0.0114)	(0.0112)	(0.0110)	(0.0117)	(0.0306)
Inverse Distance to Coast	2.9647[a]	2.3086[a]	2.1692[a]	2.5170[a]	2.2652[a]	2.2826[a]	2.1892[a]	2.3979[a]	0.2385
	(0.5736)	(0.6321)	(0.7006)	(0.5698)	(0.5856)	(0.5406)	(0.5562)	(0.5616)	(2.1131)
Ln(Oil)	0.9503[a]	1.6367[a]	0.5257	1.1319[a]	1.1739[a]	1.1916[a]	1.1165[a]	1.2054[b]	0.5201
	(0.1371)	(0.5966)	(0.5050)	(0.3309)	(0.3219)	(0.3302)	(0.2950)	(0.4982)	(0.4921)
Informal Payments		-0.0121							
		(0.0499)							
ln(Tax Days)			-0.5497[a]						
			(0.1446)						
Ln(Days without electricity)				-0.1375					
				(0.0847)					
Security costs					-0.0332				
					(0.0250)				
Access to land						-0.7493			
						(0.5783)			
Access to finance							-0.5164		
							(0.4202)		
Government Predictability								0.3835	
								(0.4431)	
Doing Business Percentile Rank									0.6704
									(1.6413)
Constant	3.5761[a]	5.1927[a]	5.1619[a]	4.6815[a]	4.7382[a]	5.1545[a]	4.9498[a]	3.9328[a]	8.6509[b]
	(0.9372)	(1.1015)	(1.2918)	(0.9542)	(1.0046)	(0.9971)	(1.0246)	(0.9724)	(3.1636)
Observations	104	73	55	75	76	80	81	72	17
Adjusted R^2	63%	73%	76%	69%	69%	70%	70%	71%	34%
Adj. R^2 without institution	63%	73%	69%	69%	69%	69%	69%	71%	39%
Adj. R^2 without education	50%	53%	60%	49%	50%	52%	52%	50%	26%

Note: a = significant at the 1% level, b = significant at the 5% level, and c = significant at the 10% level.

Table 5: National GDP per capita, Institutions, Infrastructure, and Culture (cont)

Panel B: Infrastructure and Culture

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Years of Education	0.2566[a]	0.2379[a]	0.2642[a]	0.1935[a]	0.1818[a]	0.2534[a]	0.2394[a]
	(0.0308)	(0.0338)	(0.0325)	(0.0498)	(0.0538)	(0.0347)	(0.0377)
Ln(Population)	0.0684[c]	0.0688[c]	0.0653	0.1238	0.2169[b]	0.0999	0.0807[c]
	(0.0408)	(0.0414)	(0.0407)	(0.0788)	(0.1017)	(0.0640)	(0.0450)
Temperature	-0.0189[c]	-0.0145	-0.0191[c]	-0.0283[b]	-0.0434[a]	-0.0188[c]	-0.0163
	(0.0106)	(0.0109)	(0.0108)	(0.0135)	(0.0148)	(0.0107)	(0.0108)
Inverse Distance to Coast	2.9647[a]	2.7218[a]	3.0968[a]	3.6522[a]	4.3386[a]	2.7758[a]	2.7448[a]
	(0.5736)	(0.6025)	(0.6268)	(0.7902)	(1.0486)	(0.6473)	(0.5853)
Ln(Oil)	0.9503[a]	1.0157[a]	0.8737[a]	0.9902[a]	0.9751[a]	0.9538[a]	0.8792[a]
	(0.1371)	(0.1438)	(0.1467)	(0.3207)	(0.2895)	(0.1443)	(0.1657)
Ln(Power line density)		0.1480					
		(0.1099)					
Ln(Travel time)			0.0825				
			(0.0934)				
Trust in others				1.2472			
				(0.8796)			
Civic values					0.4180		
					(0.3105)		
Ln(Number of ethnic groups)						-0.0996	
						(0.1550)	
Probability of same language							0.4195
							(0.3391)
Constant	3.5761[a]	3.6383[a]	3.0050[b]	2.3962	-0.1572	3.4625[a]	3.3864[a]
	(0.9372)	(0.9251)	(1.2448)	(2.0122)	(3.2084)	(0.9289)	(0.9548)
Observations	104	104	104	67	57	104	103
Adjusted R^2	63%	63%	63%	49%	47%	63%	62%
Adj. R^2 without infrastructure or culture	63%	63%	63%	48%	45%	63%	62%
Adj. R^2 without education	50%	54%	50%	44%	42%	51%	52%

Note: a = significant at the 1% level, b = significant at the 5% level, and c = significant at the 10% level.

Table 6: Regional GDP per capita, Institutions, Infrastructure, and Culture

Ordinary least square regressions of the log of regional GDP per capita. All regressions include years of education, logarithm of population, temperature, inverse distance to coast, and the logarithm of per capita oil production and reserves. In addition, regressions include measures of: (1) institutions (Panel A) and (2) infrastructure and culture (Panel B). Robust standard errors are shown in parenthesis. For comparison, the bottom panel shows the adjusted R^2 of two alternative specifications: (1) a regression with all regressors except the measure of institutions or culture; and (2) a regression with all regressors except education. All variables are described in Table 2.

Panel A: Institutions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Years of Education in the Region	0.2758[a]	0.3056[a]	0.3620[a]	0.3439[a]	0.3343[a]	0.3267[a]	0.3273[a]	0.3166[a]	0.4141[a]
	(0.0172)	(0.0298)	(0.0288)	(0.0481)	(0.0310)	(0.0218)	(0.0215)	(0.0207)	(0.0229)
Ln(Population in the Region)	0.0126	-0.0185	-0.0175	-0.0442	-0.0191	-0.0087	-0.0098	-0.0113	-0.0026
	(0.0168)	(0.0495)	(0.0536)	(0.0613)	(0.0432)	(0.0316)	(0.0312)	(0.0305)	(0.0229)
Temperature	-0.0140[c]	-0.0101	-0.0086	-0.0015	-0.0064	-0.0093	-0.0106	-0.0131	0.0016
	(0.0084)	(0.0096)	(0.0078)	(0.0122)	(0.0093)	(0.0086)	(0.0086)	(0.0081)	(0.0059)
Inverse Distance to Coast	0.4971[a]	0.4647	0.8290[c]	0.1810	0.2703	0.4054	0.5133[c]	0.4420	0.0913
	(0.1441)	(0.3293)	(0.4273)	(0.4312)	(0.3041)	(0.2636)	(0.2822)	(0.2788)	(0.3460)
Ln(Oil)	0.1752[a]	-0.0578	0.1555	-0.0584	-0.0473	-0.0224	-0.0040	-0.0170	0.1834
	(0.0578)	(0.1283)	(0.1319)	(0.2503)	(0.0862)	(0.1081)	(0.1113)	(0.0735)	(0.1160)
Informal Payments		-0.0089							
		(0.0353)							
ln(Tax Days)			-0.0479						
			(0.0630)						
Ln(Days without electricity)				0.0001					
				(0.0764)					
Security costs					-0.0004				
					(0.0060)				
Access to land						-0.1900			
						(0.1457)			
Access to finance							-0.0935		
							(0.1536)		
Government Predictability								-0.1251	
								(0.1426)	
Doing Business Percentile Rank									-0.6199[c]
									(0.3437)
Constant	6.3853[a]	6.5073[a]	5.7640[a]	6.8622[a]	6.4507[a]	6.3453[a]	6.2816[a]	6.4790[a]	6.3186[a]
	(0.1947)	(0.7043)	(0.8220)	(0.7867)	(0.5993)	(0.4664)	(0.4827)	(0.4629)	(0.4428)
Observations	1,469	338	255	216	352	387	381	368	172
Number of countries	104	73	55	72	76	77	76	72	17
R^2 Within	42%	58%	66%	59%	60%	62%	62%	63%	69%
R^2 Between	60%	64%	64%	53%	58%	60%	60%	63%	39%
R^2 Overall	62%	59%	60%	49%	53%	55%	55%	56%	51%
Within R^2 without institution	42%	57%	66%	59%	60%	62%	62%	62%	67%
Within R^2 without education	9%	11%	14%	10%	9%	6%	5%	7%	9%
Between R^2 without institution	60%	64%	63%	53%	58%	60%	60%	63%	41%
Between R^2 without education	42%	25%	20%	21%	26%	35%	39%	45%	50%
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: a = significant at the 1% level, b = significant at the 5% level, and c = significant at the 10% level.

Table 6: Regional GDP per capita, Institutions, Infrastructure, and Culture (Cont)

Panel B: Infrastructure and Culture

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Years of Education in the Region	0.2758[a]	0.2713[a]	0.2627[a]	0.3021[a]	0.2986[a]	0.2644[a]	0.2719[a]
	(0.0172)	(0.0187)	(0.0197)	(0.0286)	(0.0305)	(0.0181)	(0.0175)
Ln(Population in the Region)	0.0126	0.0101	0.0023	0.0091	0.0138	0.0170	0.0115
	(0.0168)	(0.0168)	(0.0184)	(0.0187)	(0.0193)	(0.0173)	(0.0157)
Temperature	-0.0140[c]	-0.0142[c]	-0.0166[c]	-0.0015	-0.0038	-0.0154[c]	-0.0140[c]
	(0.0084)	(0.0085)	(0.0085)	(0.0060)	(0.0056)	(0.0090)	(0.0080)
Inverse Distance to Coast	0.4971[a]	0.4872[a]	0.4626[a]	0.4750[c]	0.4093	0.4351[a]	0.5162[a]
	(0.1441)	(0.1427)	(0.1438)	(0.2590)	(0.2713)	(0.1358)	(0.1450)
Ln(Oil)	0.1752[a]	0.1793[a]	0.1864[a]	0.0534	0.0354	0.1922[a]	0.1772[a]
	(0.0578)	(0.0584)	(0.0582)	(0.0669)	(0.0572)	(0.0613)	(0.0591)
Ln(Power line density)		0.0199					
		(0.0198)					
Ln(Travel time)			-0.0456[c]				
			(0.0231)				
Trust in others				-0.0611			
				(0.0868)			
Civic values					-0.0040		
					(0.0231)		
Ln(Number of ethnic groups)						-0.0504[b]	
						(0.0249)	
Probability of same language							0.1723
							(0.2067)
Constant	6.3853[a]	6.4350[a]	6.9287[a]	6.0940[a]	6.0196[a]	6.5272[a]	6.2956[a]
	(0.1947)	(0.1928)	(0.3351)	(0.2863)	(0.3245)	(0.1679)	(0.2337)
Observations	1,469	1,469	1,469	699	635	1,468	1,445
Number of countries	104	104	104	65	70	104	103
R^2 Within	42%	42%	43%	49%	48%	42%	42%
R^2 Between	60%	60%	60%	50%	50%	60%	60%
R^2 Overall	62%	62%	61%	50%	47%	62%	62%
Within R^2 without institution	42%	42%	42%	49%	48%	42%	42%
Within R^2 without education	9%	13%	17%	10%	10%	14%	11%
Between R^2 without institution	60%	60%	60%	51%	50%	60%	59%
Between R^2 without education	42%	51%	47%	7%	17%	47%	50%
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: a = significant at the 1% level, b = significant at the 5% level, and c = significant at the 10% level.

Table 7: National GDP per capita and commonly used measures of institutions

Ordinary least square regressions of the log of GDP per capita at the country level. All regressions include the years of education, logarithm of population, temperature, inverse distance to coast, and the logarithm of per capita oil production and reserves. In addition, regressions include the following variables: (1) Autocracy; (2) Executive constraints; (3) Expropriation riks; (4) Proportional representation; and (5) Corruption. Robust standard errors are shown in parenthesis. For comparison, the bottom panel shows the adjusted R^2 of two alternative specifications: (1) a regression with all regressors except the measure of institutions or culture; and (2) a regression with all regressors except education.

