Human Capital and Regional Development

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

Revised, April 2012[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 97 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 productivity 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 entrepreneurial inputs and possibly human capital externalities help understand 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 5 years. We thank Gary Becker, Nicholas Bloom, Vasco Carvalho, Edward Glaeser, Gita Gopinath, Josh Gottlieb, Elhanan Helpman, Chang-Tai Hsieh, Matthew Kahn, Pete Klenow, Robert Lucas, Casey Mulligan, Elias Papaioannou, Jacopo Ponticelli, Giacomo Ponzetto, Jesse Shapiro, Chad Syverson, David Weil, seminar participants at the UCLA Anderson School, Harvard University, University of Chicago, and NBER, as well as the editors and referees of this journal for extremely helpful comments. Gennaioli thanks the Barcelona Graduate School of Economics and the European Research Council for financial support. Shleifer thanks the Kauffman Foundation for support.

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 97 percent of its GDP. We explore the influences of geography, natural resource endowments, institutions, human capital, and culture by looking within countries. We combine this analysis with an examination of productivity in several thousand establishments covered by the World Bank Enterprise Survey, for which we have both establishment-specific and regional data. In this analysis, human capital measured using education emerges as the most consistently important determinant of both regional income and productivity of regional establishments. We then use the combination of regional and establishment-level data 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 model 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 human capital of the entrepreneur/manager influences firm-level productivity independently. The distinction between entrepreneurs/managers and workers has been shown empirically to be critical in accounting for productivity and size of firms in developing countries (Bloom and Van Reenen 2007, 2010; La Porta and Shleifer 2008; Syverson 2011). 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) and Romer (1990). 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]. 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, we need to consider the mobility of firms, workers, and entrepreneurs across regions, which is presumably less expensive than that across countries. Our model follows the standard urban economics approach (e.g., Roback 1982, Glaeser and Gottlieb 2009) of labor mobility across regions with land and housing limiting universal migration into the most productive regions. This formulation allows us to analyze the conditions under which the regional equilibrium is stable and to consider jointly the education coefficients in regional and establishment level regressions.

To begin, we examine the determinants of regional income in a specification with country fixed effects. Our 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. We also consider institutions, which have been found by King and Levine (1993), De Long and Shleifer (1993), Hall and Jones (1999), and Acemoglu et al. (2001) to be significant determinants of development. We also look at culture, measured by trust (Knack and Keefer 1997), and at ethnic heterogeneity (Easterly and Levine 1997, Alesina et al. 2003). Last, we look at average education in the region. 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), Breton (2012), and Cohen and Soto (2007) are recent confirmations.

[2] We do not consider the role of human capital in shaping technology adoption (Nelson and Phelps 1966). For recent models of these effects, see Benhabib and Spiegel (1994), Klenow and Rodriguez-Clare (2005), and Caselli and Coleman (2006). For evidence, see Coe and Helpman (1995), Ciccone and Papaioannou (2009), Wolff (2011).

Across countries, the effects of education and institutions are difficult to disentangle: both variables are endogenous and the potential instruments for them are correlated (Glaeser et al 2004). By using country fixed effects, we avoid identification problems caused by unobserved country-specific factors.

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 Enterprise Surveys help account for cross-regional differences within a country. Some institutions or culture may matter only at the national level, but then large income differences within countries call for explanations other than culture and institutions. In contrast, differences in educational attainment account for a large share of the regional income differences within a country. The within country R^2 in the univariate regression of the log of per capita income on the log of education is about 25 percent; this R^2 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 similar to the Mankiw et al. (1992) estimate for a cross-section of countries. We confirm a 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 have regional data on several aspects of institutional quality, but find that their ability to explain cross-regional differences is minimal[3].

In regional regressions, human capital in a region may be endogenous because of migration. To make progress, we examine the determinants of firm-level productivity. We merge our data with World Bank Enterprise Surveys, which provide establishment-level information on sales, labor force,

[3] Recent work argues that regions within countries that were treated particularly badly by colonizers have poor institutions and lower income today (Banerjee and Iyer 2005, Dell 2010, Michalopoulos and Papaioannou 2011).

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. We estimate the production function predicted by our model using several methods, including Levinsohn-Petrin's (2003) panel approach. The micro data point to a large role of managerial/entrepreneurial human capital in raising firm productivity. We also find that regional education has a large positive coefficient, consistent with sizeable human capital externalities. However, because regional education may be correlated with unobserved region-specific productivity ~~parameters~~, we do not have perfect identification of externalities.

To assess the extent to which firm-level results can account for the role of human capital across regions, we combine estimation with calibration following Caselli (2005). We rely on previous research regarding factor shares (e.g., Gollin 2002, Caselli and Feyrer 2007, Valentinyi and Herrendorf 2008), but then combine it with coefficient estimates from regional and firm-level regressions. Our calibrations show that worker education, entrepreneurial education, and externalities all substantially contribute to productivity. We find the role of workers' human capital to be in line with standard wage regressions, which are the benchmark adopted by conventional calibration studies (e.g., Caselli 2005). Crucially, however, our results indicate that focusing on worker education alone substantially underestimates both private and social returns to education. Private returns are very high but to a substantial extent earned by entrepreneurs, and hence might appear as profits rather than wages, consistent with Lucas (1978). Although we have less confidence in the findings for externalities, our best estimates suggest that those are also sizeable. In sum, the evidence points to a large influence of entrepreneurial human capital, and perhaps of human capital externalities, on productivity.

In section II, we present a model of regional development that organizes the evidence. In section III, we describe our data. Section IV examines the determinants of both national and regional development. Section V presents firm-level evidence and section VI calibrates the model to assess its ability to explain income differences. Section VII 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 $\tilde{A}_P$ and a share 1– *p* of which has productivity $\tilde{A}_U < \tilde{A}_P$. We refer to the former regions as "productive", to the latter regions as "unproductive", and denote them by *i* = *P*, *U*. 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 owns 1 unit of housing, T units of land, K units of physical capital (and no human capital). Each labourer is endowed with $h \in \mathrm{R}_{++}$ units of human capital. In region *i* = *P*, *U* the distribution of human capital is Pareto in $[\underline{h},+\infty)$, where $\underline{h}$>1, with mean $\underline{H}_i$ denoting the initial, exogenous endowment of human capital in region *i* = *P*, *U.*

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 in $H_{i,h}$, $K_{i,h}$ and $T_{i,h}$. We model human capital externalities (Lucas 1988) by assuming that regional total factor productivity is given by:

$$A_i = \widetilde{A}_i \left(E_i(h)^{\psi} L_i\right)^{\gamma}, \quad \gamma > 0, \;\; \psi \geq 1. \tag{3}$$

According to (3), there are two determinants of regional productivity: i) region-specific factors $\tilde{A}_i$, which capture geography, institutions, and other influences, and ii) the region's human capital. In expression (3), $E_i(h)$ is the average level of human capital in region *i* and L_i is the measure of labour in that region. Parameter ψ captures the importance of the quality of human capital: when ψ = 1 only the total quantity of human capital $H_i = E_i(h)L_i$ matters for externalities; as ψ rises the quality of human capital becomes relatively more important than quantity. Parameter γ captures the overall importance of externalities. In our formulation, there are regional scale effects since $\gamma > 0$, which we will look for in the data, but we allow them to be arbitrarily small (when $\gamma \approx 0$). We take regional productivity A_i as given until we describe the spatial equilibrium in which A_i is endogenously determined by regional sorting of labourers.

Rentiers rent land and physical capital to firms, and housing to entrepreneurs and workers. In region *i*, each rentier earns $\lambda_i T$ and η_i by renting land and housing, where λ_i and η_i are rental rates, and $\rho_i K$ by renting physical capital. A region's land and housing endowments T and 1 are immobile; physical capital is fully mobile. Labourers use their 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]

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 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

[4] Assuming that migrants lose a fixed amount of human capital φ ensures that skilled laborers have the greatest incentive to migrate. If migrants lose a share of destination earnings, everybody has the same incentive to migrate. For simplicity, we assume that moving costs are a redistribution from migrants to locals (e.g., the latter provide moving services) and are non-rival with the time spent working. This ensures that the human capital employed in a region, as well as the aggregate income of laborers, do not depend on moving costs.

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 across regions and their choice between work and entrepreneurship within a region, we must compute regional wages w_i and profits $\pi_i(h_j)$. To do so, we first 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. We consider symmetric spatial equilibria in which all productive regions share the same factor allocation $\left(H_P^E, H_P^W, K_P\right)$, the same wage w_P and rental rates λ_P and η_P, and unproductive regions share the same allocation $\left(H_U^E, H_U^W, K_U\right)$, wage w_U, and rentals λ_U and η_U.

Throughout the analysis, the price of consumption is normalized to one. Endogenous regional differences in the rental rates of housing and land affect the welfare of labourers in different regions, but regional variation in value added does not depend on these prices in our model (precisely because value added just consists of the tradable consumption good).

Production and occupational choice

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

$$\max_{H_{i,h}, T_{i,h}, K_{i,h}} A_i h^{1-\alpha-\beta-\delta} H_{i,h}^{\alpha} K_{i,h}^{\delta} T_{i,h}^{\beta} - w_i H_{i,h} - \rho K_{i,h} - \lambda_i T_{i,h}, \tag{4}$$

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

$$H_{i,j} = \frac{h_j}{H_i^E} \cdot H_i^W, \quad K_{i,j} = \frac{h_j}{H_i^E} \cdot K_i, \quad T_{i,j} = \frac{h_j}{H_i^E} \cdot T. \tag{5}$$

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

Equation (5) implies that the aggregate regional output is given by:

$$Y_i = A_i \left(H_i^E\right)^{1-\alpha-\beta-\delta} \left(H_i^W\right)^{\alpha} K_i^{\delta} T^{\beta}. \tag{6}$$

Using Equation (6), one can determine wages, profits, and capital rental rates as a function of regional factor supplies via the usual (private) marginal product pricing. That is, the profit $\pi_i(h)$ earned by an individual with human capital h in region i is equal to h times the return of entrepreneurial human capital in the region, $\partial Y_i/\partial H_i^E$. The same individual can earn a wage income equal to h times the return to workers' human capital in the region $\partial Y_i/\partial H_i^W$. A labourer j with human capital h_j chooses to be an entrepreneur if and only if $(\partial Y_i/\partial H_i^E)\cdot h_j > \partial Y_i/\partial H_i^W \cdot h_j$ and a worker if $(\partial Y_i/\partial H_i^E)\cdot h_j < \partial Y_i/\partial H_i^W \cdot h_j$. In equilibrium, labourers must be indifferent between the two occupations, 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) describes the allocation of labour within in a region from the total quantities of human and physical capital (H_i,K_i).

