NBER WORKING PAPER SERIES

HUMAN CAPITAL AND REGIONAL DEVELOPMENT

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

Working Paper 17158
http://www.nber.org/papers/w17158

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2011

We are grateful to Nicolas Ciarcia, Nicholas Coleman, Sonia Jaffe, Konstantin Kosenko, Francisco Queiro, and Nicolas Santoni for dedicated research assistance over the past 4 years. We thank Gary Becker, Edward Glaeser, Gita Gopinath, Josh Gottlieb, Elhanan Helpman, Chang-Tai Hsieh, Mathew Kahn, Pete Klenow, Casey Mulligan, Elias Papaioannou, Jacopo Ponticelli, Giacomo Ponzetto, Jesse Shapiro, Chad Syverson, and seminar participants at the UCLA Anderson School for extremely helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Human Capital and Regional Development
Nicola Gennaioli, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer
NBER Working Paper No. 17158
June 2011
JEL No. L26,O11,O43,O47,R11

ABSTRACT

We investigate the determinants of regional development using a newly constructed database of 1569 sub-national regions from 110 countries covering 74 percent of the world's surface and 96 percent of its GDP. We combine the cross-regional analysis of geographic, institutional, cultural, and human capital determinants of regional development with an examination of 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 human capital externalities are essential for understanding the data.

Nicola Gennaioli
CREI
Universitat Pompeu Fabra
Ramon Trias Fargas 25-27
08005 Barcelona (Spain)
ngennaioli@crei.cat

Rafael La Porta
Dartmouth College
Tuck School
210 Tuck Hall
Hanover, NH 03755
and NBER
rafael.laporta@dartmouth.edu

Florencio Lopez-de-Silanes
EDHEC Business School
393, Promenade des Anglais BP 3116
06202 Nice Cedex 3
FRANCE
and NBER
Florencio.lopezdesilanes@edhec.edu

Andrei Shleifer
Department of Economics
Harvard University
Littauer Center M-9
Cambridge, MA 02138
and NBER
ashleifer@harvard.edu

I. Introduction

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

To organize this discussion, we present a new framework describing the channels through which human capital influences productivity, which combines three features. First, and most importantly, human capital of workers enters as an input into the neoclassical production function, but the 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 von 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). These externalities result from people in a given location spontaneously interacting with and learning from each other, so knowledge is transmitted across people without being paid for. Because our framework incorporates both the allocation of talent between entrepreneurship and work as in Lucas (1978), and human capital externalities as in Lucas (1988), we call it the Lucas-Lucas model[2]. Human capital externalities have been shown to be important in a variety of development and regional contexts (Rauch 1993, Glaeser, Scheinkman and Shleifer 1995, Angrist and Acemoglu 2000, Glaeser and Mare 2001, Moretti 2004, Iranzo and Peri 2009), although Ciccone and Peri (2006) and Caselli (2005) find them to be unimportant. By decomposing human capital effects into those of worker education, entrepreneurial/managerial education, and externalities using a unified framework, we try to disentangle different mechanisms.

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

To begin, we use regional data to examine the determinants of regional income in a specification with country fixed effects. The approach follows development accounting, as in Hall and Jones (1999), Caselli (2005), and Hsieh and Klenow (2010). Among the determinants of regional

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

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), DeLong 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, for which we have data at the regional level and which may matter at the national level (Knack and Keefer 1997), as well as ethnic heterogeneity (Easterly and Levine 1997, Alesina et al. 2003). Last, we consider the effect of average education in the region on its level of development. A substantial cross-country literature points to a large role of education. Barro (1991) and Mankiw, Romer, and Weil (1992) are two early empirical studies; de La Fuente and Domenech (2006) and Cohen and Soto (2007) are recent confirmations. Across countries, the effects of education and institutions are difficult to disentangle empirically (Glaeser et al 2004).

We find that favorable geography, such as lower average temperature and proximity to the ocean, as well as higher natural resource endowments, are associated with higher per capita income in regions within countries. We do not find that culture, as measured by ethnic heterogeneity or trust, explains regional differences. Nor do we find that institutions as measured by survey assessments of the business environment in the 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 calls 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 that half of the differences in per capita incomes across countries is attributable to education. We confirm the large role of education, but try to go further in identifying the channels. Acemoglu and Dell also conjecture that institutions shape the remainder of the local income differences. We do have regional data on several aspects of institutional quality, but find that their ability to explain cross-regional differences is minimal[3].