	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	0.2567[a]	0.2200[a]	0.2069[a]	0.1626[a]	0.2448[a]	0.1850[a]
	(0.0308)	(0.0433)	(0.0438)	(0.0480)	(0.0363)	(0.0351)
Ln(Population)	0.0683[c]	0.0354	0.0559	-0.0356	0.0732	0.0504
	(0.0410)	(0.0487)	(0.0470)	(0.0482)	(0.0533)	(0.0370)
Temperature	-0.0189[c]	-0.0179	-0.0135	0.0024	-0.0181	-0.0100
	(0.0106)	(0.0118)	(0.0109)	(0.0106)	(0.0126)	(0.0104)
Inverse Distance to Coast	2.9646[a]	2.3421[a]	2.3853[a]	2.3974[a]	2.9603[a]	1.9906[a]
	(0.5742)	(0.7800)	(0.6050)	(0.5941)	(0.6208)	(0.5463)
Ln(Oil)	0.9503[a]	0.7877[c]	1.0708[a]	0.8965[a]	1.0720[b]	0.9928[a]
	(0.1373)	(0.4564)	(0.1729)	(0.1100)	(0.4094)	(0.2013)
Autocracy		-0.5994[a]				
		(0.2184)				
Executive Constraints			0.1633[b]			
			(0.0696)			
Expropriation Risk				0.3952[a]		
				(0.0986)		
Proportional Representation					0.3972[c]	
					(0.2328)	
Corruption						0.2130[a]
						(0.0479)
Constant	3.5771[a]	5.3781[a]	3.7896[a]	3.1830[b]	3.2958[a]	4.1183[a]
	(0.9416)	(1.3861)	(1.0059)	(1.3630)	(1.0503)	(0.8118)
Observations	103	80	101	81	97	103
Adjusted R^2	63%	67%	65%	70%	63%	69%
Adj. R^2 without institution	63%	64%	63%	63%	62%	63%
Adj. R^2 without education	50%	60%	59%	67%	52%	63%

Note: a = significant at the 1% level, b = significant at the 5% level, and c = significant at the 10% level.

Table 8: Firm level productivity and Regional Human Capital

The table reports fixed effect regressions for for the following two dependent variables: (1) logarithm of sales per employee; and (2) logarithm of wages per employee. All regressions include country and industry fixed effects. Errors are clustered at the regional level. The independent variables include: (1) Temperature, (2) Inverse Distance to Coast, (3) Ln(Oil), (4) Ln(Power line density), (5) Access to finance, (6) Years of Education in the Region, (7) Ln(Population in the Region), (8) Years of Education of manager, (9) Ln(Employees), (10) Years of Education of workers, (11) Ln(Expenditure on energy / employee), and (12) Ln(Property, Plant, Equipment / employees). All variables are described in Table 2.

Panel A: Basic Specification

	Dependent Variable:							
	Logarithm of Sales per employee				***Logarithm of Wages per employee***			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Years of Education in the Region	0.0655[a]	0.0639[a]	0.0954[a]	0.0950[a]	0.0580[a]	0.0577[a]	0.0840[a]	0.0843[a]
	(0.0202)	(0.0185)	(0.0280)	(0.0279)	(0.0162)	(0.0159)	(0.0233)	(0.0234)
Ln(Population in the Region)	0.0920[a]	0.0803[a]	0.1437[a]	0.1409[a]	0.0682	0.0622	0.0135	0.0159
	(0.0321)	(0.0297)	(0.0501)	(0.0504)	(0.0425)	(0.0418)	(0.0352)	(0.0354)
Years of Education of manager	0.0534[a]	0.0352[a]	0.0257[a]	0.0243[a]	0.0315[a]	0.0215[a]	0.0118[a]	0.0131[a]
	(0.0047)	(0.0048)	(0.0062)	(0.0057)	(0.0038)	(0.0036)	(0.0044)	(0.0042)
Ln(Employees)	.	0.1497[a]	.	0.0113	.	0.0827[a]	.	-0.0095
	.	(0.0154)	.	(0.0176)	.	(0.0150)	.	(0.0108)
Years of Education of workers	0.0349[a]	0.0279[a]	0.0384[a]	0.0378[a]	0.0195[a]	0.0151[a]	0.0146[a]	0.0152[a]
	(0.0053)	(0.0054)	(0.0056)	(0.0058)	(0.0036)	(0.0036)	(0.0033)	(0.0033)
Ln(Expenditure on energy / employee)	0.3577[a]	0.3554[a]	.	.	0.2248[a]	0.2232[a]	.	.
	(0.0185)	(0.0177)	.	.	(0.0173)	(0.0172)	.	.
Ln(Property, Plant, Equipment / employees)	.	.	0.3258[a]	0.3250[a]	.	.	0.1787[a]	0.1794[a]
	.	.	(0.0132)	(0.0136)	.	.	(0.0086)	(0.0089)
Constant	5.1202[a]	5.0055[a]	4.8529[a]	4.8850[a]	5.1007[a]	5.0322[a]	6.6732[a]	6.6461[a]
	(0.3706)	(0.3373)	(1.1885)	(1.1887)	(0.5225)	(0.5199)	(0.7223)	(0.7248)
Observations	13,248	13,248	19,305	19,305	12,782	12,782	19,209	19,209
Number of Countries	29	29	22	22	27	27	22	22
Within R^2	30%	32%	31%	31%	20%	21%	13%	13%
Between R^2	90%	90%	59%	59%	88%	87%	57%	57%
Overall R^2	74%	74%	54%	54%	69%	68%	44%	44%
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: a = significant at the 1% level, b = significant at the 5% level, and c = significant at the 10% level.

Table 8 A: Firm level productivity and Regional Human Capital

	Dependent Variable:							
	Logarithm of Sales per employee				*Logarithm of Wages per employee*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Temperature	0.0102	0.0123	0.0171[c]	0.0171[c]	-0.0117	-0.0104	0.0149[a]	0.0149[a]
	(0.0129)	(0.0118)	(0.0087)	(0.0087)	(0.0115)	(0.0110)	(0.0051)	(0.0051)
Inverse Distance to Coast	0.2755	0.1574	0.8401[a]	0.8380[a]	-0.1088	-0.1830	0.0437	0.0530
	(0.3492)	(0.3450)	(0.2565)	(0.2565)	(0.3266)	(0.3294)	(0.2084)	(0.2095)
Ln(Oil)	-0.8864[a]	-0.7033[b]	0.2600	0.2639	-0.6647[b]	-0.5661[c]	0.1110	0.0945
	(0.2793)	(0.3070)	(0.5745)	(0.5723)	(0.3179)	(0.3380)	(0.3702)	(0.3681)
Ln(Power line density)	-0.0246	-0.0266	0.1157[a]	0.1155[a]	-0.0125	-0.0159	0.0285	0.0293
	(0.0317)	(0.0313)	(0.0430)	(0.0429)	(0.0333)	(0.0334)	(0.0412)	(0.0412)
Access to finance	-0.0747	-0.0727	-0.0635	-0.0633	-0.1051	-0.1059	-0.1238	-0.1248
	(0.0772)	(0.0715)	(0.0989)	(0.0990)	(0.0877)	(0.0839)	(0.0778)	(0.0783)
Years of Education in the Region	0.0685[b]	0.0721[a]	-0.0203	-0.0202	0.0666[b]	0.0708[a]	0.0602[b]	0.0599[b]
	(0.0262)	(0.0247)	(0.0321)	(0.0321)	(0.0267)	(0.0264)	(0.0298)	(0.0298)
Ln(Population in the Region)	0.1046[a]	0.0906[a]	0.0454	0.0448	0.0758[c]	0.0692	-0.0353	-0.0327
	(0.0344)	(0.0316)	(0.0381)	(0.0386)	(0.0428)	(0.0422)	(0.0289)	(0.0289)
Years of Education of manager	0.0531[a]	0.0351[a]	0.0274[a]	0.0270[a]	0.0315[a]	0.0216[a]	0.0129[a]	0.0145[a]
	(0.0047)	(0.0047)	(0.0061)	(0.0054)	(0.0037)	(0.0036)	(0.0043)	(0.0040)
Ln(Employees)	.	0.1486[a]	.	0.0030	.	0.0821[a]	.	-0.0129
	.	(0.0153)	.	(0.0175)	.	(0.0150)	.	(0.0109)
Years of Education of workers	0.0344[a]	0.0276[a]	0.0410[a]	0.0408[a]	0.0193[a]	0.0151[a]	0.0158[a]	0.0165[a]
	(0.0053)	(0.0054)	(0.0059)	(0.0061)	(0.0037)	(0.0037)	(0.0035)	(0.0034)
Ln(Expenditure on energy / employee)	0.3574[a]	0.3552[a]	.	.	0.2250[a]	0.2235[a]	.	.
	(0.0184)	(0.0176)	.	.	(0.0173)	(0.0172)	.	.
Ln(Property, Plant, Equipment / employees)	.	.	0.3193[a]	0.3190[a]	.	.	0.1759[a]	0.1768[a]
	.	.	(0.0123)	(0.0127)	.	.	(0.0084)	(0.0088)
Constant	4.9971[a]	5.0412[a]	5.7485[a]	5.7554[a]	6.1865[a]	6.1040[a]	7.2040[a]	7.1742[a]
	(0.5219)	(0.4805)	(1.0095)	(1.0110)	(0.7436)	(0.7271)	(0.6329)	(0.6343)
Observations	13,248	13,248	19,305	19,305	12,782	12,782	19,209	19,209
Number of Countries	29	29	22	22	27	27	22	22
Within R^2	31%	32%	33%	33%	20%	21%	14%	14%
Between R^2	90%	90%	57%	57%	80%	78%	55%	55%
Overall R^2	73%	72%	59%	59%	65%	64%	52%	52%
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: a = significant at the 1% level, b = significant at the 5% level, and c = significant at the 10% level.

Table 9: Regional Human Capital, the size of establishments, and participation in the economy

The table reports fixed effect regressions for for the following three dependent variables: (1) logarithm of the number of employees per establishment; (3) logarithm of the number of employees per capita; and (4) logarithm of the number of employees working in firms that employ at least 100 employees as a percent of total employment. All regressions include the number of years of education. All variables are described in Table 2

	Dependent Variable:			
	Ln(Establishments/Population)	Ln(Employees/Establishments)	Ln(Employees/Population)	Ln(Employees Big Firms/Employees)
Years of Education in the Region	0.2967[a]	0.1233[a]	0.3418[a]	0.2445[a]
	(0.0314)	(0.0227)	(0.0273)	(0.0374)
Constant	-5.8626[a]	0.8855[a]	-4.3992[a]	-3.6568[a]
	(0.2571)	(0.2093)	(0.2119)	(0.4299)
Observations	951	983	988	501
Adjusted R^2	92%	83%	94%	95%
Country Fixed Effects	Yes	Yes	Yes	Yes

Note: a = significant at the 1% level, b = significant at the 5% level, and c = significant at the 10% level.