The spatial equilibrium: consumption, housing and mobility

To compute the allocation of human capital, we must characterize labour mobility by computing 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. Since the aggregate income of labourers in region i is equal to w_iH_i, the demand for housing in the region is $\theta\cdot w_iH_i/\eta_i$. Given the unitary housing supply, the housing rental rate is equal to $\eta_i = \theta\cdot w_i\cdot H_i$. As a consequence, the utility (gross of moving costs) of a labourer in region i is equal to:

$$u_{w,i}(c,a) = \frac{w_i h}{\eta_i^\theta} = \frac{w_i^{1-\theta}}{\theta^\theta}\cdot\frac{h}{H_i^\theta}, \tag{8}$$

which rises with the wage and falls with regional human capital H_i due to higher rents. To find the spatial equilibrium, we need to find the ratio between wages paid in productive and unproductive regions, which determine the incentive to migrate. By taking capital mobility and external effects into account, in Appendix 1 we show that:

$$\frac{w_P}{w_U} = \left(\frac{\tilde{A}_P}{\tilde{A}_U}\right)^{\frac{\delta}{1-\delta}} \left(\frac{E(h_P)^{\psi} L_P}{E(h_U)^{\psi} L_U}\right)^{\frac{\gamma}{1-\delta}} \cdot \left(\frac{H_U}{H_P}\right)^{\frac{\beta}{1-\delta}} \tag{9}$$

Ceteris paribus, the wage is higher in productive regions. A higher human capital stock has a negative effect on the wage because of diminishing returns but once externalities are taken into account the net effect is ambiguous. In the remainder we assume:

A.1 $\left(\frac{\tilde{A}_P}{\tilde{A}_U}\right) \cdot \left(\frac{\underline{H}_P}{\underline{H}_U}\right)^{\gamma\psi-\beta-\delta} > 1$,

which implies that the autarky wage and interest rates are higher in productive regions, so that both capital and labour tend to move there. We can then prove the following (in Appendix 1):

Proposition 1 Under the parametric restriction:

$$(\beta - \psi\gamma)(1 - \theta) + \theta(1 - \delta) > 0, \tag{10}$$

there is a stable equilibrium allocation H_P and H_U. In this allocation:

a) There is a cutoff h_m such that agent j migrates from an unproductive to a productive region if and only if $h_j \geq h_m$. The cutoff h_m increases in the mobility cost φ.

b) Denote by $\underline{H} \equiv p\underline{H}_P + (1-p)\underline{H}_U$ the aggregate human capital. Then, when $\varphi = 0$, the equilibrium level of human capital in region i is independent of the region's initial human capital endowment. In particular, for $\psi = 1$ the full mobility allocation satisfies:

$$H_P = \tilde{H}_P^{free} \equiv \frac{A_P^{\frac{1-\theta}{(\beta-\gamma)(1-\theta)+\theta(1-\delta)}}}{E\left[A^{\frac{1-\theta}{(\beta-\gamma)(1-\theta)+\theta(1-\delta)}}\right]} \cdot \underline{H}\,. \tag{11}$$

When $\varphi > 0$ and $\psi \geq 1$, we have that $H_P < \tilde{H}_P^{free}$ and H_P increases in $\underline{H}_P$ holding $\underline{H}$ constant.

Since wages (and profits) are higher in the productive than in the unproductive regions, labour migrates to the former from the latter. The cutoff rule in a) is intuitive: more skilled people have a greater incentive to pay the migration cost because the wage (or profit) gain they experience from doing so is higher. Even if mobility costs are zero, migration to the more productive regions is not universal. This is due to the limited supply of land T, which causes decreasing returns in production, and to the limited supply of housing, which implies that migration causes housing costs to rise until the incentive to migrate disappears. Regional externalities moderate the adverse effect of fixed supplies of land and housing on mobility. In fact, for migration to be interior, condition (10) must be met, which requires external effects $\psi\gamma$ to be sufficiently small relative to: i) the diminishing returns β due to land and ii) the sensitivity θ of house prices to regional human capital.

In equilibrium, wages are higher in the more productive regions, $w_P > w_U$, but the housing rental rate is also higher there, $\eta_P > \eta_U$. As a result, our model predicts that more productive regions should remain more productive even after mobility is taken into account. When migration is costless (Equation (11)), the human capital employed in a region only depends on its productivity. In this respect, Proposition 1 shows that for our regressions to estimate the effect of human capital, mobility must be imperfect (i.e., $\varphi > 0$). When ψ = 1 and $\varphi = 0$, national output is equal to:

$$Y = \widehat{A} H^{\gamma} \left(H^{E}\right)^{1-\alpha-\beta-\delta} \left(H^{W}\right)^{\alpha} K^{\delta} T^{\beta}, \tag{12}$$

where $\widehat{A}$ is a function $\widehat{A}(\beta, \delta, \theta, \widetilde{A}_P, \widetilde{A}_U, p, \psi, \gamma)$ of exogenous parameters. More generally, under condition (10) the Lucas-Lucas model yields the following equation for firm level output:

$$y_{i,j} = \widetilde{A}_i E_i(h)^{\psi\gamma} L_i^{\gamma} h_j^{1-\alpha-\beta-\delta} H_{i,j}^{\alpha} K_{i,j}^{\delta} T_{i,j}^{\beta}, \tag{13}$$

and the following equation for regional output:

$$Y_i = \widetilde{A}_i E_i(h)^{\psi\gamma} L_i^{\gamma} (H_i^{E})^{1-\alpha-\beta-\delta} (H_i^{W})^{\alpha} K_i^{\delta} T^{\beta}. \tag{14}$$

Value added (at the regional and firm levels) does not depend on local prices after inputs are accounted for because output in our model consists only of the tradable consumption good.

Empirical Predictions of the Model

To obtain 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:

$$h_j = \exp\left(\mu_j S_j\right), \tag{15}$$

where $S_j \geq 0$ and $\mu_j \geq 0$ are two random variables (distributed according to a density $g_i(S,\mu)$ that ensures that the distribution of h_j is Pareto). The return to schooling μ_j varies across individuals, potentially due to talent. This allows us to estimate different returns to schooling for workers and entrepreneurs. Card (1999) offers some evidence of heterogeneity in the returns to schooling. In line with macro studies, in our regressions we express average human capital in the region as a first order expansion around the mean Mincerian return and years of schooling $E(h_i) \cong e^{\bar{\mu}_i \cdot \bar{S}_i}$, where $\bar{S}_i$ is average schooling while $\bar{\mu}_i$ is the average Mincerian return, both computed in region *i*.

Regional Income Differences

To test Equation (14) we must express physical capital, for which we have no data, as a function of human capital. The equalization of the return to capital implies $K_i = B A_i^{\frac{1}{1-\delta}} H_i^{\frac{1-\beta-\delta}{1-\beta}}$ where B>0 is a constant. Substituting this condition and the linearized expression for human capital into (14) we find:

$$\ln(Y_i/L_i) = C + [1/(1-\delta)]\ln\tilde{A}_i + [1+\gamma\psi - \beta/(1-\delta)]\,\bar{\mu}_i\,\bar{S}_i + [\gamma - \beta/(1-\delta)]\ln L_i, \tag{16}$$

where C is a constant absorbed by the country fixed effect. The coefficient on average regional schooling captures the product of the "technological" parameter (1+ $\gamma\psi - \beta$) and the nation-wide average $\bar{\mu}$ of the regional Mincerian returns $\bar{\mu}_i$. The coefficient $[\gamma - \beta/(1-\delta)]$ on population L_i captures the benefit γ of increasing regional workforce in terms of externalities minus the cost β of crowding the fixed land supply. A similar interpretation holds with respect to the schooling coefficient (1+ $\gamma\psi - \beta$).

If the variation in regional schooling and population is mostly due to imperfect mobility (φ>0), the estimated coefficients on schooling and population should reflect their theoretical counterparts in (16). In our model productivity also varies because of limited migration, owing to the fixed housing supply. This creates a serious concern: since in our model some human capital migrates to more productive regions, any mismeasurement of regional productivity A_i may contaminate the coefficient of regional human capital. We deal with this issue in two steps. First, we control in regression (16) for proxies of A_i. Second, we compare these results to the coefficients obtained from the firm level regressions and to the calibration exercises performed by the development accounting literature. These comparisons allow us to assess the severity of the endogeneity problem in the estimation of (16).

Firm-Level Productivity

In (13), the output of a firm j operating in region i depends on the human capital $h_{E,j}$ of the entrepreneur, as determined by his schooling $S_{E,j}$ and return to schooling $\mu_{E,ij}$, and on the average human capital $\mathrm{E}(h_{W,j})$ of workers, which again we approximate by $e^{\bar{\mu}_{W,j} \cdot \bar{S}_{W,j}}$ (where $\bar{\mu}_{W,j}$ and $\bar{S}_{W,j}$ are average values in the firm's workforce). Ceteris paribus, in our model entrepreneurs have a higher return to schooling than workers because in region *i* an entrepreneur with schooling *S* is someone whose return satisfies $e^{\mu S} \geq h_{E,i}$, 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 talented labourers whose return satisfies $\mu \geq \mu_{E,i}(S) \equiv \ln h_{E,i} / S$ while labourers with $\mu < \mu_{E,i}(S)$ become workers.