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

Most importantly, the simultaneous use of regional and firm-level data enables us to explore the effects of human capital by combining estimation with calibration. Because education is endogenous in national and regional regressions, scholars have turned to calibration techniques, using Mincerian estimates of private returns to education, to compute the parameters of the production function (e.g., Caselli 2005). We rely on previous research regarding factor shares (e.g., Gollin 2002, Caselli and Feyer 2007, Valentyi and Herrendorf 2008), but combine it with coefficient estimates from productivity

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

regressions to calculate the parameters of our production function. We find substantial consistency between regional and firm-level results, as well as plausible estimates of the parameters of the production function.

Specifically, the micro data show that establishments with employees and managers with higher education are more productive, holding capital and energy inputs constant and including region-industry fixed effects. These data establish quite clearly the huge role of managerial/entrepreneurial human capital in raising firm productivity. When regional education is added to these regressions, it has a large and positive coefficient. Of course, regional education may be correlated with omitted region-specific productivity parameters, so we do not have perfect identification. We try to control for regional characteristics, but also bring information from other studies to calibrate the parameters of the production function consistent with our estimates. Our calibrations show that worker education, entrepreneurial education, and externalities all contribute to productivity. Crucially, while we find the role of workers' human capital to be in line with standard calibration exercises (e.g., Caselli 2005), our results indicate that focusing on worker education alone enormously underestimate both private and social returns to education. Private returns are extremely high but to a substantial extent are earned by entrepreneurs, and hence might appear as profits rather than wages, consistent with Lucas (1978). Although we have less confidence in the findings for externalities, our best estimates suggest that those are also substantial, in line with Lucas (1988, 2008). In sum, the evidence points to a large influence of entrepreneurial human capital, and perhaps of human capital externalities, on productivity.

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

In the next section, we present our model of regional development that organizes the evidence. In section III, we describe our data in some detail. Section IV presents the evidence on the role of various factors in accounting for the differences in both national and regional development. Section V presents firm-level evidence to disentangle the channels of through which human capital influences productivity. We combine the regional and the micro estimates to assess the parameters of the model. Section VI presents some of the census evidence to test additional implications of the model. 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 A_G and a share $1-p$ of which has productivity $A_B < A_G$. We refer to the former regions as "productive", to the latter regions as "unproductive", and denote them by $i = G, B$. A measure 2 of agents is uniformly distributed across regions. An agent j enjoys consumption and housing according to the utility function:

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

where c and a denote consumption and housing, respectively. Half the agents are "rentiers," the remaining half are "labourers''. Each rentier 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 $i = G, B$ the distribution of human capital is Pareto in $[\underline{h},+\infty)$, with mean $\mu_i\underline{h}/(\mu_i-1)$, where μ_i, $\underline{h} > 1$. We denote by $\underline{H}_i = \mu_i\underline{h}/(\mu-1)$ the initial human capital endowment of region $i = G, B$. Differences in $\underline{H}_i$ capture exogenous variation of human capital across regions.

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 first consider the Lucas-only model where A_i captures exogenous regional differences, such as institutions or geography. We then extend the analysis to the Lucas-Lucas model, where A_i depends also on regional human capital due to the presence of externalities. Either way, productivity A_i induces human and physical capital to sort across regions.

Rentiers rent land and physical capital to firms, 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.

[4] For simplicity, we assume that moving costs are a redistribution from migrants to locals (the latter may be viewed as providing 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.

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 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. Throughout the analysis, the price of consumption is normalized to one.