Figure 3: Regions in the database

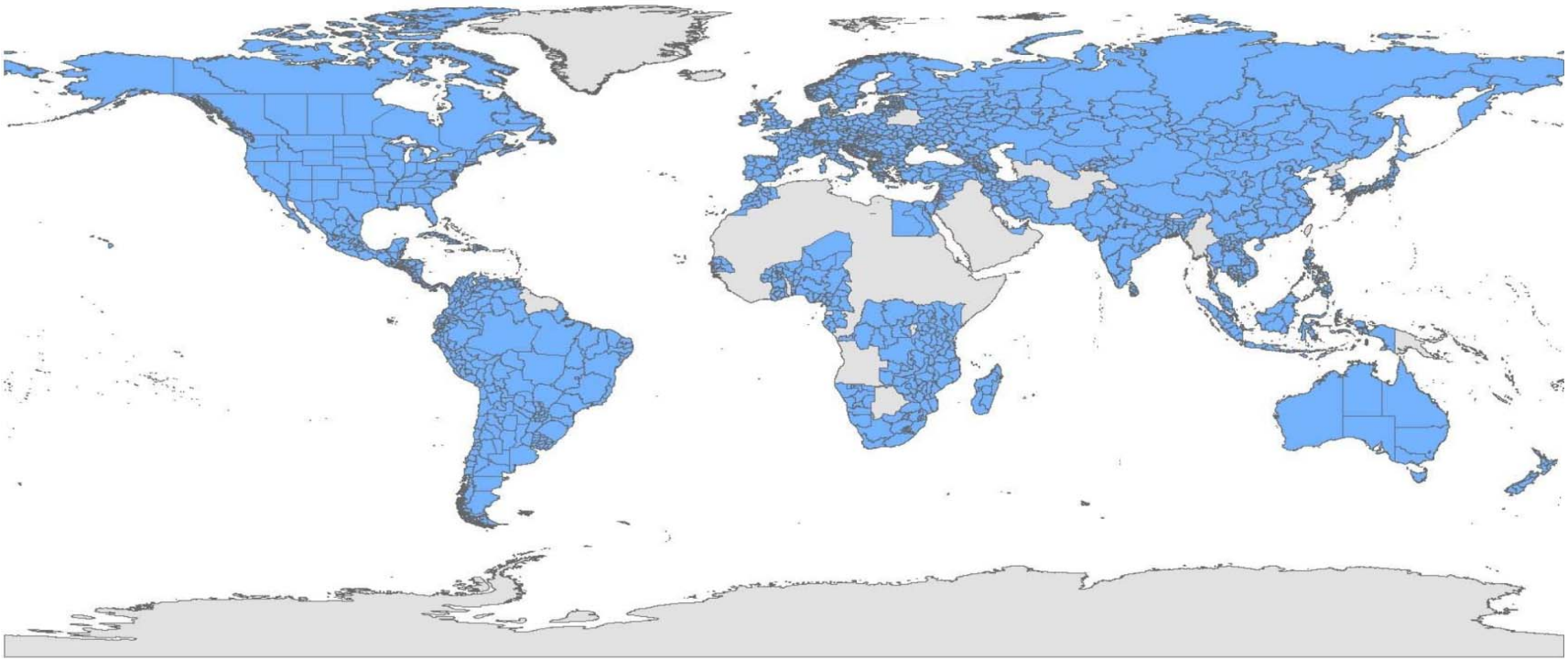


Figure 4: Within-country standard deviation of GDP per capita and development

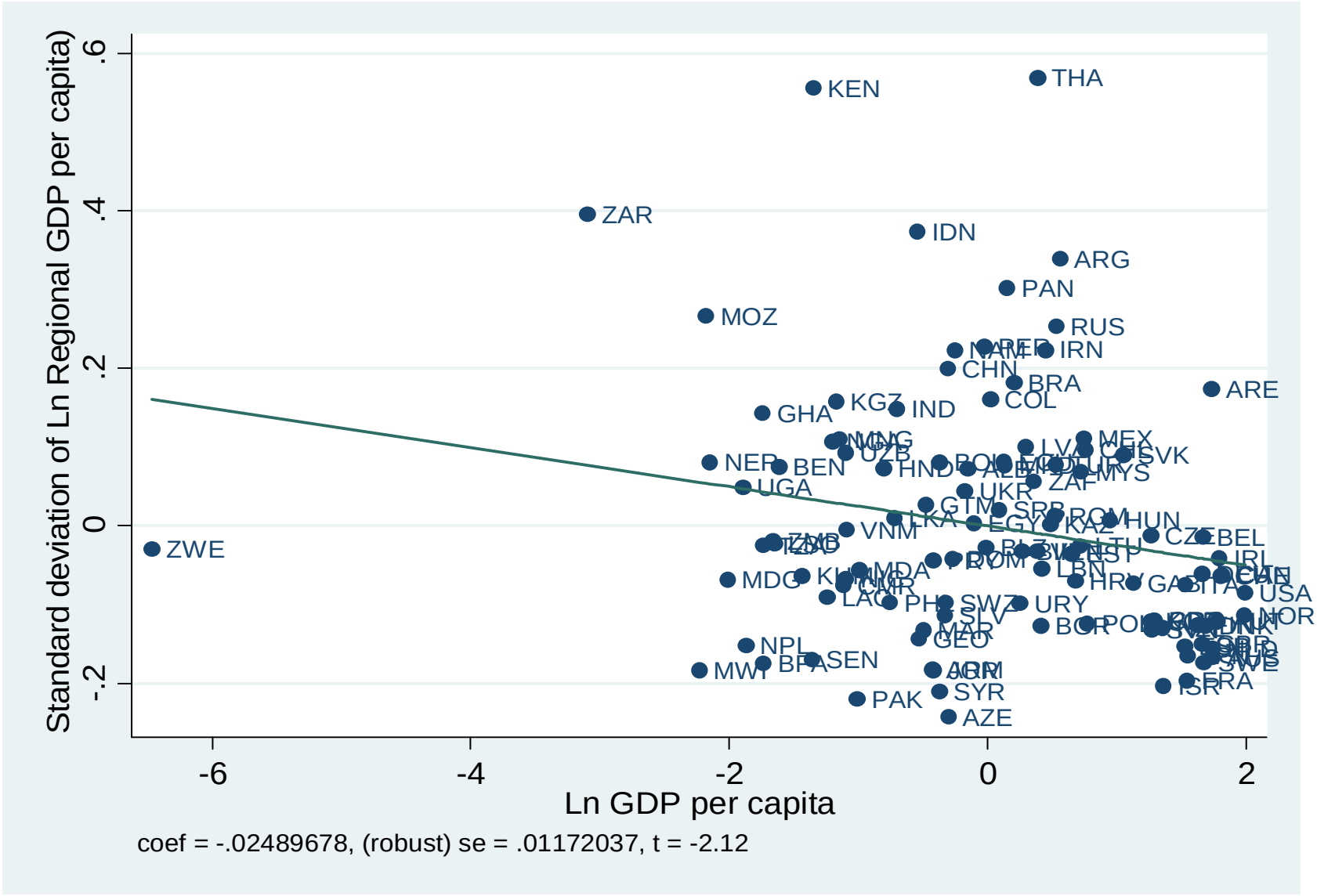


Figure 5: Within-country standard deviation of years of education and development

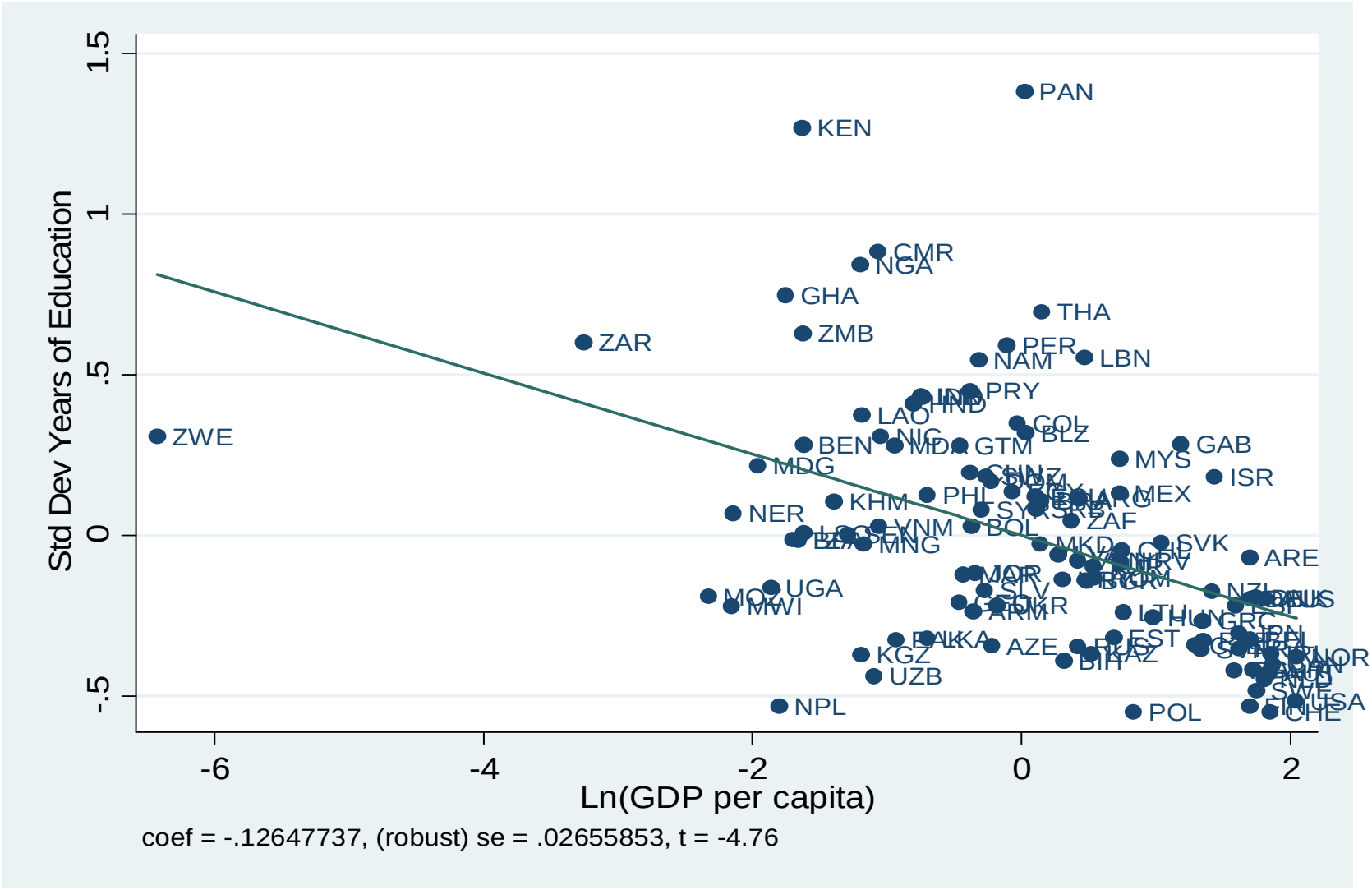


Figure 6: Partial Correlation Graph of (Log) GDP per capita and Years of Education