By writing Equation (16) in terms of firm-level output per worker $y_{i,j}/l_{i,j}$ and by exploiting the expressions for entrepreneurs' and workers human capital, we obtain the prediction:

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

$$(1-\alpha-\beta-\delta)\ln(l^{E}_{i,j}/l_{i,j}) + \alpha\ln(l^{W}_{i,j}/l_{i,j}) + \delta\ln k_{i,j} + \beta\ln t_{i,j} + \gamma\ln L_i + \gamma\psi\,\overline{\mu}_i\,\overline{S}_i\,, \qquad (17)$$

where $x_{i,j} = X_{i,j}/l_{i,j}$ denotes per-worker values, $l^{E}_{i,j}/l_{i,j}$ and $l^{W}_{i,j}/l_{i,j}$ capture the share of a firm's employment on managerial and non-managerial jobs, respectively. The coefficient on entrepreneurial schooling is the product of entrepreneurial rents $(1-\alpha-\beta-\delta)$ and the Mincerian return to entrepreneurial education $\overline{\mu}_E$. The coefficient on workers' schooling is the labour share α times $\overline{\mu}_W$, the Mincerian return of workers. The coefficient on regional schooling is the product of the externality parameter $\gamma\psi$ and the population-wide average Mincerian return $\overline{\mu}$.[5]

The estimation of (17) allows us to separate the role of the "low human capital" of workers from the "high human capital" of entrepreneurs in shaping firm productivity, as well as to get at the effect of human capital externalities by including regional human capital (and other controls). There are, however, two potential concerns. First, our model literally implies that output per-worker should be equalized across firms within a region. Realistically, though, output per-worker is equalized across firms ex-ante, but its ex-post value varies as a result of stochastic ex-post changes in the values of firm level TFP and inputs. This is the variation we appeal to when estimating (17).[6] Second, since the selection of talented entrepreneurs into more productive firms may contaminate our results, we employ the Levinsohn-Petrin (2003) instrumental variables approach.

[5] In the regional and firm level Equations (16) and (17) the average return to schooling should vary across regions. To account for this, one could 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 them to save space.

[6] Formally, if ex-ante a firm hires $X_{i,j}$ units of a factor, this results in $X_{i,j} = \varepsilon_X \cdot X_{i,j}$ units of the same factor being employed in production ex-post, where ε_X is a random shock to the value of inputs (e.g. an unpredictable change in the value of equipment, size of the workforce, and so on). Given the Cobb-Douglas production function, the firm's ex-ante optimization problem (occurring with respect to the ex-ante inputs $X_{i,j}$) does not change with respect to Equations (4) and (5). The only change is that a firm's productivity also includes expectations of the random factors ε_X. Crucially, this formulation implies that ex-ante returns are equalized, ex-post returns are not, which allows us to estimate (17) insofar as our input measures captures the ex-post values $X_{i,j}$. In estimation, we deal with the endogenous adjustment of inputs by using the Levinsohn-Petrin instrumental variables approach, and view the remaining productivity differences across firms as being the result of classical measurement error.

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 <u>states</u> or <u>provinces</u>, which are further subdivided into <u>counties</u> or <u>municipalities</u>. 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 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

---

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

[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, and travel time 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 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 within a region.

on geography, and climate 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. The online data Appendix 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 constraint on the sample is the availability of human capital data. All countries have periodic censuses and thus have sub-national data on human capital, but these data are hard to find.

Figure 1 portrays the 1,569 regions in our sample. It shows that coverage is extensive outside of North and sub-Saharan Africa. Sample coverage rises with a country's surface area, total GDP, but not GDP per capita. For example, we only have data for 7 of the smallest by surface area 50 countries, 9 of the 50 lowest GDP in 2005 countries, but for 26 of the lowest 50 GDP per capita countries.

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 Kazakhstan), and IPUMS (Israel).[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

[9] We are missing regional income per capita for Bangladesh and Costa Rica and national income per capita in PPP terms for Cuba. When regional income data for 2005 is missing, we interpolate regional income shares using as much data as is available for the period 1990-2008 or, when interpolation is not possible, the closest available year.

(1 country) to fill-in missing values. We measure regional income in current purchasing-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 the GDP at the country level.

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.

In addition, we examine productivity and its determinants using data from the Enterprise Survey for as many as 6,314 establishments in 20 countries and 76 of the regions in our sample.[11] Sample size is sharply reduced because we estimate alternative OLS specifications on a fixed sample of firms. The Enterprise Survey covers establishments owned by formal firms with five or more employees. We collect firm-level controls such as age, foreign ownership, as well as the number of establishments owned by the firm. We also collect establishment-level data on sales, exports, cost of raw materials, cost of labor, cost of electricity, and book value of assets (i.e. property, plant, and equipment). Critically, some of the Enterprise Surveys keep track of the highest educational attainment of the establishment's top manager as well as of that of its average worker. Panel data at the firm level is available for only 7 of the countries in our sample. Finally, we collect the two-digit SIC code (e.g., food, textiles, chemicals, etc.) of the establishments in our sample. These exclude OECD countries, as well as informal firms. We relate regional economic development to: (1) geography, (2) education, (3) institutions, and (4) culture. We restrict attention to regional variables available for at least 40 countries and 200 regions.

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

[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 in the OLS regressions. In contrast, we use all available years in the panel regressions.

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]

We gather data on the educational attainment of the population 15 years and older for 106 countries and 1,519 regions from EPDC Data Center (55 countries), Eurostat (17 countries), National Statistics Offices (27 countries) and IPUMS (8 countries); see the online data appendix for sources. We also gather data on the educational attainment of the population 66 years and older from IPUMS for 39 countries. 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 (2010). 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 and 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

[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 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.

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 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 (13.05%). 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 largely 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).[13] 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.

[13] 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

To shed light on the channels through which education affects regional income, we gather census data on occupations for as many as 565 regions in 35 countries. We focus on the incidence of directors and officers as well as employers in the workforce.

We create an index of the quality of institutions based on seven variables from the Enterprise Survey and one from the Sub-national Doing Business Reports. The Enterprise Survey covers as many as 80 of the countries and 428 of the regions in our sample.[14] The Enterprise Survey asked business managers to quantify: (1) informal payments in the past year, (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. The Enterprise Survey also asks managers to rate a variety of obstacles to doing business, including: (5) access to land, and (6) access to finance.[15] For each of these obstacles to doing business, we keep track of the percentage of the respondents that rate the item as a major or a very severe obstacle to business. The final Enterprise Survey variable we use is government predictability (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). We also use the overall ranking of the business environment from sub-national *Doing Business* reports, which summarizes government regulations in a range of areas, including starting a new business, enforcing contracts, registering property, and dealing with licenses. The index of the quality of institutions is the latent variable that captures the common variation in these eight variables (the online appendix presents the results for individual variables).

---

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] The main reason why we have more regions with measures of institutions than regions with productivity data is because many Enterprise Surveys lack data on the education of managers. For the computation of our index of institutional quality, we required a minimum of 10 establishments answering the particular institutions question.

[15] 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. We also gathered data on public infrastructure (e.g. power lines, air fields, highways, roads) from the US Geological Survey Global GIS database as well as the average travel time between cells in a region and the nearest city of 50,000 or more from the Global Environment Monitoring Unit. These variables are generally insignificant in regional income regressions (see the online appendix).

To measure culture, we gather data on trust in others from the World Value Survey (WVS) for as many as 69 countries and 745 regions.[16] Specifically, we focus on the percentage of 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?"[17] In addition, as a rough proxy for ethnic fractionalization, we gather data on the number of ethnic groups that inhabited each region in 1964 for up to 1,568 regions and 110 of our sample countries.[18]

In addition to running regressions using regional data, we examine GDP per capita at the country level, which comes 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. 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.

Table 1 summarizes our data. For each variable used in the regional regressions, Table 1 shows the number of regions for which we have data, the number of countries, the median value of the country mean, the median range and standard deviation within a country, and the ratio of the variable in the region with the highest vs. lowest GDP per capita. 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 3.77 for Africa, 5.63 for Asia, 3.74 for Europe, 4.60 for North America, and 5.61 for South America. The country with the highest ratio of incomes in the richest to that in the poorest region is Russia (43.30); the country with the lowest ratio is Pakistan (1.32).

[16] The WVS was collected between 1981 and 2005. When data from WVS for a country are available for multiple years, we use the most recent data. We set to missing 38 WVS observations in five countries (France, Japan, Philippines, Russia, and the United States) because the sub-national units in WVS are very coarse.

[17] From WVS, we also examined proxies for civil values (Knack and Keefer, 1997), for confidence in various institutions, for what is important in people's lives, as well as for characteristics valued in children. We also examined proxies for broad cultural attitudes with regards to authority, tolerance for other people, and family. Finally, we examined the percentage of respondents that participate in professional and civic associations. The results for these variables are qualitatively similar to those for trust in others that we discuss in the text.

[18] We also gathered data on 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. The results for linguistic fractionalization are qualitatively similar to the results for ethnic fractionalization that we discuss in the text.

Interestingly, this ratio is 5.16 for the United States, 2.59 for Germany, 1.93 for France, and 2.03 for Italy. 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 income inequality is not unusual.

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 2.74 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 Kenya (12.99), where education is 8.00 in Nairobi but only 0.62 in the North Eastern region. The lowest ratio is .62 in Malawi, where the Central region has lower education than the Central region (1.73 vs. 2.79) despite having higher income per capita ($739 vs. $555). Perhaps not surprisingly, there is more variation between rich and poor regions in the fraction of the population with a college degree than in the level of education. On average, the ratio of the fraction of the population with a college degree in the richest region to that in the poorest region is 4.70. To continue with the example of Kenya, 19.5% of the population older than 15 years in Nairobi has a college degree while only .9% of the comparable population in the North Eastern region completed college.

The patterns of inequality among regions within countries are interesting for other variables as well. Table 1 shows large differences in the incidence of employers as well as directors and officers in the workforce. There is also considerable variation across regions in both culture and institutions. On average, the quality of institutions is *lower* in the richest region than in the poorest one, which suggests that regional differences in institutions may have trouble explaining differences in economic development. Differences in endowments between rich and poor regions, such as temperature and distance to coast, are small.

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 because in our model we cannot express national output in closed form. More importantly, the estimated coefficients of education in the cross-country regressions may pick up the effect of omitted variables. The inclusion of country fixed effects in the regional regressions alleviates this concern. With respect to regional income, our benchmark is Equation (16). We have measures of average education at the regional level, but we do not have either national or regional data on physical capital or other inputs, so these variables only appear in the firm-level regressions in Section V.