The Lucas-only model: production and occupational choice

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

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

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

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

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

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

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

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

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

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

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

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

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

Equation (7) does not say whether entrepreneurial human capital is allocated to few firms run by very skilled entrepreneurs or to many firms run by less skilled ones. Assuming a fixed income cost of setting up a firm would pin down the number of firms: now the most skilled labourers become entrepreneurs, as in Lucas (1978). In this case, though, a wedge between profits and wages would arise to compensate the marginal entrepreneur for bearing the fixed cost. This wedge complicates regional mobility by creating non trivial choices between becoming an entrepreneur in one region versus a worker in another. To simplify the analysis, we focus on the limiting case where the cost of setting up a firm tends to zero. In this case, there is no wedge between profits and wages, so that moving decisions only depend on regional wages, but it is still the case that the most skilled labourers become entrepreneurs. With density $F_i(h)$ of human capital in region i, there is a threshold h_i^E defined as:

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

In region i, labourers become entrepreneurs if and only if $h \geq h_i^E$, and the number of firms is $1 - F_i(h_i^E)$, where the density $F_i(h)$ and the quantity H_i of human capital in the region are endogenously determined by labour mobility, which we study next. The number of firms is irrelevant for most of our results, and becomes relevant only for the predictions of Proposition 3.

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

We consider symmetric spatial equilibria in which all productive regions share the same factor allocation (H_G, K_G), the same wage w_G and rental rates λ_G and η_G, and unproductive regions share the same allocation (H_B, K_B), wage w_B, and rentals λ_B and η_B. The uniformity of the capital rental ρ across regions pins down the allocation of physical capital as a function of the regional allocation of human capital. By exploiting Equations (7) and (6) one finds that ρ is constant across regions provided:

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

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

To find 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 regions is $\theta\cdot w_iH_i/\eta_i$. With the unitary supply of housing, the housing rental rate is equal to:

$$\eta_i = \theta\cdot w_i\cdot H_i \,. \tag{10}$$

As a consequence, the utility (gross of moving costs) of a labourer in region *i* is equal to:

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

which rises with the wage and falls with regional human capital H_i due to higher rents. We assume:

A.1 $\underline{H}_G/\underline{H}_B < (A_G/A_B)^{1/(\delta+\beta)}$,

which bounds the relative abundance of human capital in productive regions so that the autarky wage rate is higher there and thus both capital and labour tend to move there. We can establish:

Proposition 1 In equilibrium, the human capital allocation H_G and H_B has the following features:

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 $\underline{H} \equiv p\underline{H}_G + (1-p)\underline{H}_B$ the aggregate human capital. Then, when $\varphi = 0$ the equilibrium allocation satisfies:

$$H_G = H_G^{free} \equiv \frac{A_G^{\frac{1-\theta}{\beta(1-\theta)+\theta(1-\delta)}}}{E\left[A^{\frac{1-\theta}{\beta(1-\theta)+\theta(1-\delta)}}\right]} \cdot \underline{H} \cdot \tag{17}$$

When $\varphi > 0$, we have that $H_G < H_G^{free}$ and H_G increases in $\underline{H}_G$ holding $\underline{H}$ constant.

Since wages (and profits) are higher in the productive than in the unproductive regions, migration occurs 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. In equilibrium, wages are higher in the more productive regions, $w_G > w_B$, but the housing rental rate is also higher there, $\eta_G > \eta_B$. Critically, while (17) shows that under free migration the human capital employed in a region only depends on that region's productivity, when mobility is imperfect (i.e. $\varphi > 0$) a region exogenously endowed with more human capital will employ more human capital in equilibrium. This property will be important for the empirical analysis.

More generally, in our model productive regions attract both human and physical capital. In the proof of Proposition 1 we show that under perfect mobility this implies that national output is equal to:

$$Y = \hat{A}\left(H^E\right)^{1-\alpha-\beta-\delta}\left(H^W\right)^{\alpha} K^{\delta} T^{\beta}, \tag{18}$$

where $\hat{A}$ is a function $\hat{A}(\beta, \delta, \theta, A_G, A_B, p)$ of exogenous parameters.

The Lucas-Lucas spatial equilibrium

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

$$\tilde{A}_i = A_i\left(E_i(h)^{\psi} L_i\right)^{\gamma}, \quad \gamma > 0,\ \psi \geq 1, \tag{19}$$

where $E_i(h)$ is the average level of human capital in region *i* and L_i is the measure of labour in that region. Productivity depends not only on exogenous regional conditions but also on the sorting of human capital. Parameter ψ captures the importance of the quality of human capital: when $\psi = 1$ only the total quantity of human capital $H_i = \mathrm{E}_i(h)L_\mathrm{i}$ matters for externalities; as ψ becomes larger the quality of human capital becomes relatively more important than quantity. Parameter γ captures the overall importance of externalities. In our regional and firm level regressions, we employ the flexible specification of Equation (20), which the appendix shows to yield the following result:

Proposition 2 With human capital externalities, under the parametric restriction:

$$(\beta - \psi\gamma)(1 - \theta) + \theta(1 - \delta) > 0, \tag{20}$$

there is a stable equilibrium allocation H_G and H_B. 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) 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:

$$H_G = \tilde{H}_G^{free} \equiv \frac{A_G^{\frac{1-\theta}{(\beta-\gamma)(1-\theta)+\theta(1-\delta)}}}{E\left[A^{\frac{1-\theta}{(\beta-\gamma)(1-\theta)+\theta(1-\delta)}}\right]} \cdot \underline{H} \cdot \tag{21}$$

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

The main change introduced by externalities is that now the effect of fixed supplies of land and housing on hindering mobility are moderated by regional externalities. In fact, for migration to be interior condition (20) 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. When ψ = 1 and φ = 0, national output is equal to:

$$Y = \widehat{A} H^{\gamma} \left(H^{E}\right)^{1-\alpha-\beta-\delta} \left(H^{W}\right)^{\alpha} K^{\delta} T^{\beta} \ , \tag{22}$$

where $\widehat{A}$ is a function $\widehat{A}(\beta,\delta,\theta,A_G,A_B,p,\psi,\gamma)$ of exogenous parameters. More generally, under condition (20) the Lucas-Lucas model yields the following equation for firm level output:

$$y_{i,j} = A_i E_i(h)^{\psi\gamma} L_i^{\gamma} h_j^{1-\alpha-\beta-\delta} H_{i,j}^{\alpha} K_{i,j}^{\delta} T_{i,j}^{\beta} \ , \tag{23}$$

and the following equation for regional output:

$$Y_i = A_i E_i(h)^{\psi\gamma} L_i^{\gamma} (H_i^{E})^{1-\alpha-\beta-\delta} (H_i^{W})^{\alpha} K_i^{\delta} T^{\beta} \ . \tag{24}$$

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:

$$h_j = \exp\left(\mu_j S_j\right), \tag{25}$$

where $S_j \geq 0$ and $\mu_j \geq 0$ are two random variables (distributed to ensure that the distribution of h_j is Pareto). The return to schooling μ_i 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. Human capital in region *i* is then equal to $\int_{\underline{h}}^{+\infty} h dF_i(h) = \iint_{S\geq 0,\mu\geq 0} e^{\mu S} g_i(S,\mu) dS d\mu$, where $dF_i(h)$ is the density of region *i* labourers with human capital *h* and $g_i(S,\mu)$ is the density of region *i* labourers having human capital $h = e^{\mu S}$, so that $\iint_{S\geq 0,\mu\geq 0} g_i(S,\mu) dS d\mu = L_i$. 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:

$$E(h_i) \cong e^{\overline{\mu}_i \cdot \overline{S}_i} \quad , \tag{26}$$

where $\overline{S}_i$ is average schooling while $\overline{\mu}_i$ is the average Mincerian return, both computed in region *i*.

Regional Income Differences

To test Equation (24) we need to specify a regression in terms of observables, which entails regressing regional per capita income on human capital and population (we do not have regional data on physical capital). From the Equation for ρ in (6) we obtain the condition $K_i = B A_i^{\frac{1}{1-\delta}} H_i^{\frac{1-\beta-\delta}{1-\beta}}$ where B > 0 is a constant. By substituting this condition as well as Equation (26) into Equation (24) we find that:

$$\ln(Y_i/L_i) = C + [1/(1-\delta)]\ln A_i + [1+\gamma\psi - \beta/(1-\delta)]\, \overline{\mu}_i\, \overline{S}_i + [\gamma - \beta/(1-\delta)]\ln L_i, \tag{27}$$

where C is a constant absorbed by the country fixed effect. Estimating (27) using OLS implies that the coefficient on average regional schooling should be interpreted as the product of the "technological" parameter (1+ $\gamma\psi - \beta$) and the nation-wide average $\overline{\mu}$ of the regional Mincerian returns $\overline{\mu}_i$.