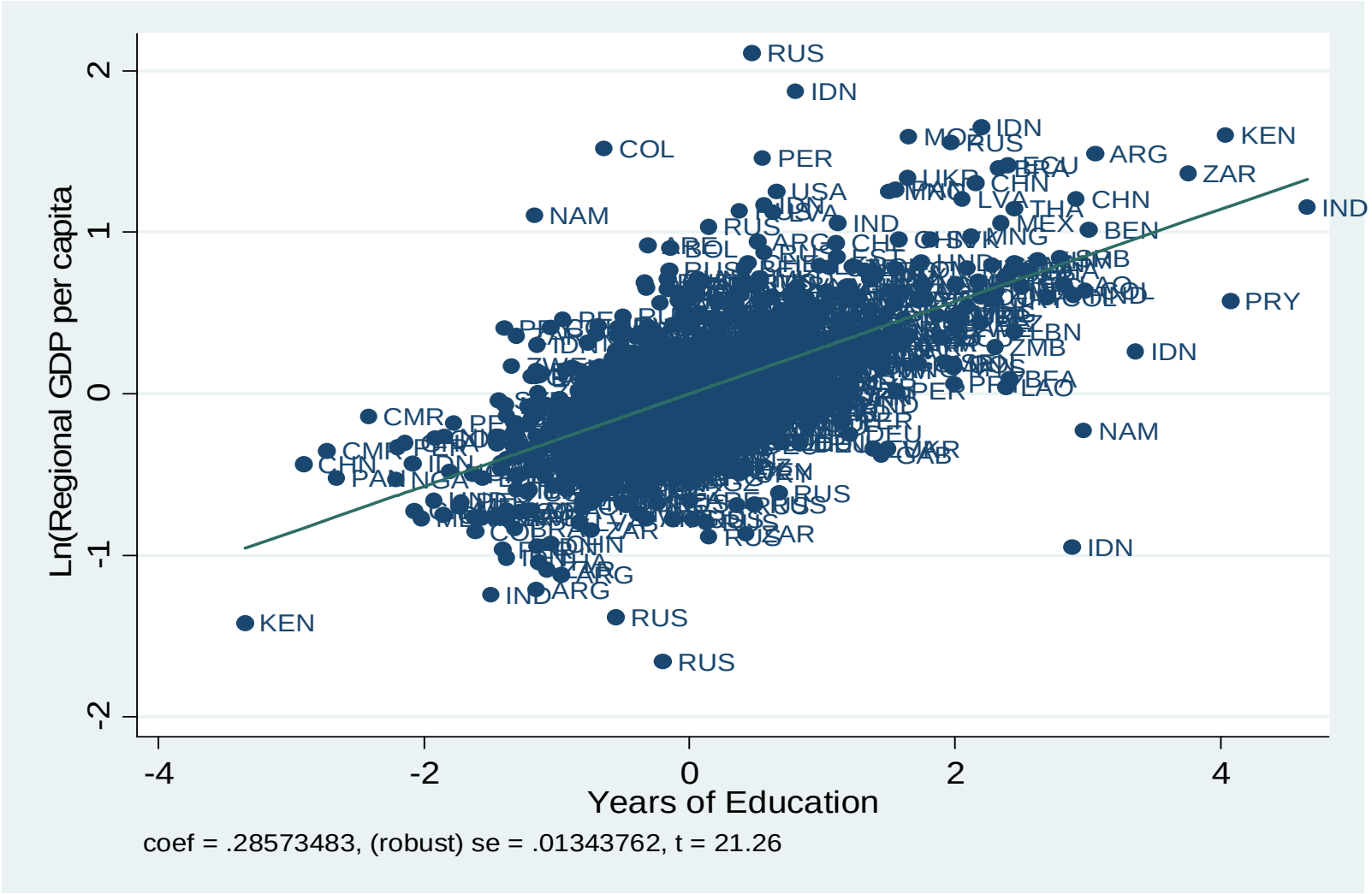
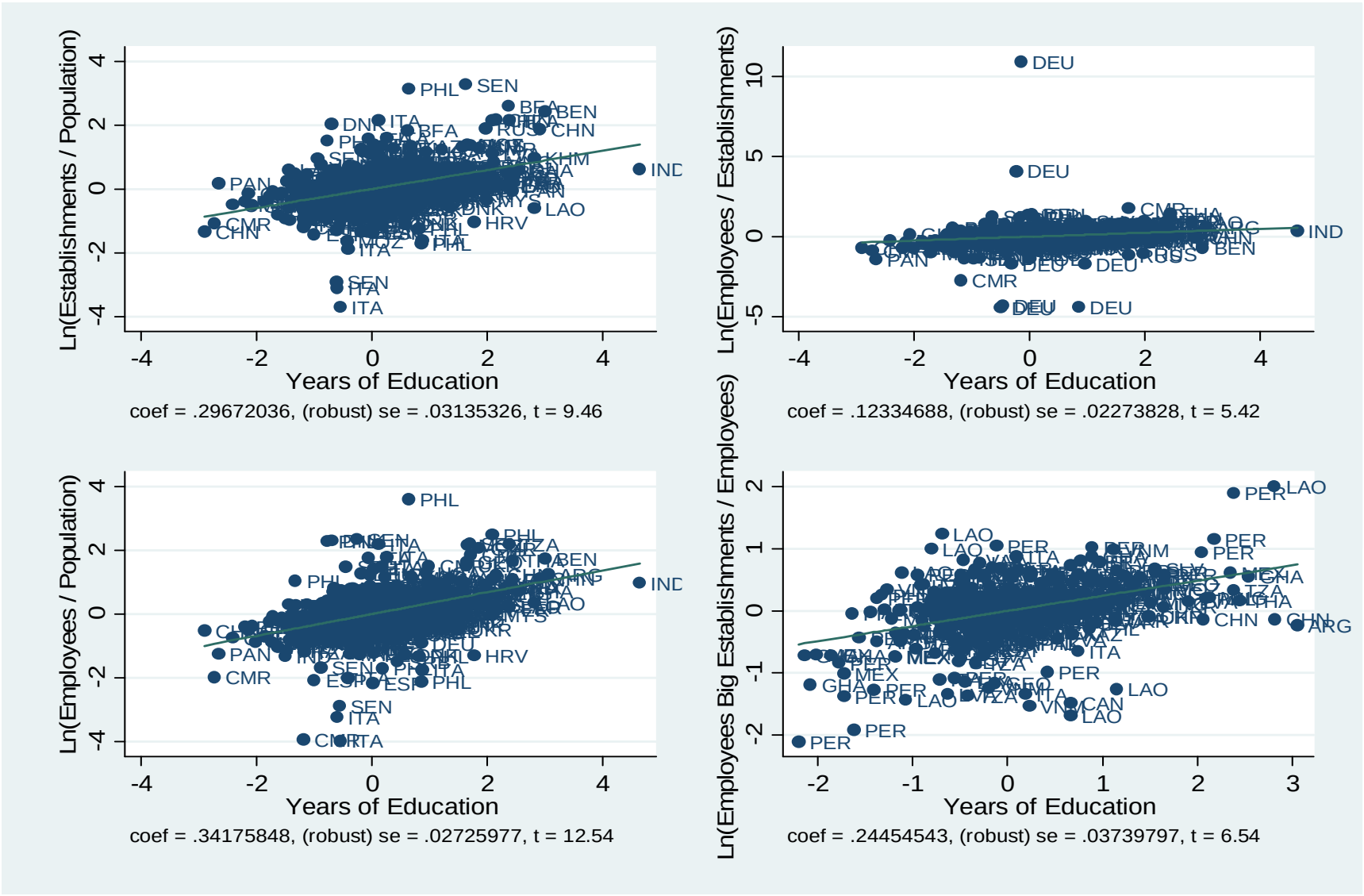


Figure 7: Years of Education and Participation in the Official Economy



Appendix A: Reporting level for countries in our dataset

The table identifies the reporting level for the regions in our database. The table splits countries in three main groups: (1) countries where data is reported at the first-order administrative regions; (2) countries where data is reported for economic or statistical regions and where first-order administrative regions are equivalent to provinces, states or derpartments; and (3) countries where data is reported for economic or statistical regions and where first-order administrative regions are equivalent to counties, boroughs, cities, districts or municipalities. The table also subdivides countries based on the reason why the first-order administrative regions are different than the reporting regions for each of these three groups of countries.

	Number of countries	First-order administrative regions	Regions in our dataset	Country names (number of first-order administrative regions lost)
1. Reporting done at the first-order administrative level:	**79**	**1,362**	**1,328**	
Our regions match first-order administrative level:	60	934	934	
Differences due to :				
Missing information for some region	7	148	130	France (4 overseas departments), Grece (1 self-governing monastic state), India (2 union territories & 1 island), Morocco (2 disputed territories), Pakistan (1 Tribal area), Tanzania (5 islands), Venezuela (2)
Aggregation of some regions	6	183	168	Croatia (1), Mozambique (1), New Zealand (3), Russia (3), Serbia (6), Switzerland (1)
Political change during sample period	6	97	96	Canada (1), Chile (2), Denmark(-10), Ecuador (2), Peru (2), Senegal (4)
2. Reporting done for economic or statistical regions. First-order administrative regions are equivalent to provinces, states or departments.	**22**	**691**	**177**	
Most data collected for statistical regions	6	78	44	Belgium(-8), Cezch Republic(6), Finland(1), Nepal(9), Portugal (13),Sweden(13)
GDP per capita collected for statistical regions	4	88	37	Dominican Republic(23), Kazakhstan(10), Cambodia(9), South Korea(9)
Education collected for statistical regions	12	525	96	Burkina Faso(32), Bulgaria(22), Egypt(22), Gabon(5), Guatemala(14), Nigeria(31), Philippines(65), Thailand(71), Turkey(69), Romania(34), Uzbekistan(9), Vietnam(55)
3. Reporting done for economic or statistical regions. First-order administrative regions are equivalent to counties, boroughs, cities, districts, or municipalities.	**9**	**782**	**64**	
Most data collected for statistical regions	7	725	52	Azerbajan (66), Great Britain (217), Ireland(32), Macedonia(76), Malawi(25), Slovenia(181), Uganda(76)
Education collected for statistical regions	2	57	12	Hungary(13), Moldova(32)
Total in the sample	**110**	**2,835**	**1,569**	

APPENDIX OF DATA SOURCES: REGIONAL GDP

Code	Country	Type of Data	Source	Available link
ALB	Albania	GDP	Data from HDR 2002	
ARE	United Arab Emirates	GDP	Data from HDR 1997 in arabic	
ARG	Argentina	GDP	1990-2001 Data from Ministry of interior	http://www.ec.gba.gov.ar/Estadistica/FTP/pbg/pbg3.html
ARM	Armenia	Expenditure	National Statistics Office	http://www.armstat.am/file/article/marz_07_e_22.pdf
AUS	Australia	GDP	Data from OECDStats	http://stats.oecd.org/WBOS/index.aspx
AUT	Austria	GDP	Data from OECDStats	http://stats.oecd.org/WBOS/index.aspx
AZE	Azerbaijan	Income	National Statistics Office	http://74.125.47.132/search?q=cache:http://www.azstat.org/statinfo/budget_households/en/003.shtml
BEL	Belgium	GDP	Data from OECDStats	http://stats.oecd.org/WBOS/index.aspx
BEN	Benin	GDP	Data from HDR 2007/2008 and 2003	
BFA	Burkina Faso	GDP	Data from HDR for GDP per capita.	
BGD	Bangladesh	NA		
BGR	Bulgaria	GDP	Data from HDR 2003, 2002 and 2001	
BIH	Bosnia and Herzegovina	GDP	National Statistics Offices	http://www.bhas.ba/Arhiva/2007/brcko/PODACI%201-08.pdf
BLZ	Belize	Expenditure	Data from LSMS 2002	http://www.statisticsbelize.org.bz/dms20uc/dm_filedetails.asp?action=d&did=13
BOL	Bolivia	GDP	National Statistics Office	http://www.ine.gov.bo/indice/visualizador.aspx?ah=PC0104010201.HTM
BRA	Brazil	GDP	National Statistics Office	http://www.ibge.gov.br/home/estatistica/economia/contasregionais/2002_2005/contasregionais2002_2005.pdf
CAN	Canada	GDP	Data from OECDStats	http://stats.oecd.org/WBOS/index.aspx
CHE	Switzerland	GDP	National Statistics Office	http://www.bfs.admin.ch/bfs/portal/fr/index/infothek/lexikon/bienvenue___login/blank/zugang_lexikon.Document.20896.xls
CHL	Chile	GDP	National Statistics Office	http://www.bcentral.cl/publicaciones/estadisticas/actividad-economica-gasto/aeg07a.htm
CHN	China	GDP	Data from National Statistics Yearbooks 2006, 2002, 1998 and 1996	http://www.stats.gov.cn/english/statisticaldata/yearlydata/YB1998e/C3-8E.htm
CMR	Cameroon	Expenditure	National Statistics Office	http://www.statistics-cameroon.org/archive/ECAM/ECAM2001/survey0/data/ECAM2001/Documentation/ECAM%20II%20-%20Rapport%20principal.pdf
COL	Colombia	GDP	National Statistics Office	http://www.dane.gov.co/index.php?option=com_content&task=category&sectionid=33&id=148&Itemid=705
CRI	Costa Rica	NA		
CUB	Cuba	Wages	Monthly wages from HDR 1996	
CZE	Czech Republic	GDP	Data from OECDStats	http://stats.oecd.org/WBOS/index.aspx
DEU	Germany	GDP	Data from OECDStats	http://stats.oecd.org/WBOS/index.aspx
DNK	Denmark	GDP	Data from OECDStats	http://stats.oecd.org/WBOS/index.aspx
DOM	Dominican Republic	GDP	National Statistics Office	
ECU	Ecuador	GDP	National Statistics Office	
EGY	Egypt	GDP	Data from HDRs 2008, 2005, 2004, 2003, 2001; data from 2006 excluded	
ESP	Spain	GDP	Data from OECDStats	http://stats.oecd.org/WBOS/index.aspx
EST	Estonia	GDP	Data from National Statistics Office	http://pub.stat.ee/px-web.2001/I_Databas/Economy_regional/23National_accounts/01Gross_Domestic_product_(GDP)/14Regional_gross_domestic_product/14Regional_gross_domestic_product.asp
FIN	Finland	GDP	Data from OECDStats	http://stats.oecd.org/WBOS/index.aspx
FRA	France	GDP	Data from OECDStats	http://stats.oecd.org/WBOS/index.aspx
GAB	Gabon	Expenditure	Data from HDR 2005	
GBR	United Kingdom	GDP	Data from OECDStats	http://stats.oecd.org/WBOS/index.aspx
GEO	Georgia	GDP	Data from HDR 2002.	http://www.undp.org.ge/nhdr2001-02/chpt1.htm
GHA	Ghana	Income	Data from Living Standards Measurement Survey Reports for 1998/9 and 1991/2	http://siteresources.worldbank.org/INTLSMS/Resources/3358986-1181743055198/3877319-1190221709991/G3report.pdf
GRC	Greece	GDP	Data from OECDStats	http://stats.oecd.org/WBOS/index.aspx
GTM	Guatemala	GDP	Data from HDR 2007/2008 annex	http://cms.fideck.com/userfiles/desarrollohumano.org/File/8012264236003654.pdf
HND	Honduras	GDP	Data from HDR 2006	
HRV	Croatia	GDP	Data from National Statistics Office	http://www.dzs.hr/default_e.htm
HUN	Hungary	GDP	Data from OECDStats	http://stats.oecd.org/WBOS/index.aspx
IDN	Indonesia	GDP	Data from National Statistics Office	http://www.bps.go.id/sector/nra/grdp/table1.shtml
IND	India	GDP	National Statistics Office	http://mospi.nic.in/6_gsdp_cur_9394ser.htm
IRL	Ireland	GDP	Data from OECDStats	http://stats.oecd.org/WBOS/index.aspx
IRN	Iran	GDP	Data from National Statistics Office	http://www.sci.org.ir/content/userfiles/_sci_en/sci_en/sel/year85/f21/CS_21_4.HTM
ISR	Israel	GDP	National Statistics Office	
ITA	Italy	GDP	Data from OECDStats	http://stats.oecd.org/WBOS/index.aspx
JOR	Jordan	GDP	Data from HDR 2004	
JPN	Japan	GDP	Data from OECDStats	http://stats.oecd.org/WBOS/index.aspx
KAZ	Kazakhstan	Income	LSMS 1996, World Bank	http://siteresources.worldbank.org/INTLSMS/Resources/3358986-1181743055198/3877319-1181930718899/finrep1.pdf
KEN	Kenya	GDP	Data from HDRs for 2006, 2005, 2003, 2001 and 1999	
KGZ	Kyrgyz Republic	GDP	Data from HDR 2005, 2001	