Table 2 presents our basic regional results in perhaps the most transparent way. It reports the results of univariate regressions of regional income on its possible determinants, all with country fixed effects. Such specifications are loaded in favor of each variable seeming important since it does not compete with any other variable. We report both the within country and between countries R^2 of these regressions. The first column shows that education explains 58% of between country variation of per capita income, and 38% of within country variation of per capita income. Figure 2 shows, for Brazil, Colombia, India, and Russia the striking raw correlation between regional schooling and per capita income. The results are qualitatively similar if we use the fraction of the population with a high school degree or that with a college degree. Regional population explains only 3% of between country variation of per capita income and 1% of within country variation of per capita income.

Although several other variables in Table 2 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 reserves explain a trivial amount of variation at either level. The index of institutional quality explains 25% 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 the index explains 0% of within country variation of per capita incomes. Although some of the individual components of the index, such as access to finance or the number of days it takes to file a tax return, explain as much as 25% of cross-country variation, none explains more than 2% of within country variation of per capita incomes (see online appendix).[19] Cultural variables account for a substantial share of between country variation but none accounts 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).

Tables 3 and 4 show the multivariate regression results at the national and regional level. Table 3 presents regressions of national 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 together explain an impressive 50% of the variance. Education is also statistically significant, with a coefficient of .26, raising the R^2 to 63%. Next we add, one at a time, two measures of institutions (our index and expropriation risk) and two measures of culture (trust in others and the number of ethnic groups). Education remains highly statistically significant in each specification, and its coefficient does not fall much. At the country level, both institutional quality and expropriation risk are statistically significant with coefficients of 0.32 and 0.36, respectively. In contrast, proxies for culture are statistically insignificant. The final specification combines geography, education, institutions, and culture in one regression. Although we lose roughly two thirds of the observations, there are no surprising results: the coefficient on years of education drops to 0.15 but remains the most powerful

[19] Consistent with the results on institutions, two indicators of infrastructure – density of power lines and travel time between cities—explain a substantial fraction of the cross-country variation but much less within-country variation (see online appendix). Density of power lines account for 36% of cross country variation but only 5% of within country variation. Travel time accounts for 15% of cross country variation but only 7% of within country variation.

predictor of GDP per capita, while distance to the coast, oil reserves, and risk of expropriation are also statistically significant, although their combined explanatory power is low.

The last two rows of Table 3 show the adjusted R^2 of each regression if we omit the institutional (or cultural) variable, as well as the adjusted R^2 if we omit education. The impact on R^2 of dropping education ranges from a sharp reduction in the specifications that controls for the quality of institutions and the number of ethnic groups (columns 3 and 6) to a modest increase in the specification that includes risk of expropriation (column 4). The risk of expropriation has a 76% sample correlation with years of schooling. These results illustrate the difficulty of disentangling the effect of institutions and human capital in cross-country regressions (see Glaeser et al. 2004).[20]

Table 4 presents the corresponding results at the regional level, including country fixed effects. Among the geography variables, inverse distance to coast is the most robust predictor of regional income per capita. The education coefficient is slightly higher than in Table 3, and is highly significant, as illustrated in Figure 3~~5~~. When we include our proxies for institutions and culture one at a time, we find a small adverse effect of ethnic heterogeneity on income and no effect of the quality of institutions or of trust in others.[21] Institutional quality is insignificant and its incremental explanatory power is tiny. Combining our proxies for human capital, institutions and culture in one specification, we find that the coefficient on years of education rises from 0.27 to 0.37 and is highly significant while inverse distance to the coast is the only other variable that is statistically significant (at the 10% level). The last four rows of Table 4 show the within and between country adjusted R^2 of each regression if we omit the institutional or cultural variable, as well as the analog statistics if we omit education. While geography,

[20] Risk of expropriation has the highest explanatory power among standard measures of institutions, such as constraints on the executive, proportional representation, and corruption (see the online appendix).

[21] The region's ranking in the Doing Business report is the only component of the quality of institutions variable that is statistically significant but its incremental explanatory power is tiny (see online appendix). In results reported in the online appendix, we also find a small adverse effect of travel time but no role for other infrastructure variables such as the density of power lines. Finally, we find no role for cultural variables such as linguistic fractionalization and civic values.

institutions, and culture jointly explain a respectable fraction of the cross-country variation, they explain at most 16 percent of the within-country variation. In contrast, education explains a large fraction of the variance both across and within countries.

The final regression in Table 4 addresses the concern that the coefficient on education is biased because richer regions invest more in education. To address this simultaneity bias, we include in the regression years of education for the population over 65 years old rather than for the population over 14 years as we do in all other regressions. The results show that the estimated coefficient on years of education for the population over 65 years old is highly statistically significant and only marginally lower than the coefficient of the standard measure of education in column 2 (0.25 vs. 0.28). These results should alleviate concerns about the simultaneity bias. We discuss the omitted variable bias when we present firm-level regressions in the next section.

We have conducted several robustness checks of our basic findings, and here summarize them but do not present the results. First, we eliminated regions that include national capitals from the regressions; the results are not materially affected. Second, we included measures of regional population density in the specifications; density is typically insignificant and other results are not importantly affected. Third, we have tested the robustness of these results using data on regional luminosity instead of per capita income (see Henderson, Storeygard, and Weil 2009 and 2011). 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.

The low explanatory power of institutions is puzzling: since institutional quality rises with income, the endogeneity of institutions should if anything raise the coefficients. The measures we use (but also the components of the aggregate index) are standard and theoretically appropriate. In general, subjective assessments correlate much better with measures of development than objective measures of institutions (Glaeser et al. 2004). Even subjective assessments of institutions have low explanatory power in the sample of developing countries covered by the Enterprise Survey (see online appendix). The weakness of institutional variables may result in part from different data and in part from our focus on poorer countries, for which institutional variables indeed matter less.

Due to potential migration of better educated workers to more productive regions, we cannot interpret the large education coefficients - which appear to come through with a similar magnitude across a range of specifications – as the causal impact of human capital on regional income. We next estimate the role of human capital in the production function by looking at firm level evidence based on Enterprise Surveys, which allows us to partially address this problem by including region fixed effects as well as by taking advantage of panel data. By combining estimation and calibration, we then assess the extent to which the role of human capital at the firm level can account for its role across regions.

V. Establishment-Level Evidence.

In Table 5, we turn to the micro evidence and estimate essentially Equation (17). We use the Enterprise Survey data described in Section III. We estimate OLS regressions using a single cross-section of 6,314 firms in 20 countries and panel regressions using 2,922 firms in 7 countries.[22] We report results using a rough measure of value added, namely the logarithm of sales net of raw material and energy inputs, as the dependent variable.[23] We use the log of the number of employees as a proxy for ~~of~~ $l_{i,j}$. We measure capital (which includes both land $t_{i,j}$ and physical capital $k_{i,j}$) by the log of property, plant and equipment but also use the log of expenditure on energy as a proxy for it. We also include firm-level controls such as age, number of establishments, exports, and equity ownership by foreigners.

[22] Panel data for two of the countries in our sample (Brazil and Malawi) is available but we can't use it because data on schooling is missing for one of the years.
[23] Results are qualitatively similar if we use the log of sales as the dependent variable (see online appendix).

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. We thus implicitly assume that the establishment's top manager plays the role of the entrepreneur in our Lucas-Lucas model. As we explained in Section II, the Mincer model implies that schooling should enter the specification in levels, rather than in logs. We include geographic variables to control for exogenous differences in productivity.[24] To capture scale effects in regional externalities, we control for the log of the region's population L_i.

In Table 5, we begin with three OLS specifications. In the most parsimonious specification in the first column, we include proxies for geography and regional education, worker and manager schooling, log number of employees, log of property, plant, and equipment, and industry fixed effects (for 16 industries). Errors are clustered at the regional level. The estimated coefficient on capital is only 0.24 while the estimated coefficient on labor is .86. To address concerns over measurement error, the second specification adds the log of energy expenditure as a proxy for physical capital. The estimated coefficient on labor drops to 0.68 while the sum of the estimated coefficients on capital and energy is 0.42. The third specification adds to the previous one four firm-level controls, namely log firm age, a dummy variable if the firm has multiple establishments, the percentage of sales that are exported, and the percentage of the equity owned by foreigners. These firm-level controls have the expected signs and are highly statistically significant. Yet, including these controls does not materially change any of the coefficients of interest.

Depending on the specification, the coefficient on management schooling ranges from 0.026 to 0.015 while the coefficient on worker schooling takes values between .017 and .015. The similarity in

[24] Consistent with the findings for regional data, measures of regional institutions and infrastructure are usually insignificant, and hence we do not focus on these results. The coefficient on management schooling may be biased insofar as our regional proxies leave out much of the variation in A_i. To address this issue, we estimate (17) by controlling for the full set of region x industry dummies. The results on years of schooling of managers and workers are robust to including region x industry fixed effects (see online appendix).

the magnitude of the management and worker schooling coefficients drives our calibration exercise. In the context of Equation (17), this implies that $(1–\alpha–\beta–\delta)\,\overline{\mu}_{E,i}$ is roughly equal to $\alpha\,\overline{\mu}_{W,i}$. The return on entrepreneurial schooling must thus be substantially higher than that on worker schooling because the labor share α is typically much higher than the entrepreneurial share $(1–\alpha–\beta–\delta)$.

The coefficient on regional schooling is statistically significant across specifications and varies in a narrow range between .07 and .09. In so far as there is large measurement error in workers' schooling at the firm level, regional education may provide a more precise proxy for workers' skills, creating a false impression of human capital externalities. This, however, is unlikely to be the case since the average education of workers does not vary much across firms within regions. Consistent with agglomeration economies, the coefficient on regional population is positive, ranging from .10 and .12 depending on the specification. Finally, the coefficients on geography variables are generally insignificant. Thus, the most obvious proxies for omitted regional productivity do not appear to be important. These results on geography should partially address the concern that regional schooling picks up the effect of omitted regional productivity.

In the OLS results in Table 5, the coefficients on production inputs (including managerial and worker education) may be biased by unobservable differences in firm-level productivity. In the last column of Table 5, we follow Levinsohn and Petrin's (2003) panel data approach and use expenditure on energy to control for the unobserved correlation between production inputs and productivity. These panel data results need to be interpreted with caution because we have at most three observations per establishment. None of the regional variables come in significant, most likely because we only have panel data for 22 regions in 7 countries. Turning to the firm-level variables, the results are consistent with our earlier findings. The coefficient on labor is .62 while that on property, plant, and equipment is .34. The estimated coefficients on managerial and worker schooling are close to their respective OLS levels: the coefficient on management schooling rises to .027 from .015 under OLS while the coefficient on worker schooling rises to .032 from .015 under OLS.