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 $(1+ \gamma\psi - \beta)$. The estimation of Equation (27) raises a serious concern: since in our model human capital migrates to more productive regions, any mismeasurement in 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 (27) 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 (27).

Firm-Level Productivity

In (23), 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 (23) 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:

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

$$(1-\alpha-\beta-\delta)\ln(l^{E}_{i,j}/l_{i,j}) + \alpha\ln(l^{W}_{i,j}/l_{i,j}) + \delta\ln k_{i,j} + \beta\ln t_{i,j} + \gamma\ln L_i + \gamma\psi\,\overline{\mu}_i\,\overline{S}_i\,, \qquad (28)$$

where $x_{i,j} = X_{i,j}/l_{i,j}$ denote 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 total employment on managerial and non-managerial jobs, respectively.

When estimating (28) using OLS, the coefficient on the entrepreneur's schooling should be interpreted as the product of entrepreneurs' rents $(1-\alpha-\beta-\delta)$ and a nation-wide average Mincerian entrepreneurs' return to education $\overline{\mu}_E$. The coefficient on worker's average schooling should be interpreted as the labour share α times a nation-wide average Mincerian returns to workers $\overline{\mu}_W$. The coefficient on regional schooling should be interpreted as the product of the externality parameter $\gamma\psi$ and the population-wide average Mincerian return $\overline{\mu}$.[5]

The estimation of (28) allows us to separate the role of the "low human capital" of workers from the "high human capital" of entrepreneurs in shaping firm productivity. Since the selection of talented entrepreneurs into more productive regions/industries may contaminate our results, we first estimate (28) by controlling for the full set of region-industry dummies. In addition, we estimate (28) by directly controlling for region specific variables to assess the importance of externalities from the coefficient γ on population and the coefficient $\gamma\psi\overline{\mu}$ on regional schooling. In this case, however, migration of human capital to more productive regions may give the false impression of positive externalities, creating the identification issue present in estimating (27). We deal with this problem by controlling for proxies for A_i and by comparing our estimation results to standard calibrations of externalities.

[5] Both the regional level Equation (27) and the firm level Equation (28) imply that the average return to human capital should vary across regions. One way to empirically account for such possibility is to run random coefficient regressions. We have performed this analysis and the results change very little (the results on human capital become slightly stronger). We do not report the results to save space.

In our model the human capital of more productive regions is of better quality, namely $E_G(h) > E_B(h)$, because migrants are more skilled than average.[6] The black bold line in Figure 2 below is the skill distribution in the unproductive region, which is truncated at h_m. The skill distribution in the productive region coincides with the black bold line until h_m and jumps to the red line for $h > h_m$.

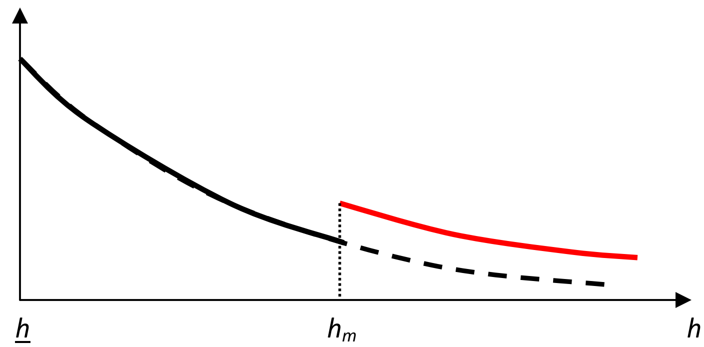

Figure 2: *regional distributions of human capital*

Since in productive regions the right tail of the skill distribution is fatter, in these regions there are likely to be fewer but higher skilled entrepreneurs running larger firms than in unproductive regions. The result below (proved in the Appendix) identifies one set of conditions in which this indeed true:

Proposition 3 Consider a full mobility equilibrium for ψ = 1 and $\mu_G = \mu_B$. If p is sufficiently large, there are two thresholds z_1 and z_2 such that for $(A_G / A_B)^{\frac{1-\theta}{(\beta-\gamma)(1-\theta)+\theta(1-\delta)}} \in (z_1, z_2)$ we have $l_G/f_G > l_B/f_B$. That is, productive regions have: i) a larger average firm, ii) a larger share of workers in the population.