Code	Country	Type of Data	Source	Available link
KHM	Cambodia	Expenditure	Data from Poverty profile of Cambodia 2004; Daily consumption	http://www.mop.gov.kh/Situationandpolicyanalysis/PovertyProfile/tabid/191/Default.aspx
KOR	Korea, Rep.	GDP	Data from OECDStats	http://stats.oecd.org/WBOS/index.aspx
LAO	Lao PDR	C+I+G	Data from HDR 2006; Consumption, Investment and Government Expenditure	
LBN	Lebanon	GDP	Data from HDR 2001	
LKA	Sri Lanka	GDP	Data from HDR 1998 and National Statistics Office	http://www.cbsl.gov.lk/pics_n_docs/08_statistics/_docs/xls_real_sector/table1.17.xls
LSO	Lesotho	GDP	Data from HDR 2006	
LTU	Lithuania	GDP	Data from National Statistics Office	http://db1.stat.gov.lt/statbank/SelectVarVal/Define.asp?MainTable=M2010210&PLanguage=1&PXSId=0&ShowNews=OFF
LVA	Latvia	GDP	National Statistics Office	http://data.csb.gov.lv/Dialog/varval.asp?ma=02-02a&ti=2-2.+GROSS+DOMESTIC+PRODUCT+BY+STATISTICAL+REGION,+CITY+AND+DISTRICT&path=../DATABASEEN/ekfin/Annual%20statistical%20data/02.%20Gross%20domestic%20product/&lang=1
MAR	Morocco	GDP + Expenditure	Data from HDR 1999, 2003 and Enquete Nationale sur la Consommation et les Depenses des Menages 2000/2001	
MDA	Moldova	Wages	Data from 2007 Statistical Yearbook; monthly salary	http://www.statistica.md/public/files/Yearbook/Venit_pop_1999_2006_en.doc
MDG	Madagascar	GDP	Data from HDR 2003, 2000	
MEX	Mexico	GDP	Data from OECDStats	http://stats.oecd.org/WBOS/index.aspx
MKD	Macedonia, FYR	GDP	Data from National Statistics Office	http://www.stat.gov.mk/english/statistiki_eng.asp?ss=09.01&rbs=2
MNG	Mongolia	GDP	National Statistics Office	
MOZ	Mozambique	GDP	Data from HDR 2007, 2001	
MWI	Malawi	Expenditure	Data from Malawi INTEGRATED HOUSEHOLD SURVEY 2004-2005 and 1998	
MYS	Malaysia	GDP	Data from Chapter 5 of EIGHTH MALAYSIA PLAN 2001 - 2005	http://www.epu.jpm.my/new%20folder/development%20plan/RM8.htm
NAM	Namibia	Expenditure	Data from Namibia Household Income & Expenditure Survey 2003/2004; data is expenditu	http://www.npc.gov.na/publications/prenhies03_04.pdf
NER	Niger	GDP	Data from HDR 2004	
NGA	Nigeria	Income	2006 Annual Abstract of Statistics.	http://nigerianstat.gov.ng/annual_report.htm
NIC	Nicaragua	Expenditure	Data from HDR 2002	
NLD	Netherlands	GDP	Data from OECDStats	http://stats.oecd.org/WBOS/index.aspx
NOR	Norway	GDP	Data from OECDStats	http://stats.oecd.org/WBOS/index.aspx
NPL	Nepal	GDP	Data from HDR 2004, 2001 and 1998	
NZL	New Zealand	GDP	Data from OECDStats	http://stats.oecd.org/WBOS/index.aspx
PAK	Pakistan	GDP	Data from HDR 2003	
PAN	Panama	GDP	Data from National Statistics Office	http://www.contraloria.gob.pa/dec/
PER	Peru	GDP	Cuentas Nacionales del Peru, Producto Bruto Interno por Departmentos 2001-2006	http://www1.inei.gob.pe/biblioineipub/bancopub/est/lib0763/cuadros/c037.xls
PHL	Philippines	GDP	National Statistics Office	
POL	Poland	GDP	Data from OECDStats	http://stats.oecd.org/WBOS/index.aspx
PRT	Portugal	GDP	Data from OECDStats	http://stats.oecd.org/WBOS/index.aspx
PRY	Paraguay	GDP	Data from Atlas de Desarrollo Humano Paraguay 2007	http://www.undp.org.py/dh/?page=atlas
ROM	Romania	GDP	Data from National Statistics Office	http://www.insse.ro/cms/files/pdf/en/cp11.pdf
RUS	Russia	GDP	National Statistics Office	http://translate.google.com/translate?ie=UTF8&oe=UTF&u=http://www.gks.ru/bgd/regl/b07_14p/IssWWW.exe/Stg/d02/10-02.htm&hl=en&ie=UTF8&sl=ru&tl=en
SEN	Senegal	GDP	Data from HDR 2001	
SLV	El Salvador	GDP	Data from HDR 2007/2008, 2005, 2003, 2001; 1996 values were in 1994 prices	
SRB	Serbia	Income	Data from National Statistics Municipal Database	http://www.statserb.sr.gov.yu/Pod/epok.asp
SVK	Slovak Republic	GDP	Data from OECDStats	http://stats.oecd.org/WBOS/index.aspx
SVN	Slovenia	GDP	Data from National Statistics Office	http://www.stat.si/eng/novica_prikazi.aspx?id=1318
SWE	Sweden	GDP	Data from OECDStats	http://stats.oecd.org/WBOS/index.aspx
SWZ	Swaziland	GDP	Data from HDR 2008	
SYR	Syrian Arab Republic	GDP	Data from HDR 2005	
THA	Thailand	GDP	Data from Statistical Year Book Thailand 2002	http://web.nso.go.th/eng/en/pub/pub0.htm
TUR	Turkey	GDP	National Statistics Office	http://www.turkstat.gov.tr/VeriBilgi.do?tb_id=56&ust_id=16
TZA	Tanzania	GDP	National Statistics Office	
UGA	Uganda	GDP	Data from HDR 2007	
UKR	Ukraine	GDP	Data from National Statistics Office	http://www.ukrstat.gov.ua/operativ/operativ2008/vvp/vrp/vrp2008_e.htm
URY	Uruguay	GDP	Data from HDR 2005	
USA	United States	GDP	Data from OECDStats	http://stats.oecd.org/WBOS/index.aspx
UZB	Uzbekistan	GDP	Data from HDR 2007/8, 2000 and 1998	
VEN	Venezuela	GDP	Data from HDR 2000	
VNM	Vietnam	Wages	National Statistics Office	http://www.gso.gov.vn/Modules/Doc_Download.aspx?DocID=2097
ZAF	South Africa	GDP	National Statistics Office (table 16)	http://www.statssa.gov.za/publications/statsdownload.asp?PPN=P0441&SCH=4048
ZAR	Congo, Dem. Rep.	GDP	Data from HDR 2008	
ZMB	Zambia	GDP	Data from HDR 2007 and 2003	
ZWE	Zimbabwe	GDP	Data fom HDR 2003	

APPENDIX OF DATA SOURCES: REGIONAL YEARS OF EDUCATION

Code	Country	Source	Available Link
ALB	Albania	NA	
ARE	United Arab Emirates	Ministry of Economy, 2005 Census	http://www.economy.ae/English/economicandstatisticreports/statisticreports/pages/census2005.aspx
ARG	Argentina	Education Policy and Data Center (EPDC)	http://epdc.org/
ARM	Armenia	Education Policy and Data Center (EPDC)	http://epdc.org/
AUS	Australia	National Statistics Office	http://www.abs.gov.au/
AUT	Austria	Eurostat	http://epp.eurostat.ec.europa.eu/NavTree_prod/everybody/BulkDownloadListing
AZE	Azerbaijan	Education Policy and Data Center (EPDC)	http://epdc.org/
BEL	Belgium	Eurostat	http://epp.eurostat.ec.europa.eu/NavTree_prod/everybody/BulkDownloadListing
BEN	Benin	Education Policy and Data Center (EPDC)	http://epdc.org/
BFA	Burkina Faso	Education Policy and Data Center (EPDC)	http://epdc.org/
BGD	Bangladesh	Education Policy and Data Center (EPDC)	http://epdc.org/
BGR	Bulgaria	Eurostat	http://epp.eurostat.ec.europa.eu/NavTree_prod/everybody/BulkDownloadListing
BIH	Bosnia and Herzegovina	Education Policy and Data Center (EPDC)	http://epdc.org/
BLZ	Belize	Education Policy and Data Center (EPDC)	http://epdc.org/
BOL	Bolivia	Education Policy and Data Center (EPDC)	http://epdc.org/
BRA	Brazil	Integrated Public Use Microdata Series International (IPUMS)	https://international.ipums.org/international/
CAN	Canada	National Statistics Office	http://www40.statcan.gc.ca/l01/cst01/educ43a-eng.htm
CHE	Switzerland	Swiss Labor Force Survey (SLFS) SFSO	http://www.bfs.admin.ch/bfs/portal/de/index/themen/15/04/ind4.informations.40101.401.html
CHL	Chile	National Statistics Office	http://espino.ine.cl/CuadrosCensales/apli_excel.asp
CHN	China	National Statistics Office	http://www.stats.gov.cn/ndsj/information/nj97/C091A.END
CMR	Cameroon	Education Policy and Data Center (EPDC)	http://epdc.org/
COL	Colombia	National Statistics Office	http://190.25.231.246:8080/Dane/tree.jsf
CRI	Costa Rica	Education Policy and Data Center (EPDC)	http://epdc.org/
CUB	Cuba	NA	
CZE	Czech Republic	Eurostat	http://epp.eurostat.ec.europa.eu/NavTree_prod/everybody/BulkDownloadListing
DEU	Germany	Eurostat	http://epp.eurostat.ec.europa.eu/NavTree_prod/everybody/BulkDownloadListing
DNK	Denmark	National Statistics Office	http://www.statbank.dk/statbank5a/SelectVarVal/Define.asp?Maintable=RASU1&PLanguage=1
DOM	Dominican Republic	Education Policy and Data Center (EPDC)	http://epdc.org/
ECU	Ecuador	National Statistics Office	http://190.95.171.13/cgibin/RpWebEngine.exe/PortalAction?&MODE=MAIN&BASE=ECUADOR21&MAIN=WebServerMain.inl
EGY	Egypt	Education Policy and Data Center (EPDC)	http://epdc.org/
ESP	Spain	Eurostat	http://epp.eurostat.ec.europa.eu/NavTree_prod/everybody/BulkDownloadListing
EST	Estonia	National Statistics Office	http://pub.stat.ee/px-web.2001/Dialog/varval.asp?ma=PC414&ti=ECONOMICALLY+ACTIVE+POPULATION+BY+AGE,+EDUCATIONAL+ATTAINMENT+AND+ETHNIC+NATIONALITY*&path=../I_Databas/Population_census/06Economically_active_population/(=1
FIN	Finland	Eurostat	http://epp.eurostat.ec.europa.eu/NavTree_prod/everybody/BulkDownloadListing
FRA	France	Eurostat	http://epp.eurostat.ec.europa.eu/NavTree_prod/everybody/BulkDownloadListing
GAB	Gabon	Education Policy and Data Center (EPDC)	http://epdc.org/
GBR	United Kingdom	Eurostat	http://epp.eurostat.ec.europa.eu/NavTree_prod/everybody/BulkDownloadListing
GEO	Georgia	National Statistics Office (special request of data)	
GHA	Ghana	Education Policy and Data Center (EPDC)	http://epdc.org/
GRC	Greece	Eurostat	http://epp.eurostat.ec.europa.eu/NavTree_prod/everybody/BulkDownloadListing
GTM	Guatemala	Education Policy and Data Center (EPDC)	http://epdc.org/
HND	Honduras	Education Policy and Data Center (EPDC)	http://epdc.org/
HRV	Croatia	National Statistics Office	http://www.dzs.hr/Eng/censuses/Census2001/Popis/E01_01_07/E01_01_07.html
HUN	Hungary	Eurostat	http://epp.eurostat.ec.europa.eu/NavTree_prod/everybody/BulkDownloadListing
IDN	Indonesia	Education Policy and Data Center (EPDC)	http://epdc.org/
IND	India	Education Policy and Data Center (EPDC)	http://epdc.org/
IRL	Ireland	Eurostat	http://epp.eurostat.ec.europa.eu/NavTree_prod/everybody/BulkDownloadListing
IRN	Iran	NA	
ISR	Israel	Integrated Public Use Microdata Series International (IPUMS)	https://international.ipums.org/international/
ITA	Italy	Eurostat	http://epp.eurostat.ec.europa.eu/NavTree_prod/everybody/BulkDownloadListing
JOR	Jordan	Education Policy and Data Center (EPDC)	http://epdc.org/
JPN	Japan	National Statistics Office	http://www.e-stat.go.jp/SG1/chiiki/ToukeiDataSelectDispatchAction.do
KAZ	Kazakhstan	Education Policy and Data Center (EPDC)	http://epdc.org/
KEN	Kenya	Education Policy and Data Center (EPDC)	http://epdc.org/
KGZ	Kyrgyz Republic	Education Policy and Data Center (EPDC)	http://epdc.org/
KHM	Cambodia	Education Policy and Data Center (EPDC)	http://epdc.org/
KOR	Korea, Rep.	NA	
LAO	Lao PDR	Education Policy and Data Center (EPDC)	http://epdc.org/
LBN	Lebanon	Ministry of Social Affairs	http://www.cas.gov.lb/images/PDFs/Educational%20status-2004.pdf
LKA	Sri Lanka	Education Policy and Data Center (EPDC)	http://epdc.org/