We added additional controls to these regressions, and obtained similar results. Most of the specifications confirm both the general findings, and parameter estimates, in Table 5. There does not appear to be much evidence of significant omitted regional effects, although since we do not have all of the determinants of regional productivity, our assessment of external effects might be exaggerated.

In light of this evidence, it is interesting to go back to the regional data and ask: If entrepreneurs/managers are so important in determining firm-level productivity, can we also find evidence of their influence on regional income? To address this issue, Table 6 uses an approach similar to that in Table 4 but focuses on the composition of human capital and the structure of the workforce in explaining regional differences in GDP per capita. We run regressions with and without years of education but always include the standard geography controls. We first examine whether the share of the population with a college degree –a measure of skilled labor—plays a special role (Vanderbussche et al. 2006). To this end, we divide the population in each region according to their highest educational attainment into three groups: (1) less than high school, (2) high school, and (3) college or higher. We then include in the regressions the share of the population with high school and, separately, that with college degree (the omitted category is the population with less than high school). To make the estimated coefficients comparable to those for years of education in Table 4, we multiply the shares of the population with college and high school degrees by 16 and 12, respectively (their weights in our standard measure of years of education). The estimated coefficient is higher for the (scaled) share of the population with college than with high school (0.25 vs. 0.20) but cannot reject the hypothesis that the two coefficients are equal (the F-statistic is 1.28).

The evidence in Table 5, like our model, suggests a positive correlation between regional income and the share of educated workers becoming managers. We use data on the fraction of the workforce classified by the census as directors and officers to explore this prediction. The data is noisy because occupational categories are not standardized across countries and data is available for only 28 countries (not all countries have census data online and not all censuses have detailed occupational data). With these caveats in mind, we find that, controlling for the percentage of the population with college and high school, increasing by one percentage point the fraction of the workforce classified as directors and officers is associated with an 8% increase in GDP per capita. This finding is robust to including the level of education. Focusing on the share of directors and officers that also have a college degree yields similar results: a percentage point increase in the fraction of college-educated directors and officers is associated with an increase in GDP per capita of 11% to 12%, depending on the specification. Consistent with our model, the incidence of doctors and government bureaucrats is uncorrelated with regional income per capita (see online appendix).

As an alternative way of looking at occupations, we include in the regressions the share of the workforce classified as employers. The results for employers suggest that increasing by one percentage point the share of employers in the workforce is associated with a 3 percent increase in GDP per capita when we control for educational attainment but the estimated coefficient drops in value (from 0.03 to 0.02) and becomes insignificant when we control for the level of education.

VI. Calibration.

Can the effects estimated from firm level regressions account for the large role of schooling in the regional regressions? How do these effects compare with the calibrations performed in development accounting? We first discuss the predictions of our model under a set of standard calibration values for the labor share α, the capital share ($\delta + \beta$), and the housing income share θ, but also consider a range of parameter values (particularly for the labor share α). The standard calibration for the U.S. labour share is about α =.6. We however calibrate α =.55 to reflect the fact that in developing countries the labour share tends to be lower than in the U.S., in part because a fraction of labour income remunerates entrepreneurship (Gollin 2002). We follow the standard calibration for the overall capital share and set it to .35, which falls between our firm level and panel estimates. These calibrations imply that managerial/entrepreneurial input accounts for $(1\text{–}\alpha\text{–}\beta\text{–}\delta)$ = (1–.55–.35) =.1 of value added.

From our estimated regressions we impose the following restrictions:

i) $\alpha\,\overline{\mu}_W$ =.03 and $(1\text{–}\alpha\text{–}\beta\text{–}\delta)\,\overline{\mu}_E$ =.025 (from Table 5, column 4).

ii) γ = .05 (from Table 5, column 4)

iii) $\gamma\,\psi\,\overline{\mu}$ = .074 (from Table 5, columns 1,2,3)

iv) $\gamma - \beta/(1-\delta)$ =.01 (from Table 4, column 2)

v) $[1+\gamma\psi-\beta/(1-\delta)]\,\overline{\mu}$ =.27 (from Table 4, column 2)

These specifications should not be viewed as "structural estimates" of model parameters, but rather as a means of finding what parameter values are in the ballpark of our regressions estimates. Note that our starting estimates for regional externalities in the firm level regressions do not come from the Levinsohn-Petrin method, which yields zero. We come back to this issue below.

Using these calibrated parameters, the above equations can be solved to yield:

$$\overline{\mu}_W = .055;\ \overline{\mu}_E = .25\ ;\ \overline{\mu} = .20;\ \delta = .32;\ \psi = 7.25;\ \beta = .03;$$

At these parameter values, the spatial equilibrium is stable, since $(\beta-\psi\gamma)(1-\theta)+\theta(1-\delta)$ = $(-.33)(.6)+(.4)(.68)>0$. Interestingly, some of these parameter values fall in the ballpark of existing micro-estimates. The land share β is just below estimates based on income accounts (Valentinyi and Herrendorf 2008). The return to worker schooling of 5-6% is consistent with micro evidence on workers' Mincerian returns (Psacharopoulos 1994). This finding suggests that our firm level productivity regressions reduce identification problems at least as far as firm-level variables are concerned.

The critical new finding is that our estimation results point to a Mincerian return $\overline{\mu}_E$ = .25 for entrepreneurs. This 25% estimate is higher than those found by Goldin and Katz (2008) for returns to college education for workers. However, entrepreneurial returns might be ignored in surveys focusing on wages as returns to education. The few existing analyses of entrepreneurial education document substantially higher returns to education for managers than for workers (Parker and van Praag 2005, van Praag et al. 2009).[25] The high returns to entrepreneurial education, compared to the relatively low returns to worker education, might explain the difficulty encountered by the development accounting literature when trying to use human capital to explain productivity differences across space (Caselli 2005, Hsieh and Klenow 2010). Individuals selected into entrepreneurship appear to have vastly more human capital than workers, driving up productivity. Of course, entrepreneurial talent may be more important than schooling in explaining this finding. Our analysis cannot address this issue (which would require better data and an endogenous determination of the connection between schooling and talent), but it still identifies a critical role of management and entrepreneurship in determining productivity.

The spatial differences in the stocks of human capital implied solely by returns to worker education are considerably lower than those implied by blended returns of workers and entrepreneurs. The average population-wide Mincerian return $\overline{\mu}$ of 20% is in fact substantially above the return to workers, and lies in between our estimates of workers' and entrepreneurs' values.[26]

[25] Using U.S. and Dutch individual-level data, these studies find that one extra year of schooling increases entrepreneurial income by 18% and 14%, respectively. This is much higher than the 3% found in our firm-level data (in our model entrepreneurial income is a constant share of a firm's output), implying gigantic Mincerian returns under an entrepreneurial share of .1. Note, however, that these studies rely on small start-ups (in the Dutch data) or on self employed individuals (in the U.S. data). In both cases, the entrepreneurial share is likely to be higher than .1, moving Mincerian returns closer to our benchmark of 25%.

[26] Although we lack direct data on the number of entrepreneurs in the economy, we can make a back-of-the-envelope calculation to assess whether our firm level evidence is consistent with a population-wide 20% Mincerian return. If: (1) an average entrepreneur is as educated as the entrepreneurs in the enterprise survey on average, i.e. has 14 years of schooling; and (2) an average worker in the economy is as educated as the average worker in the sample, i.e. has roughly 7 years of schooling, then to obtain an average population-wide Mincerian return of 20% entrepreneurs need to account for 10.14% of the workforce. Formally, the fraction of entrepreneurs *f* solves the equation: $\exp[0.2*(14*f+7*(1-f))] = f*\exp(14*.25)+(1-f)*\exp(7*.055)$.

Consider now the role of externalities. The education externality parameter ψ we use is 7.25, although recall that Levinsohn-Petrin estimate is zero. This implies that a given increase in regional human capital generates 7.25 times more externalities if it is due to an increase in the average amount of human capital than to a larger number of people with average education. These estimates imply that raising the educational level from the sample mean of 6.58 years by one year increases regional TFP by about 7.5~~6~~%. The magnitude of human capital externalities has been heavily discussed in the literature. As Lange and Topel (2006) indicate in their survey, the results have been fairly diverse. For instance, Caselli (2005) and Ciccone and Peri (2006) find externalities to be unimportant. Rauch (1993) estimates a 3-5% effect, somewhat lower than our estimate. Acemoglu and Angrist (2000) estimate that a one year increase in average schooling is associated with a 1-3% increase in average wages. Moretti (2004) examines the impact of spillovers associated with the share of college graduates living in a city and finds that a 1-percent increase in the share of college graduates in the population leads to an increase in output of roughly half a percentage point. By way of comparison, under our variable definitions, a 1-percent increase in the share of college graduates in the population is associated with (at most) an additional .16 years of education and thus with a 1.2% (=.16x0.075) increase in regional TFP. Iranzo and Peri (2009) estimate that one extra year of college per worker increase the state's TFP by a very significant and large 6-9%, whereas the effect of an extra year of high school is closer to 0-1%. These estimates suggest a potentially sizeable effect of schooling for productivity via social interactions or R&D spillovers, consistent with Lucas (1985, 2009~~8~~) as well as with the literature in urban economics (e.g., Glaeser and Mare 2001, Glaeser and Gottlieb 2009). Externalities (whose empirical identification is admittedly much harder) may also improve the explanatory power of human capital, although we show below that they only help a lot when entrepreneurial returns are high.

We now assess the explanatory power of entrepreneurial inputs and externalities by using our parameter estimates to perform a standard development accounting exercise. To do so, define a factor-based model of national income as $\hat{Y}=E(h)^{\psi\gamma}L^{\gamma}H^{1-\beta-\delta}K^{\delta+\beta}$, which is national income predicted by our model when: i) all regions in a country are identical and all countries are equally productive, and ii) in line with standard development accounting we consider only physical and human capital, thereby attributing land rents to physical capital. This model with no regional mobility provides a benchmark to assess the role of physical and human capital when productivity differences are absent. Following Caselli (2005), one measure of the success of the model in explaining cross-country income differences is

$$success = \frac{\mathrm{var}(\log(\hat{Y}))}{\mathrm{var}(\log(Y))},$$

where Y is observed GDP per worker. Using Caselli's dataset, the observed variance of (log) GDP per worker is 1.32. Ignoring human capital externalities (i.e., assuming $\psi=\gamma=0$) and using the standard 8% average Mincerian return on human capital for both workers and entrepreneurs (i.e., setting $\bar{\mu}=8\%$), the variance of log($\hat{Y}$) equals 0.76, i.e. physical and human capital explain 57% (0.76/1.32) of the observed variation in income per worker. This calculation reproduces the standard finding that, under standard Mincerian returns, a big chunk of the cross country income variation is accounted for by the productivity residual.