Code	Country	Source	Available Link
LSO	Lesotho	Education Policy and Data Center (EPDC)	http://epdc.org/
LTU	Lithuania	National Statistics Office	http://db.stat.gov.lt/sips/Dialog/varval.asp?ma=gs_dem17en&ti=Population+by+educational+attainment+and+age+group%A0%28aged+10+years+and+over%29&path=../Database/cen_en/p71en/demography/(=2
LVA	Latvia	National Statistics Office	http://data.csb.gov.lv/Dialog/varval.asp?ma=tsk03a&ti=EDUCATIONAL+ATTAINMENT+OF+POPULATION&path=../DATABASEEN/tautassk/Results%20of%20Population%20Census%202000%20in%20brief/(=1
MAR	Morocco	Education Policy and Data Center (EPDC)	http://epdc.org/
MDA	Moldova	Education Policy and Data Center (EPDC)	http://epdc.org/
MDG	Madagascar	Education Policy and Data Center (EPDC)	http://epdc.org/
MEX	Mexico	Education Policy and Data Center (EPDC)	http://epdc.org/
MKD	Macedonia, FYR	Education Policy and Data Center (EPDC)	http://epdc.org/
MNG	Mongolia	Integrated Public Use Microdata Series International (IPUMS)	https://international.ipums.org/international/
MOZ	Mozambique	Education Policy and Data Center (EPDC)	http://epdc.org/
MWI	Malawi	Education Policy and Data Center (EPDC)	http://epdc.org/
MYS	Malaysia	Integrated Public Use Microdata Series International (IPUMS)	https://international.ipums.org/international/
NAM	Namibia	Education Policy and Data Center (EPDC)	http://epdc.org/
NER	Niger	Education Policy and Data Center (EPDC)	http://epdc.org/
NGA	Nigeria	Education Policy and Data Center (EPDC)	http://epdc.org/
NIC	Nicaragua	Education Policy and Data Center (EPDC)	http://epdc.org/
NLD	Netherlands	Eurostat	http://epp.eurostat.ec.europa.eu/NavTree_prod/everybody/BulkDownloadListing
NOR	Norway	National Statistics Office	http://statbank.ssb.no/statistikkbanken/Default_FR.asp?PXSid=0&nvl=true&PLanguage=1&tilside=selecttable/hovedtabellHjem.asp&KortnavnWeb=utniv
NPL	Nepal	Education Policy and Data Center (EPDC)	http://epdc.org/
NZL	New Zealand	National Statistics Office	http://wdmzpub01.stats.govt.nz/wds/ReportFolders/reportFolders.aspx
PAK	Pakistan	Education Policy and Data Center (EPDC)	http://epdc.org/
PAN	Panama	Education Policy and Data Center (EPDC)	http://epdc.org/
PER	Peru	Education Policy and Data Center (EPDC)	http://epdc.org/
PHL	Philippines	Education Policy and Data Center (EPDC)	http://epdc.org/
POL	Poland	Eurostat	http://epp.eurostat.ec.europa.eu/NavTree_prod/everybody/BulkDownloadListing
PRT	Portugal	Eurostat	http://epp.eurostat.ec.europa.eu/NavTree_prod/everybody/BulkDownloadListing
PRY	Paraguay	National Statistics Office	http://celade.cepal.org/cgibin/RpWebEngine.exe/EasyCross?&BASE=CPVPRY2002&ITEM=INDICADO&MAIN=WebServerMain.inl
ROM	Romania	Eurostat	http://epp.eurostat.ec.europa.eu/NavTree_prod/everybody/BulkDownloadListing
RUS	Russian Federation	National Statistics Office	http://74.125.65.132/translate_c?hl=en&ie=UTF-8&sl=ru&tl=en&u=http://www.perepis2002.ru/index.html%3Fid%3D15&prev=_t&usg=ALkJrhiZr6thPp3doxH9mXdDZgf-DA1fyw
SEN	Senegal	Education Policy and Data Center (EPDC)	http://epdc.org/
SLV	El Salvador	VI Censo de la Poblacion y V de Vivienda 2007	http://www.digestyc.gob.sv/cgibin/RpWebEngine.exe/Crosstabs
SRB	Serbia	National Statistics Office	http://webrzs.statserb.sr.gov.yu/axd/en/Zip/CensusBook4.zip
SVK	Slovak Republic	National Statistics Office	http://px-web.statistics.sk/PXWebSlovak/DATABASE/En/02EmploMarket/01EconPopActiv/EA_total.px
SVN	Slovenia	National Statistics Office	http://www.stat.si/pxweb/Database/Census2002/Administrative%20units/Population/Education/Education.asp
SWE	Sweden	National Statistics Office	http://www.ssd.scb.se/databaser/makro/SubTable.asp?yp=tansss&xu=C9233001&omradekod=UF&huvudtabell=Utbildning&omradetext=Education+and+research&tabelltext=Population+16-74+years+of+age+by+highest+level+of+education,+age+and+sex.+Year&preskat=O&prodid=UF0506&starttid=1985&stopptid=2007&Fromwhere=M(=2&langdh=2
SWZ	Swaziland	Education Policy and Data Center (EPDC)	http://epdc.org/
SYR	Syrian Arab Republic	Education Policy and Data Center (EPDC)	http://epdc.org/
THA	Thailand	Education Policy and Data Center (EPDC)	http://epdc.org/
TUR	Turkey	National Statistics Office	http://www.tuik.gov.tr/isgucueng/Kurumsal.do
TZA	Tanzania	Education Policy and Data Center (EPDC)	http://epdc.org/
UGA	Uganda	Education Policy and Data Center (EPDC)	http://epdc.org/
UKR	Ukraine	National Statistics Office	http://stat6.stat.lviv.ua/PXWEB2007/Database/POPULATION/1/06/06.asp
URY	Uruguay	National Statistics Office	http://www.ine.gub.uy/microdatos/engih2006/persona.zip,
USA	United States	National Statistics Office	http://factfinder.census.gov/servlet/DatasetMainPageServlet?_program=ACS&_submenuId=&_lang=en&_ts=
UZB	Uzbekistan	Education Policy and Data Center (EPDC)	http://epdc.org/
VEN	Venezuela	Integrated Public Use Microdata Series International (IPUMS)	https://international.ipums.org/international/
VNM	Vietnam	Education Policy and Data Center (EPDC)	http://epdc.org/
ZAF	South Africa	National Statistics Office	http://www.statssa.gov.za/timeseriesdata/pxweb2006/Dialog/varval.asp?ma=Highest%20level%20of%20education%20grouped%20by%20province&ti=Table:+Census+2001+by+province,+highest+level+of+education+grouped,++population+group+and+gender.+&path=../Database/South%20Africa/Population%20Census/Census%202001%20-%20NEW%20Demarcation%20boundaries%20as%20at%209%20December%202005/Provincial%20level%20-%20Persons/=1
ZAR	Congo, Dem. Rep.	Education Policy and Data Center (EPDC)	http://epdc.org/
ZMB	Zambia	Education Policy and Data Center (EPDC)	http://epdc.org/
ZWE	Zimbabwe	Education Policy and Data Center (EPDC)	http://epdc.org/