To isolate the role of entrepreneurial capital, we compute $\hat{Y}$ assuming no human capital externalities (i.e., $\psi=\gamma=0$) while still keeping a population-wide Mincerian return $\bar{\mu}$ of 20%, consistent with our firm-level estimates. It is not surprising that average Mincerian returns of about 20% greatly improve the explanatory power of human capital. Indeed, under this assumption *success* rises to 81%. This improvement is solely due to accounting for managerial schooling. We note that this result is quite sensitive to our assumption of labor share of 55%. If the labor share were lower, the residual income share allocated to entrepreneurial rents would be correspondingly higher. This would reduce our estimate of the returns to entrepreneurial education, and therefore of average Mincerian returns. Finally, to assess the incremental explanatory power of human capital externalities, we compute

$\hat{Y}$ assuming our estimated values (i.e., ψ=7.25 and γ=.05), while retaining the assumption that the average Mincerian return equals 20%. Under these new assumptions, the model generates too much productivity variation, and s*uccess* rises to 103%.

Table 7 presents sensitivity results for the calibration exercise in this section. We focus on the predictions of the model when the labor share ranges between 50 and 60 percent while keeping the capital share β+δ constant at 35 percent, i.e. increases in the labor share of workers are offset by reductions in the labor share of entrepreneurs. Panel A presents results under the assumption that both $(1-\alpha-\beta-\delta)\bar{\mu}_{E,i}$ and $\alpha\bar{\mu}_{W,i}$ equal 0.03 while Panel B presents results under the assumption that they equal 0.02. In both panels, we assume that entrepreneurs are 5% of the workforce and have 14 years of education while workers have 7 years. We continue to use γ=.05, ψ=7.25, β=.03, and δ=.32. Table 7 shows that the average Mincerian return increases sharply with α. As α rises from 50 to 60 percent, the average Mincerian return rises from 11 to 74 percent in Panel A (i.e. when $\alpha\bar{\mu}_{W,i}$ =.03) and from 6 to 37 percent in Panel B (i.e. when $\alpha\bar{\mu}_{W,i}$ =.02). These changes in Mincerian returns take place because $\bar{\mu}_{E,i}$ compounds during 14 years and it triples as the labor share rises from 50 to 60 while $\bar{\mu}_{W,i}$ compounds for 7 years and falls modestly (from 6 to 5 percent in Panel A and from 4 to 3.3 percent in Panel B).

It is clear from Table 7 that $\bar{\mu}_{E,i}$ needs to be high (i.e. in excess of 25%) for our model to add meaningful explanatory power beyond that of models that do not account for entrepreneurial inputs. Externalities play second fiddle; they have a minor impact on the success ratio when $\bar{\mu}_{E,i}$ is low and, conversely, they only come into play when $\bar{\mu}_{E,i}$ is high. This raises the question of how plausible are high levels of $\bar{\mu}_{E,i}$. To assess this issue, Table 7 reports the ratio of the entrepreneur-to-worker income for different Mincerian returns. When $\bar{\mu}_{E,i}$ is 25%, the entrepreneur-to-worker income ratio equals

22.3 in Panel A and 25.9 in Panel B. This ratio rises to 73.1 in Panel A and 83.9 in Panel B when $\overline{\mu}_{E,i}$ equals to 33%. Such levels of income inequality seem plausible for developing countries (Towers and Perrin 2005). In contrast, income inequality is too low when $\overline{\mu}_{E,i}$ is 20% (i.e. 10.8x and 12.7x).

To appreciate the importance of entrepreneurial inputs in understanding cross-country income difference, compare Mozambique and the US. Income per worker is roughly 33 times higher in the US than in Mozambique ($57,259 vs. $1,752), while the stock of physical capital per capita is 185 times higher in the US than in Mozambique ($125,227 vs. $676). The average number of years of schooling for the population 15 years and older is 1.01 years Mozambique and 12.69 years in the United States. These large differences in schooling imply that the (per capita) stock of human capital is 10.3 higher ($H_{US}/H_{MOZ}=e^{.20*(12.69-1.01)}$) in the US than in Mozambique if the average Mincerian return is 20%. In contrast, the (per capita) stock of human capital is only 2.5 times higher ($H_{US}/H_{DRC}=e^{.08*(12.69-1.01)}$) in the US than in Mozambique if the average Mincerian return is 8%. Using weights of 1/3 and 2/3 for physical and human capital, these differences in physical and human capital imply that income per capita should be 27 times higher in the US than in Mozambique ($27 = 10.3^{2/3}$x$185^{1/3}$), which is much closer to the actual value of 33 times than the 10.6 multiple implied by 8% Mincerian return ($10.6=2.5^{2/3}$x$185^{1/3}$).

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, ii) entrepreneurial inputs area fundamental and relatively neglected channel for understanding the role of schooling in shaping productivity differences, and iii) human capital externalities may magnify the impact of entrepreneurial inputs. Our parameter estimates point to very large returns to entrepreneurial schooling (perhaps due to entrepreneurs' general talent) and to large social returns to education at the regional level.

VII. Conclusion.

Evidence from more than 1,500 sub-national regions of the world suggests that regional education is the critical determinant of regional development, and the only such determinant that explains a substantial share 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 perhaps regional externalities. The latter come primarily from the level of education (the quality of human capital) in a region, and not from its total quantity (the number of people with some education).

A simple Cobb-Douglas production function specification used in development accounting would have difficulty accounting for all this evidence. Instead, we presented a Lucas-Lucas model of an economy, which combines the allocation of talent between work and entrepreneurship, human capital externalities, and migration of labor across regions within a country. The empirical findings we presented are both consistent with the general predictions of this model, and provide plausible values of the model's parameters. In addition, we follow Caselli (2005) in assessing the ability of the model to account for variation of output per worker across countries. 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) and possibly also social returns to education through external spillovers, are large. To the extent that earlier estimates of return to education have missed the benefits of educated managers/entrepreneurs, they may have underestimated the returns to education.

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 more important than having educated workers. Consistent with earlier observations of Banerjee and Duflo (2005) and La Porta and Shleifer (2008), economic development occurs in regions that concentrate entrepreneurs, who run productive firms. These entrepreneurs may also 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.

Bibli.ography

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, and James 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.

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.

Barro, Robert and Jong-Wha Lee, April 2010, "A New Data Set of Educational Attainment in the World, 1950-2010." NBER Working Paper No. 15902.

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

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

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

Bloom, Nicholas, and John Van Reenen (2007). “Measuring and Explaining Management Practices across Firms and Countries.” *Quarterly Journal of Economics* 122(4): 1351-1408.

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

Breton, Theodore R. (2012). “Were Mankiw, Romer, and Weil Right? A Reconciliation of the Micro and Macro Effects of Schooling on Income.” *Macroeconomic Dynamics*, 1-32.

Card, David (1999). “The Causal Effect of Education on Earnings.” In Orley Ashenfelter and David Card, eds., *Handbook 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 John Wilbur Coleman II (2006). “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.

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.

Coe, David, and Elhanan Helpman (1995). “International R&D Spillovers.” *European Economic Review* 39(5): 859-887.

Cohen, Daniel, and Marcelo Soto (2007). “Growth and Human Capital: Good Data, Good Results.” *Journal of Economics Growth*, 12 (1):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 Jones, and Benjamin Olken (2009). “Temperature and Income: Reconciling New Cross-Sectional and Panel Estimates.” NBER Working Paper No. 14680.

De Long, 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, and Joshua 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, and David Mare (2001). “Cities and Skills.” *Journal of Labor Economics* 19(2):316-342.

Glaeser, Edward, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer (2004). “Do Institutions Cause Growth?” *Journal of Economic Growth*, 9 (2):271-303.

Goldin, Claudia, and Lawrence Katz (2008). *The Race Between Education and Technology*. Cambridge, MA: Harvard University Press.

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

Hall, Robert, and Charles 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.” *American Economic Review*, forthcoming.

Henderson, Vernon, Adam Storeygard, and David N. Weil. 2011. "A Bright Idea for Measuring Economic Growth." *American Economic Review*, 101(3): 194–99.

Hsieh, Chang-Tai, and Peter 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.

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

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

Klenow, Peter, 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.

Lange, Fabian and Robert Topel, (2006). "The Social Value of Education and Human Capital”, Labor Markets and Economic Growth". *Handbook of the Economics of Education. Vol. 1*, p. 459-509.

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

Levinsohn, James and Amil Petrin (2003). “Estimating production functions using inputs to control for unobservables.” *Review of Economic Studies* 70(2): 317–342.

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

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

Lucas, Robert (2009). “Ideas and Growth.” *Economica*, 76(301): 1–19.

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

Michalopoulos, Stelios and Elias Papaioannou, 2011, “The Long-Run Effects of the Scramble for Africa”, NBER Working Paper No. 17620.

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

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

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

Parker, Simon and Mirjam van Praag, (2005). “Schooling, Capital Constraint and Entrepreneurial Performance.” Tinbergen Institute Discussion Paper.

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, NJ: Princeton University Press.

Rauch, James (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.

Syverson, Chad (2011). “What Determines Productivity?” *Journal of Economic Literature*, forthcoming.

Towers and Perrin (2005), “Managing Global Pay and Benefits 2005-2006”.

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

Vandenbussche, Jerome, Philippe Aghion, and Costas Meghir (2006), Growth, distance to frontier and composition of human capital.” *Journal of Economic Growth* 11: 97-127.

van Praag, Mirjam, Arjen van Witteloostuijn, and Justin van der Sluis (2009). “Returns for Entrepreneurs vs. Employees: The Effect of Education and Personal Control on the Relative Performance of Entrepreneurs vs. Wage Employees.” IZA Discussion Paper no. 4628.