APPENDIX OF DATA SOURCES: ECONOMIC CENSUS DATA

Country	Code	Year	Source	Big Firms	Industry	Calculations	Links
Albania	ALB	2009	Data is from the Albanian Institute of Statistics' Economic Indicators, Business Register, table titled: Active Enterprises by Counties and Size.	50 Employees	non-agriculture	Data includes the number of establishments in four size categories. Data for the number of employees in each size category is estimated using an assumption of 2.5 employees for establishments with 1 to 4 employees, 7 for establishments with 5 to 9, 29.5 for establishements with 10 to 49 and 50 employees for establishments with over 50 employees. Data excludes establishments with no employees.	http://www.instat.gov.al/
Argentina	ARG	2003	Data is from the Instituto Nacional de Estadística y Censos, Censo Nacional Económico, Industria Manufacturera, table 2.	na	non-agriculture	Data includes the number of establishments and employees by three size categories. The number of employees are approximated here by the use of Jobs Held data.	http://www.indec.mecon.ar/
Armenia	ARM	2008	Data is from the National Statistical Service of the Republic of Armenia, Main Statistical Indicators 2005-2008	na	all sectors	No size data was found available for establishments. For employee data we use figures from the labor force survey.	http://www.armstat.am/file/RegStat/default-eng.html
Australia	AUS	2006	Data is from the Australian Bureau of Statistics, statistical catalogue number 81. Industry wide statistics, table 8165.0 Counts of Australian Businesses, including Entries and Exits, Jun 2003 to Jun 2007, Businesses by Industry Class by Main State by Employment Size Ranges - 2006-07.	200 Employees	non-agriculture (i.e. excl. 0111-0219)	Data includes the number of establishments by four size categories. Data for the number of employees in each size category is estimated using an assumption of 10 employees for establishments with 1 to 19 employees, 109.5 for establishments with 20 to 199, and 200 employees for establishments with over 200 employees.	http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/8165.0Jun%202003%20to%20Jun%202007?OpenDocument
Bangladesh	BGD	2001 & 2003	Data is from Bangladesh's Economic Census of 2001 & 2003.	na	non-agriculture and manufacturing	Data includes the total number of establishments and employees.	http://discovery.lib.harvard.edu/?itemid=|library/m/aleph|011346898
Benin	BEN	1980 & 2002	Data is from the National Institute of Statistics and Economic Analysis, Recensement des entreprises, Les Entreprises Artisanales 1980. And the Troisieme Recensement General de la Population et de L'Habitation 2002.	na	establishments data excludes the agriculture sector, employee data includes all sectors	Data includes the total number of establishments from 1980, and total number of employees from 2002. Because the regions of Benin were split in 1999, data from 1980 is disaggregated (establishments for each pre-1999 region are allocated to the two new 1999 sub regions pro rata based on formal sector employees).	http://www.insae-bj.org/
Bosnia	BIH	2009	Data is from the Federal Bureau Of Statistics, Cantons in numbers.	na	non-agriculture and manufacturing	Data includes the total number of establishments and employees.	http://www.bhas.ba/new/
Brazil	BRA	2006	Data from the Brazilian Central Registry of Businesses.	na	manufacturing [Indústrias de transformação]	Data includes the total number of establishments and employees.	http://www.ibge.gov.br/english/default.php
Burkina Faso	BFA	1998	Data is from the National Institute of Statistics and Demography, Recensement Industriel & Commercial.	na	excludes agriculture and hunting	Data includes the total number of establishments.	http://www.insd.bf/fr/
Cambodia	KHM	2009	Data is from the Preliminary Results of the Nation-wide Establishment Listing of Cambodia 2009.	na	all sectors	Data includes the total number of establishments.	http://www.nis.gov.kh/
Cameroon	CMR	1986	Data is from the Institut National de la Statistique, Recensement Industriel de 1986.	na	maufacturing	Data includes the total number of establishments and employees.	http://www.statistics-cameroon.org/
Canada	CAN	2010	Data is from Statistics Canada, CANSIM, table 281-0044 and Catalogue no. 72-002-X	100 Employees	all sectors	Data includes number of employees by seven size categories. Data for the number of establishments in each size category is estimated using an assumption of 2 employees for establishments with 0 to 4 employees, 12 for establishments with 5 to 19, 34.5 for establishments with 20 to 49 employees, 74.5 employees for establishments with 50 to 99, 199.5 for establishments with 100 to 299, 399.5 for establishments with 300 to 499 and 500 employees for establishments with over 500 employees.	http://www40.statcan.gc.ca/l01/cst01/labr83d-eng.htm
China	CHN	2004	Data is from China Online, China Economic Census Yearbook 2004, Enterprise 2, table 2-8 Number of Corporation Legal Person Units by Region and Group Interval of Empolyed Persons and 1-2 Number of Legal Person Units, Establishments and Employed Persons by Region.	100 Employees	all sectors	Data includes the number of establishments by ten size categories. The number of employees in each size category is estimated using the assumption of 3.5 employees for establishment with 0 to 7 employees, 13.5 for establishments with 8 to 19, 34.5 for establishments with 20 to 49, 74.5 for establishments with 50 to 99, 199.5 for establishments with 100 to 299, 349.5 for establishments with 300 to 499, 749.5 for establishments with 500 to 999, 1999.5 for establishments with 1000 to 2999, 3999.5 for establishments with 3000 to 4999, and 5000 employees for establishments with over 5000 employees.	http://chinadataonline.org/member/ecocensus/vbListDetail.asp?ID=1#
Colombia	COL	2001-02	Data is from the National Administrative Department of Statistics (DANE), Recensement Economique 2001-2002 Royaume du Maroc.	200 Employees	total data excludes agriculture; size data appears to include all sectors	Data includes the number of establishments and employees for ten size categories.	http://190.25.231.246:8080/Dane/tree.jsf
Costa Rica	CRI	2008	Data is from the INEC's Directory of Institutional Units and Establishments. [INEC. Directorio de Unidades Institucionales y Establecimientos.]	100 Employees	all sectors	Data includes the number of establishments and employees by five size categories.	http://www.inec.go.cr/Web/Home/pagPrincipal.aspx
Croatia	HRV	2008	Data is from the Central Bureau of Statistics, Business Entities According To Nkd 2002. By Counties at 31 December 2008, and Employed By The Activity And Counties at 31 March 2008. [Poslovni Subjekti Prema Nkd-U 2002, Po Županijama Stanje 31. Prosinca 2008. Zaposleni Po Područjima Djelatnosti Županijama Stanje 31. Ožujka 2008.]	na	non-agriculture and manufacturing	Data includes the total number of employees and establishments (excludes establishments with no employees).	http://www.dzs.hr/
Cuba	CUB	2008	Data is from the National Statistics Office of Cuba, the Provincial Statistical Yearbook 2008.	na	all sectors	Data includes the total number of business entities, coorperatives and other organisations and employees.	http://www.one.cu/
Denmark	DNK	2007	Data is from the Statistical Office of Denmark, General economic statistics, Business Demography, table GF4: General enterprise statistics by region, industry (DB07, 10-grouping) and unit	na	non-agricultural	Data includes the total number of enterprises and employees. Enterprises are used in place of establishments here.	http://www.statbank.dk/statbank5a/default.asp?w=1680
Dominican Republic	DOM	2002	Data is from the National Statistics Office of the Dominican Republic, Perfil Sociodemográfico Provincial, Censo Nacional de Población y Vivienda VIII.	na	excludes agriculture, fishing and forestry	Data includes the total number of employees.	http://www.one.gob.do/index.php?module=articles&func=view&catid=217
Ecuador	ECU	2007	Data is from the National Institute of Statistics and Censuses, the results of the Manufacturing and Mining, Table N.17.	na	manufacturing and mining	Data includes the total number of establishments and employees.	http://www.inec.gov.ec/web/guest/publicaciones/anuarios/inv_eco/man_min
Egypt	EGY	2008	Data is from the Central Agency for Public Mobilization and Statistics, Data from Egypt in Figures, table Estimates of Employed Persons (15 years old and over) By Sex, Governate.	na	all sectors	Data includes the estimated total number of employees age 15 years old and older.	http://www.msrintranet.capmas.gov.eg/ows-img2/htms/pdf/work/25.pdf
El Salvador	SLV	2005	Data is from the Ministerio de Economia Direccion General de Estadistica y Censos, VII Censos Economicos.	5 Employees	non-agriculture and manufacturing	Data includes the total number of establishments and employees.	http://www.minec.gob.sv/
Estonia	EST	2009	Data is from the National Statistical Office of Mongolia, Statistical database: Economy, Entities, Business demography, ER28: Affiliates Statistical Profile By County And Number Of Employees. [Statistika andmebaas: Majandus, Majandusüksused, Ettevõtluse demograafia, ER28: Statistilisse Profiili Kuuluvad Ettevõtted Töötajate Arvu Ja Maakonna Järgi.]	250 Employees	all sectors	Data includes the number of establishments by four size categories. Data for the number of employees in each size category is estimated using an assumption of 5 employees for establishments with less than 10 employees, 30 for establishments with 10 to 49, 150 for establishments with 50 to 249, and 250 employees for establishments with over 250 employees.	http://pub.stat.ee/px-web.2001/Dialog/varval.asp?ma=ER28&ti=STATISTILISSE+PROFIILI+KUULUVAD+ETTEV%D5TTED+T%D6%D6TAJATE+ARVU+JA+MAAKONNA+J%C4RGI+&path=../Database/Majandus_regionaalne/10Majandusuksused/02Ettevetjad/&lang=2
Georgia	GEO	2010	Data is from the National Statistics Office of Georgia, special request.	100 Employees	non-agriculture	Data includes the number of establishments and employees above and below the 100 employee size threshold.	http://www.geostat.ge/
Ghana	GHA	2003-05	Data is from the Ghana Statistical Service's 2003 National Industrial Census, Phase I and II, Table 10: Establishments by Size and Region and Table 11: Persons Engaged by Size and Region.	100 Employees	industrial sectors	Data includes the number of establishments and employees by nine size categories.	http://www.statsghana.gov.gh/Industrial_Census.html

Country	Code	Year	Source	Big Firms	Industry	Calculations	Links
India	IND	2005	Data is from the Government of India Ministry of Statistics and Programme Implementation, Economic Census 2005, tables: Tables A 3.4, A 4.4, A 3.6, A-3.22, A 4.6, and A 4.22.	na	non-agriculture and manufacturing	Data includes the total number of establishments in non-agriculture and manufacturing sectors and employees in non-agriculture employees.	http://mospi.nic.in/Mospi_New/upload/economic_census_2005/index_6june08.htm
Iran	IRN	2007	Data is from the Statistical Centre of Iran, Iran Manufacturing by Ostan 2006 and 2007, table 7.5. Manufacturing Establishments By Ostan And Type Of Ownership: 1384 and 7.8. Number Of Workers In Manufacturing Establishments By Ostan And Size: 1384.	100 Employees	manufacturing	Data includes the number of establishments and employees by three size categories.	http://www.amar.org.ir/Default.aspx
Israel	ISL	2008	Data is from the Central Bureau of Statistics, Israel in Figures 2008.	na	all sectors	Data includes the total number of establishments and employees.	http://www1.cbs.gov.il/reader/cw_usr_view_Folder?ID=141
Italy	ITA	2008	Data is from the National Statistical Institute, 8th General Census of Industry and Services October 2001, table 8.	250 Employees	non-agriculture and manufacturing	Data includes the total number of establishments and employees in the manufacturing sector and non-agriculture sectors. The data also includes non-agriculture employees by four size categories.	http://dwcis.istat.it/cis/index.htm
Japan	JPN	2006	Data is from the online Portal Site, Official Statistics of Japan, the Establishment and Enterprise Census of Japan, Results of Establishments for Japan, table 6: Establishments and Employees by Sex; by Industry (Major Groups) and Size of Employees (10 Groups); for Japan, Prefectures, 16 Major Cities, 14 Major Metropolitan Areas and Special Areas.	100 Employees	manufacturing	Data includes the number of establishments and employees by ten size categories (exludes establishments with no employees).	http://www.e-stat.go.jp/SG1/estat/ListE.do?bid=000001008300&cycode=0
Jordan	JOR	1998 & 2006	Data is from the Department of Statistics (DOS), Employment Survey 1998 and the Establishment Census 2006	na	non-agricultural	Data includes the total number of establishments and employees.	http://www.dos.gov.jo/dos_home_e/main/index.htm
Kazakhstan	KAZ	2010	Data is from the Agency of Statistics Regions of Kazakhstan Brochure.	100 Employees	all sectors	Data includes the number of establishements by three size categories. The number of employees in each size category are calculated with the assumption of 25 employees for small establishments, 75 employees for medium establishments and 100 employees for large establishments.	http://www.stat.kz/Pages/default.aspx
Kyrgyz Republic	KGZ	2008	Data is from the National Statistical Committee of Kyrgyz Republic, Key indicators (Statistical Yearbook Kyrgyz Republic 2004-2009), "Number of small and medium-sized enterprises on the territory" and "The number of workers in small and medium-sized enterprises on the territory".	na	all sectors	Data includes the total number of establishments and employees for small and medium size businesses with foreign investment.	http://www.stat.kg/rus/part/msp.htm
Lao PDR	LAO	2004	Data is from the Ministry of Industry-Handicraft	99 Employees	all sectors	Data includes the number of establishments by three size categories (establishments with no employees are excluded). The number of employees in each size category is estimated using an assumption of 5 employees for establishment with 1 to 10 employees, 55 for establishments 10 to 99, and 100 employees for establishments with more than 99 employees.	http://www.nsc.gov.la/Statistics/Selected%20Statistics/Industry.htm
Latvia	LVA	2003	Data is from the Central Statistical Bureau "Statistikas uzņēmumu reģistrs, Tabula: SR09. EKONOMISKI AKTĪVIE UZŅĒMUMI UN UZŅĒMĒJSABIEDRĪBAS, STATISTISKAJOS REĢIONOS, PILSĒTĀS UN RAJONOS, 1997. - 2003.g." [Statistical Business Register, Table SR09. Economically Active Enterprises And Business Companies, Statistical Regions, Cities And Districts, 1997. - 2003]	250 Employees	non-agriculture	Data includes the number of establishments in three size categories. The number of employees is estimated using the assumption of 25 employees for establishments with 0 to 49 employees, 149.5 for establishments with 50 to 249, and 250 employees for establishments with over 250 employees.	http://data.csb.gov.lv/DATABASE/rupnbuvn/Ikgad%C4%93jie%20statistikas%20dati/Statistikas%20vien%C4%ABbu%20re%C4%A3istrs/Statistikas%20vien%C4%ABbu%20re%C4%A3istrs.asp
Lebanon	LBN	2004	Data is from the Lebanese Census of Buildings Dwellings and Establishments 2004.	100 Employees	non-agriculture and all sectors	Data includes the number of establishments from all sectors by six size categories. The number of employees with over 100 employees is estimated with the assumption of 100 employees. The total number of establishments and employees exclude the agriculture sector.	http://www.cas.gov.lb/index.php?option=com_content&view=frontpage&Itemid=28
Lithuania	LTU	2010	Data is from the Statistics Lithuania, Database of Indicators, Business statistics, Small and medium sized enterprises in operation, Table M4010241: Number of enterprises in operation and number of employees at the beginning of the year by administrative territory, size class of enterprises	100 Employees	excluding agriculture, forestry and fishing; and all sectors	Data includes the number of establishments and employees by ten size categories for all sectors. The total number of employees and establishments exclude agriculture, forestry and fishing sectors.	http://db1.stat.gov.lt/statbank/SelectTable/omrade0.asp?Subjectcode=S4&PLanguage=1&Shownews=OFF&tree=false
Macedonia, FYR	MKD	2009	Data is from the State Statistical Office of the Republic of Macedonia, Statistical Databases, Regional Statistics, Business Entities.	50 Employees	all sectors	Data includes the number of establishments by six size categories. The number of employees in each size category is estimated using the assumption of 5 employees for establishments with 1 to 9 employees, 24.5 for establishments with 10 to 49, 149.5 for establishments with 50 to 249 and 250 employees for establishements with over 250 employees.	http://www.stat.gov.mk/PXWeb2007bazi/Database/Regional%20Statistics/databasetree.asp
Madagascar	MDG	2003	Data is from INSTAT, Recensement au niveau des Communes.	na	non-agricultural	Data includes the number of establishments.	http://www.instat.mg/pdf/rgph_6.pdf
Malaysia	MYS	2005	Data is from the Department of Statistics, Economic Census of the manufacturing sector.	na	manufacturing	Data includes the number of total establishments and Engaged Persons. Here the number of Engaged Persons is used as a proxy for employees.	http://www.statistics.gov.my/portal/index.php?lang=en
Mexico	MEX	2009	Data is from the National Institute of Statistics, Geography, and Informatics, Censos Económicos 2009.	na	excludes agriculture, livestock and forestry	Data includes the number of establishments and employees by twelve size categories.	http://www.inegi.org.mx/est/contenidos/espanol/proyectos/censos/ce2009/default.asp?s=est&c=14220
Moldova	MDA	2008	Data is from the National Bureau of Statistics, Territorial statistics, Principalii Indicatori Ai Activităţii Întreprinderilor Industriale, În Profil Teritorial.	na	manufacturing	Data includes the total number of employees.	http://www.statistica.md/pageview.php?l=en&idc=349&
Mongolia	MNG	2007	Data is from the Businesss Register of Mongolia table 11.3 Number of Active Legal Units, by Aimags and the Capital, Employment Size Class,	50 Employees	all sectors	Data includes the number of establishments by four size categories. The number of employees for each size category is estimated using the assumption of 5 employees for establishments with 1 to 9 employees, 14.5 for establishments with 10 to 19, 34.5 for establishments with 20 to 49, and 50 employees for establishments with over 50 employees.	http://www.statis.mn/v3/index2.php?page=free_access
Morocco	MAR	2001-02	Data is from the Department of Statistics, Economic Census, TABLEAU 1 : Répartition des établissements et de l'effectif d'emploi selon les régions et les secteurs d'activités	na	non-agriculture and manufacturing	Data includes the total number of establishments and employees in the manufacturing sector and non-agriculture sectors.	http://www.recensement-eco.hcp.ma/article.php3?id_article=11
Mozambique	MOZ	2004	Data is from the National Institute of Statistics, CEMPRE, 2004, table 5, 14 and 21. African Development Bank & African Development Fund, Mozambique Private Sector Country Profile August 2008, table 5.	100 Employees	all sectors	Data includes the total number of establishments by two size categories. The number of employees in each size category is estimated using a national total and regional averages for each size category.	http://www.ine.gov.mz/
Nepal	NPL	2006-07	Data is from the Central Bureau of Statistics, Census of Manufacturing Establishments 2006/2007.	100 Employees	maufacturing	Data includes the number of establishments and employees by six size categories. Data excludes establishments with fewer than 10 employees.	http://www.cbs.gov.np/
New Zealand	NZL	2002 [1997 update]	Data is from Statistics New Zealand, New Zealand Business Demographic Statistics, Table 2.	100 Employees	non-agriculture	Data includes the number of Enterprises and FTE Engaged Persons by five size categories. Enterprises are used as a proxy for establishments and FTE Engaged Persons are used as a proxy for employees. Data only reflects Enterprises that meet certain significance tests including minimum revenue or two or more employees. Data has been adjusted to estimate 2002, and was originally collected from a 1997 survey.	http://www.stats.govt.nz/browse_for_stats/businesses/business_characteristics/nz-business-demography-stats-std-tables.aspx
Norway	NOR	2010	Data is from Statistics Norway, table 2 Establishments, by size groups and county.	100 Employees	all sectors	Data includes the number of establishments by eight size categories. The number of employees in each size category are estimated using an assumption of 2.5 employees for establishments with 1 to 4 employees, 7 for establishments with 5 to 9, 14.5 for establishments with 10 to 19, 34.5 for establishments with 20 to 49, 74.5 for establishments with 50 to 99, 174.5 for establishments with 100 to 249, and 250 employees for establishments with over 250 employees.	http://www.ssb.no/english/subjects/10/01/bedrifter_en/tab-2010-01-29-02-en.html