Wolff, Edward (2011). “Spillovers, Linkages, and Productivity Growth in the US Economy, 1958-2007.” Working Paper No. 16864. Cambridge, MA: National Bureau of Economic Research.

Appendix 1.

Solution of the Model and proof of Proposition 1

Given Equation (6) for regional output, we can determine wages, profits, and capital rental rates as a function of regional factor supplies via the usual (private) marginal product pricing. That is:

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

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 $\pi_i(h) = \pi_i \cdot h$.

By exploiting the breakdown of human capital into its different components in Equation (7), one finds that ρ is constant across regions provided:

$$\frac{K_P}{K_U} = \left(\frac{A_P}{A_U}\right)^{\frac{1}{1-\delta}} \left(\frac{H_P}{H_U}\right)^{\frac{1-\beta-\delta}{1-\delta}}.$$

Using this condition and Equation (3), it is easy to see that the relative wage is given by Equation (9).

Consider now the determinant of spatial mobility. By A.1, labour moves from unproductive to productive regions. Formally, Equation (11) implies that an agent with human capital h_j migrates if $w_P^{1-\theta}(h_j - \varphi) / H_P^{\theta} \geq w_U^{1-\theta} h_j / H_U^{\theta}$, where φ captures migration costs. This identifies a human capital threshold h_m such that agent j migrates if and only if $h_j \geq h_m$. By exploiting the wage equation in (6) and the equilibrium condition (9), threshold h_m can be implicitly expressed as:

$$h_m \cdot \left[1 - \left(\frac{w_U}{w_P}\right)^{1-\theta} \left(\frac{H_P}{H_U}\right)^{\theta}\right] = \varphi \, . \qquad \text{(Ap.1)}$$

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

$$p \cdot H_P + (1-p) \cdot H_U = \underline{H}. \qquad \text{(Ap.2)}$$

After accounting for externalities, the equilibrium condition (Ap.1) can be written as:

$$h_m \cdot \left[1 - \left(\frac{A_U}{A_P}\right)^{\frac{1-\theta}{1-\delta}} \left(\frac{L_P}{L_U}\right)^{\gamma(\psi-1)\frac{1-\theta}{1-\delta}} \left(\frac{H_P}{H_U}\right)^{\frac{(\beta-\gamma\psi)(1-\theta)+(1-\delta)\theta}{1-\delta}}\right] = \varphi \cdot \tag{Ap.3}$$

The previous migration-threshold implies that the human capital stock in each productive region is:

$$H_P = \underline{H}_P + \frac{1-p}{p}\int_{h_m}^{+\infty} h \cdot (\mu_B \underline{h}^{\mu_B} h^{-\mu_B - 1}) \cdot dh = \underline{H}_P + \underline{H}_U \frac{1-p}{p}\left(\frac{\underline{h}}{h_m}\right)^{\mu_B - 1}. \tag{Ap.4}$$

Using Equation (Ap.4) and (Ap.3), it is immediate to express h_m as a function of H_P and thus recover:

$$\frac{L_P}{L_U} = \frac{1 + \frac{p}{1-p} \cdot \left(\frac{H_P - \underline{H}_P}{\underline{H}_U} \cdot \frac{p}{1-p}\right)^{\frac{\mu_U}{\mu_U - 1}}}{1 - \left(\frac{H_P - \underline{H}_P}{\underline{H}_U} \cdot \frac{p}{1-p}\right)^{\frac{\mu_P}{\mu_P - 1}}}. \tag{Ap.5}$$

Under full mobility (φ = 0), using (Ap.3) one finds that the equilibrium is determined by the condition:

$$\left(\frac{A_P}{A_U}\right)^{\frac{1-\theta}{1-\delta}} \left[\frac{1 - \left(\frac{H_P - \underline{H}_P}{\underline{H}_U} \cdot \frac{p}{1-p}\right)^{\frac{\mu_U}{\mu_U - 1}}}{1 + \frac{p}{1-p} \cdot \left(\frac{H_P - \underline{H}_P}{\underline{H}_U} \cdot \frac{p}{1-p}\right)^{\frac{\mu_P}{\mu_P - 1}}}\right]^{\gamma(\psi-1)\frac{1-\theta}{1-\delta}} = \left[\frac{(1-p)H_P}{\underline{H} - pH_P}\right]^{\frac{(\beta-\gamma\psi)(1-\theta)+(1-\delta)\theta}{1-\delta}}. \tag{Ap.6}$$

The left hand side is decreasing in H_P. If $(\beta - \psi\gamma)(1-\theta) + \theta(1-\delta) > 0$, the right hand side - which captures the cost of migrating to productive regions, increases in H_P. As a result, when $(\beta - \psi\gamma)(1-\theta) + \theta(1-\delta) > 0$ even under full mobility in the stable equilibrium there is no universal migration to productive regions. Indeed, if all human capital moves to productive regions, then $H_P = \underline{H}/p$ and the right hand side of (Ap.10) becomes infinite. Full migration is not an equilibrium. No migration is not an equilibrium either, as in this case A.1 implies that (Ap.10) cannot hold. When ψ = 1 (and φ = 0) the equilibrium has:

$$H_i = \frac{A_i^{\frac{1-\theta}{(\beta-\gamma)(1-\theta)+\theta(1-\delta)}}}{E\left[A^{\frac{1-\theta}{(\beta-\gamma)(1-\theta)+\theta(1-\delta)}}\right]} \cdot \underline{H}. \tag{Ap.7}$$

With imperfect mobility φ >= 0, the equilibrium fulfils the condition:

$$\frac{\varphi}{\underline{h}}\left(\frac{p}{1-p}\right)^{\frac{1}{\mu-1}}\left(\frac{H_P-\underline{H}_P}{\underline{H}_U}\right)^{\frac{1}{\mu-1}}=1-\left(\frac{A_U}{A_P}\right)^{\frac{1-\theta}{1-\delta}}\left[\frac{1+\frac{p}{1-p}\left(\frac{H_P-\underline{H}_P}{\underline{H}_U}\cdot\frac{p}{1-p}\right)^{\frac{\mu_P}{\mu_P-1}}}{1-\cdot\left(\frac{H_P-\underline{H}_P}{\underline{H}_U}\cdot\frac{p}{1-p}\right)^{\frac{\mu_U}{\mu_U-1}}}\right]^{\gamma(\psi-1)\frac{1-\theta}{1-\delta}}\left[\frac{(1-p)H_P}{\underline{H}-pH_P}\right]^{\frac{(\beta-\gamma\psi)(1-\theta)+(1-\delta)\theta}{1-\delta}}.$$

When $(\beta - \psi\gamma)(1- \theta) + \theta(1- \delta)> 0$, an increase in $\underline{H}_P$ (holding $\underline{H}$ constant) shifts down the left hand side and shifts up the right hand side above. As a result, the equilibrium is unique.

Appendix 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	
Income per capita	Income per capita in PPP constant 2005 international dollars in the region in 2005. We GDP as a measure of income for all countries except 20. For those 20 countries, 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). For each country, we scale regional income 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 income per capita data, we interpolate using all available data for the period 1990-2008. When interpolating income values is not possible, we use the regional distribution of the closest year with regional income 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 income.	*Regional Income*: 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
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.
Share Pop with high school degree	Share of the population aged 15 years or older whose highest educational level is ISCED 3 or 4.	See Years of education.
Share Pop with college degree	Share of the population aged 15 years or older whose highest educational level is ISCED 5 or 6.	See Years of education.
Years of education 65+	The average years of schooling from primary school onwards for the population aged 65 years or older. To compute this variable, we follow the same procedure as used for the previously described years of schooling variable at the regional level.	https://international.ipums.org/international/index.html
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.
% Directors and officers in workforce	Percentage of the economically-active population aged 15 years through 65 that most closely matches the employment category of company officers and general directors in the most recent population census.	https://international.ipums.org/international/index.html
% Employers in the workforce	Percentage of the economically-active population aged 15 years through 65 classified as employers in the most recent population census.	https://international.ipums.org/international/index.html

Variable	Description	Sources and links
	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
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 production per capita)	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	Word Bank's Doing Business Subnational Reports.

Variable	Description	Sources and links
	taxes. Higher values indicate more burdensome regulation. Data is for the most recent year available, ranging from 2007 through 2010.	http://doingbusiness.org/Reports/Subnational-Reports/
Institutional Quality	Latent variable of: (1) (minus) Informal payments, (2) (minus) Ln(Tax days), (3) (minus) Ln(Days without electricity), (4) (minus) Security costs, (5) (minus) Access to land, (6) (minus) Access to finance, (7) Government predictability, and (8) (minus) Doing Business percentile rank. Higher values indicate better institutions.	
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/.
	IV. 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/
Ln(Nbr 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
	V. Enterprise Survey Data	
Ln(Sales – Raw Materials - Energy)	The logarithm of the establishment's sales minus expenditure on raw materials and energy (in current PPP dollars). Data is for the last complete fiscal year, ranging from 2002 through 2009.	World Bank's Enterprise Surveys. https://www.enterprisesurveys.org/
Ln(Expenditure on Energy)	The logarithm of the establishment's expenditure on energy (in current PPP dollars). Data is for the last complete fiscal 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 current top manager of the establishment. To compute this variable, we use data on the highest educational attainment of the top manager and follow the same procedure as used for the previously described years of schooling variable at the regional level.	World Bank's Enterprise Surveys. https://www.enterprisesurveys.org/
Years of Education of workers	The number of years of schooling of a typical production worker employed in the establishment. Respondents answers may take the following values: (a) 0-3 years, (b) 4-6 years, (c) 7-9 years, (d) 10-12 years, (e) 13 years and above. To compute this variable, we use the midpoint of each range or 13 years as appropriate.	World Bank's Enterprise Surveys. https://www.enterprisesurveys.org/
Ln(1+ Employees)	The logarithm of the total number of employees in the establishment. Data is for the last complete fiscal year, ranging from 2002 through 2009.	World Bank's Enterprise Surveys. https://www.enterprisesurveys.org/
Ln(Property, plant, and equipment)	The logarithm of the establishment's book value of property, plant and equipment (in current PPP dollars). Data is for the last complete fiscal year, ranging from 2002 through 2009.	World Bank's Enterprise Surveys. https://www.enterprisesurveys.org/
Ln(1 + Firm Age)	The logarithm of one plus the number of years that the establishment had been operating in the country at the time of the survey , ranging from 2002 through 2009	World Bank's Enterprise Surveys. https://www.enterprisesurveys.org/
Multiple Establishments	Equal to one if either the establishment was part of a larger firm or the firm had more than one establishment at the time of the survey; equals zero otherwise.	World Bank's Enterprise Surveys. https://www.enterprisesurveys.org/
Percent Export	Percentage of the establishment's sales that were directly or indirectly exported. Data is for the last complete fiscal year, ranging from 2002 through 2009.	World Bank's Enterprise Surveys. https://www.enterprisesurveys.org/
Percent equity owned by foreigners	Percent of the firm's equity owned by private foreign individuals, companies, or organizations at the time of the survey, ranging from 2002 through 2009.	World Bank's Enterprise Surveys. https://www.enterprisesurveys.org/