Country	Code	Year	Source	Big Firms	Industry	Calculations	Links
Pakistan	PAK	2005	Data is from the Statistics Division, Federal Bureau of Statistics Division, Economic Census 2005.	100 Employees	excludes agriculture, forestry, hunting and fishing	Data includes the number of establishments by nine size categories. The number of employees in each size category is estimated using the assumption of 125.5 employees for establishments with 101 to 150 employees, 175.5 for establishments with 151 to 200, 225.5 for establishments with 201 to 250, 375.5 for establishments with 251 to 300 and 300 employees for establishments with over 300 employees.	http://www.statpak.gov.pk/depts/fbs/publications/ec_2005/ec_2005.html
Panama	PAN	2002	Data is from the Census and Statistics Directorate, V Censos Nacionales Economicos.	na	non-agriculture	Data includes the total number establishements and employees.	http://www.contraloria.gob.pa/
Paraguay	PRY	2002	Data is from the General Directorate of Statistics, Surveys and Censuses, Resultados Preliminares de la Encuesta Industrial.	na	manufacturing	Data includes total number of establishments.	http://www.dgeec.gov.py/
Peru	PER	1993-94	Data is from the National Institute of Statistics and Informatics, III Censo Nacional Económico 1993-1994 (CENEC), Peru: Numero De Establecimientos Censados E Informantes, Por Estrato De Personal Ocupado; Segun Departamento (Composicion Porcentual).	20 Employees	non-agriculture and manufacturing	Data includes number of establishments by three size categories, for non-agricultural sectors and appears to exlcude establishments with no employees. The number of establishements with 20 or more employees is estimated using the assumption of 20 employees. Data also includes total establishments and employees for the non-agriculture and manufacturing sectors.	http://www.inei.gob.pe/
Philippines	PHL	2006	Data is from the 2006 Census of Philippine Business and Industry, by special request from the National Statistics Office, Republic of the Philippines.	20 Employees	non-agricultural	Data includes the number of establishments and employees by two size categories.	http://www.census.gov.ph/data/sectordata/databusind.html
Russia	RUS	2008	Data is from the Russian State Committee for Statistics, Socio-Economic Indicators 2009.	na	non-agriculture and manufacturing	Data includes establishment and employee totals for non-agriculture sectors and the manufacturing sector.	http://www.gks.ru/eng/
Senegal	SEN	2005	Data is from the National Agency of Statistics and Demography, Service Regional de la Statistique et de la Demographie.	na	non-agricultural	Data includes the total number of establishments and employees.	http://www.ansd.sn/
Serbia	SRB	2007	Data is from the Statistical Office of the Republic of Serbia, Basic Results of Business Activity of Enterprises and Entrepreneurs.	na	non-agriculture and manufacturing	Data includes the total number of establishments and employees.	http://webrzs.stat.gov.rs/axd/en/
Slovakia	SVK	2009	Data is from the Statistical Office of the Slovak Republic.	na	manufacturing	Data includes the total number of establishments and employees.	http://px-web.statistics.sk/PXWebSlovak/index_en.htm
South Africa	SFA	2007	Data is from Statistics South Africa, Labour Force Survey, table Labour market indicators (working-age population, 15-64 years) by province.	na	all sectors	Data inclues the total number of employees from a Labor Force survey.	http://www.statssa.gov.za/publications/P0210/P0210September2000,2001,2002,2003,2004,2005,2006,2007.pdf
South Korea	KOR	2004	Data is from Statistics Korea, Report of the Census on Establishments, table By province, industrial classification and type of legal organization.	na	non-agricultural	Data includes the total number of establishments and employees.	http://kostat.go.kr/nso_main/nsoMainAction.do?method=sub&catgrp=eng2009&catid1=g03&catid2=g03a&catid3=g03ac&catid4=g03acq
Sri Lanka	LKA	2003	Data is from the Department of Census and Statistics - Sri Lanka, Census of Industry 2003/2004, Table A : No.of Manufacturing Establishments and Persons Engaged by District and Type of Industry Scale.	10 Employees	manufacturing	Data includes the number of establishments and employees by two size categories.	http://www.statistics.gov.lk/industry/census%20of%20industries_2004.pdf
Switzerland	CHE	2005	Data from the The Portal Statistics Switzerland by the Federal Statistical Office (FSO), Recensement des entreprises, Structure économique, table 7. Résultats pour les grandes régions et les cantons, Etablissements and Emplois.	na	non-agricultural	Data includes the total number of establishments and employees.	http://www.bfs.admin.ch/bfs/portal/fr/index/infothek/publ.html
Syria	SYR	2007	Data is from the Central Bureau of Statistics, workers and industry statistical tables.	na	all sectors	Data includes the total number of establishments and employees.	http://www.cbssyr.org/work/2009/semi-1/TAB13.htm
Tanzania	TZA	2007	Data is from National Bureau of Statistics, Central Register of Establishments, Business Survey Tanzania Mainland Report.	100 Employees	excludes agriculture, hunting and forestry	Data includes the number of establishments and employees by eight size categories.	http://www.nbs.go.tz/index.php?option=com_phocadownload&view=category&id=72:industry&Itemid=106
Thailand	THA	2007	Data is from the National Statistical Office 2007 Industrial Census.	200 Employees	manufacturing	Data includes the number of establishments numbers by six size categories. The number of employees in each size category is estimated with the assumtion of 8 employees for establishments with 1 to 15 employees, 20.5 for establishments with 16 to 25, 28 for establishments with 26 to 30, 40.5 for establishments with 31 to 50, 125.5 for establishments with 51 to 200, and 200 employees for establishments with over 200 employees. The total number of employees are actual data and not estimated. Data excludes establishments with no employees. Establishments with 10 and fewer employees were sampled, establishments with over 11 employees were surveyed.	
Turkey	TUR	1992	Data is from Turkey's Business Statistics, General Census of Industry and Business Establishments, 2002 General Census Of Industry And Establishments, 1.1.2. Number of establishments, annual average number of persons engaged by provinces and economic activity branches.	na	non-agricultural	Data includes the total number of establishments and Persons Engaged. Persons Engaged is used as a proxy for employees.	http://www.turkstat.gov.tr/VeriBilgi.do?tb_id=28&ust_id=9
Uganda	UGA	2006	Data is from the Uganda Bureau of Statistics, Report on the Uganda Business Register 2006/7, Table 3.4.4 Distribution of Businesses by Region by Employment Size band.	100 Employees	all sectors	Data includes the number of establishments by six size categories and excludes establishments with no employees. The number of employees in each size category is estimated with the assumption of 2.5 employees for establishments with 1 to 4 employees, 7 for establishements with 5 to 9, 14.5 for establishments with 10 to 19, 34.5 for establishments with 20 to 49, 74.5 for establishments with 50 to 99, and 100 employees for establishements with over 100 employees.	
Ukraine	UKR	2008	Data is from the State Statistics Committee Of Ukraine, Activity of Enterprises: Statistical Publication 2008, Table 8.7 Number of employees the size of the enterprise, by regions, Table 8.3 Number of enterprises per 10000 of present population the size of the enterprise, by regions, and Table 8.2 Share of large, medium-sized and small enterprises, by regions.	50 Employees	all industries	Data includes the number of employees in three size categories. The number of establishments in each region and size size category were implied from the number of enterprises per 10000 of present population (Table 8.3), the 2008 regional population (Statistical publication Regions of Ukraine 2009, Table 3.1) and the percentage share of firms by size and by region (Table 8.2).	http://www.ukrstat.gov.ua/
United Arab Emirates	ARE	1995	Data from the United Arab Emirates Statistical Abstract, Chapter 11 Economical Establishments.	na	all sectors, and excluding agriculture forestry and hunting	Data includes the number of establishments and employees by ten size categories in all sectors. The total number of establishments and employees excludes agriculture, forestry and hunting sectors.	http://www.economy.ae/English/EconomicAndStatisticReports/StatisticReports/StatisticAbstract/Pages/sa2007.aspx
United States	USA	2006	Data is from the U.S. Census Bureau, Statistics of U.S. Businesses.	100 Employees	all sectors and non-agriculture	Data includes the total number of establishments and employees by twelve size categories for all sectors. The total number of establishments and employees exclude agriculture.	http://www.census.gov/epcd/www/smallbus.html
Vietnam	VNM	2007	Data is from the General Statistics Office of Vietnam, Results of the 2007 Establishment Census, Table 39 Number of individual business establishments by size of employee, by size of employee.	101 Employees	non-agricultural	Data includes the number of establishments by nine size categories. Number of employees in each size category is estimated with the assumption of 3.5 employees for establishments with 2 to 5 employees, 8 for establishments with 6 to 10, 15.5 for establishments with 11 to 20, 35.5 for establishments with 21 to 50, 75.5 for establishments with 51 to 100, 150.5 for establishments with 101 to 200, 350.5 for establishments with 201 to 500, and 500 for establishments with over 500 employees.	http://www.gso.gov.vn/default_en.aspx?tabid=515&idmid=5&ItemID=9359