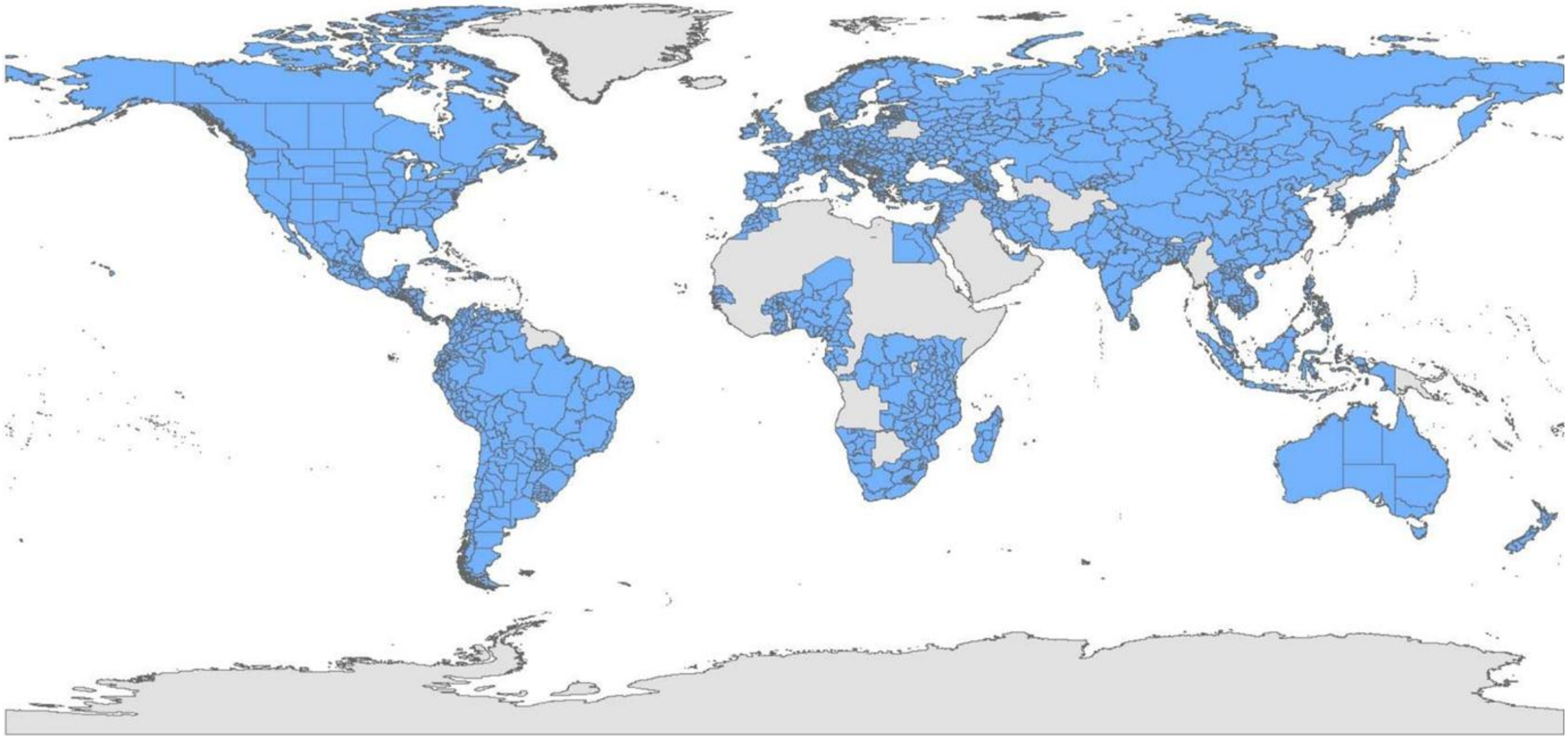

Figure 1. Countries shaded in blue are included in our sample.

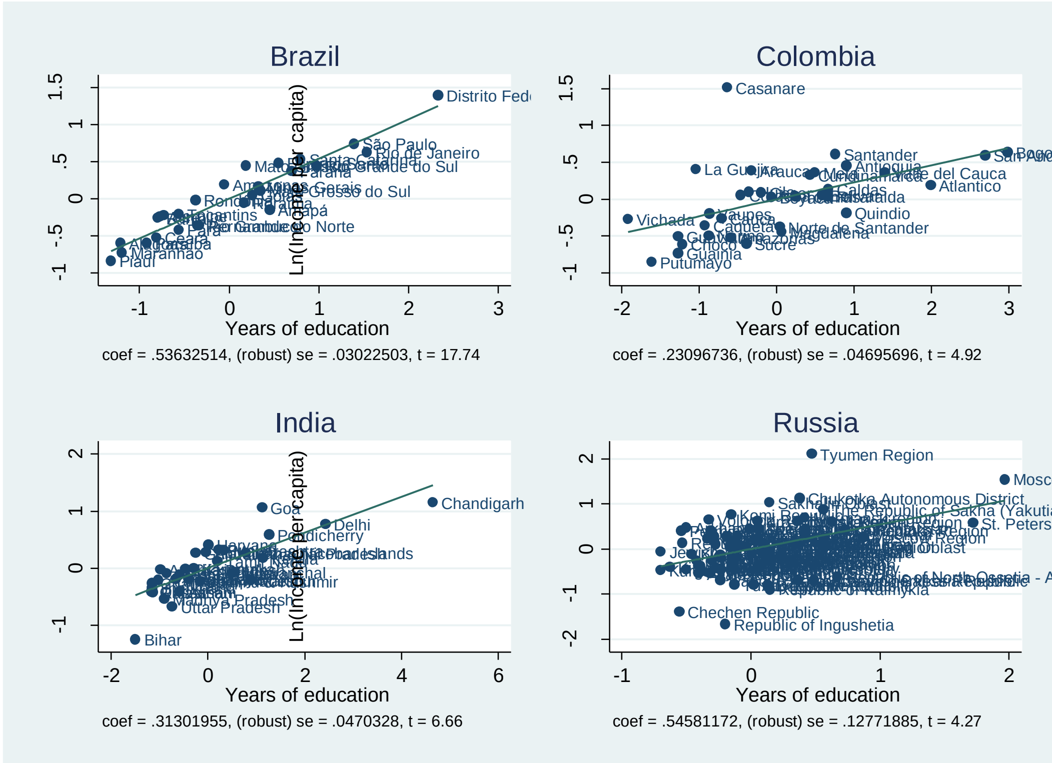

Figure 2. Partial correlation plot of (log) Regional income per capita and Years of education in Brazil (top left), Colombia (top right), India (bottom left), and Russian Federation (bottom left).

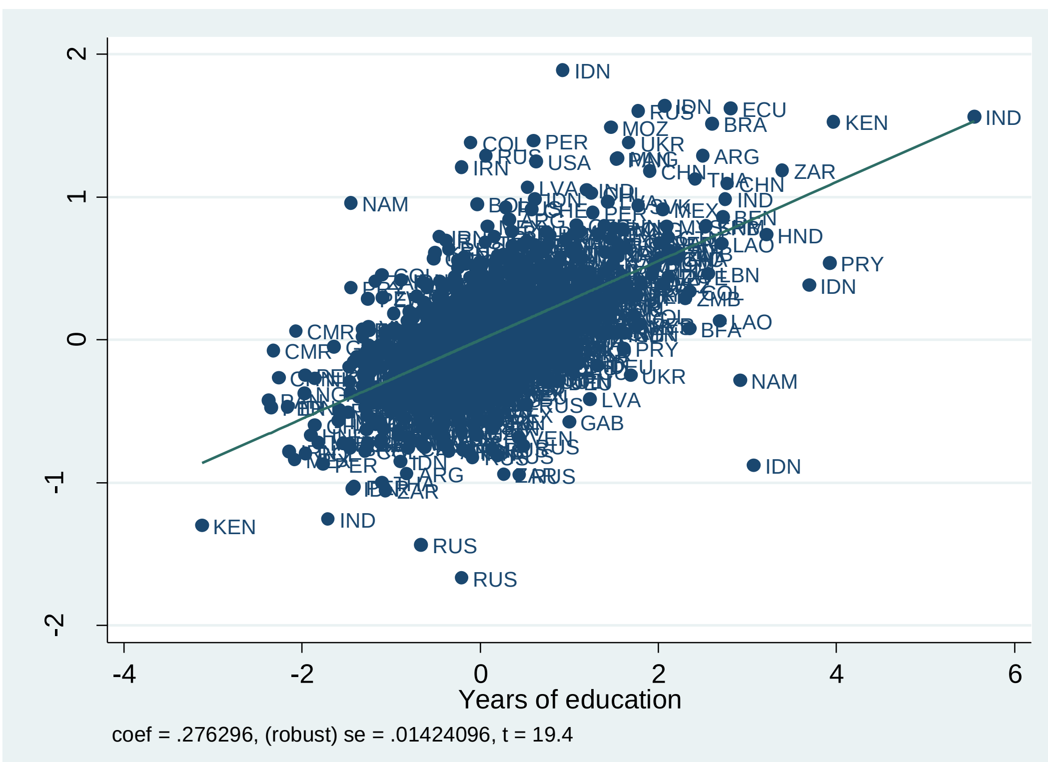

Figure 3. Partial correlation plot of (log) Regional income per capita and Years of education controlling for temperature, distance to coast, oil, population, and country dummies.