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

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

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

[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, travel time, and soil characteristics by first summing the (average) values of the relevant variable for all grid cells lying within a region and then dividing by the number of cells lying within a region. We computed regional averages for the density (e.g., power lines) and natural resources variables (oil and gas) by first summing the relevant variable for all grid cells within a region and then dividing by the region's population. We averaged the responses within a region for all the variables from the Enterprise and World Value Surveys. We sum up the number of unique ethnic groups and computed the probability that people within a region speak the same language based on the total number of people in each "language" area.

have GDP data for 27 first-level administrative regions in Brazil, corresponding to its 26 states plus the Federal District, but survey data on institutions for 248 municipalities. For our empirical analysis, we aggregate the data on institutions by taking the simple average of all observations for establishments located in the same first-level administrative division. Similarly, we aggregate the GIS geo-spatial data on geography, climate, and infrastructure at the first-administrative level using the Collins-Bartholomew World Digital Map.

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

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

Our final dataset has regional income data for 107 countries in 2005, drawn from sources including National Statistics Offices and other government agencies (42 countries), Human Development Reports (36 countries), OECDStats (26 countries), the World Bank Living Standards Measurement Survey (Ghana and Kazhakstan), and IPUMS (Israel). When regional income data for 2005 is missing, we use log-linear interpolation based on as much data as it is available for the period 1990-2008 or, when interpolation is not possible, the closest available year.[9] Our measure of regional income per capita is typically based on value added but we use data on income (6 countries), expenditure (8 countries), wages (3 countries), gross value added (2 countries), and consumption, investment and government expenditure (1 country) to fill-in missing values. We measure regional GDP in current purchasing-power-parity dollars as we lack data on regional price indexes. To ensure consistency with the national GDP figures reported by World Development Indicators, we adjust regional income values so that --when weighted by population-- they total GDP at the country level. Not surprisingly, adjustments exceeding 20% were necessary in 19 out of the 20 countries for which use GDP proxies rather than actual GDP. Adjustments exceeding 20% were also necessary in 13 countries (Democratic Republic of Congo, Lebanon, Lesotho, Madagascar, Malaysia, Nepal, Niger, Philippines, Senegal, Swaziland, Syria, Uganda, and Venezuela) where our GDP data are in real terms.

We compute regional income per capita using population data from *Thomas Brinkhoff: City Population*, which collects official census data as well as population estimates for regions where official

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

census data are unavailable.[10] We adjust these regional population values so that their sum matches the country's population in the World Development Indicators database. This adjustment exceeds 10% for 6 countries: Bangladesh (+13%), Benin (+11%), Democratic Republic of Congo (-10.5%), Gabon (-25%), Swaziland (+16%) and Uzbekistan (-22%).

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

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

[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.

data, the Enterprise Survey data only covers registered establishments. A limitation of the Enterprise Survey data is that it largely excludes OECD countries (Ireland and Mexico are the exceptions).

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

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

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

[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.

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

secondary level. For each region, we compute average years of schooling as the weighted sum of the years of school required to achieve each educational level, where the weights are the fraction of the population aged 15 an older that has completed each level of education.

To illustrate these calculations consider the Mexican state of Chihuahua. The EPDC data on the highest educational attainment of the population 15 years and older in Chihuahua in 2005 shows that 4.99% of the that population had no schooling, 13.76% had incomplete primary school, 22.12% had complete primary school, 5.10% had incomplete lower secondary school, 23.04% had complete lower secondary school, 17.94% had complete upper secondary school, and 13.05% had complete tertiary school. Next, based on UNESCO's mapping of the national educational system of Mexico, we assign six years of schooling to people who have completed primary school and 12 years of schooling to those that have completed secondary school. Finally, we calculate the average years of schooling in 2005 in Chihuahua as the sum of: (1) six years times the fraction of people whose highest educational attainment level is complete primary school (22.12%), incomplete lower secondary (5.1%), or complete lower secondary school (23.04%); (2) 12 years times the fraction of people whose highest attainment level is complete upper secondary school (17.94%); and (3) 16 years times the fraction of people whose highest attainment level is complete tertiary school. Accordingly, we estimate that the average years of schooling of the population 15 and older in Chihuahua in 2005 is 7.26 years (=6*0.5026+12*0.1794+16*0.1305).

We compute years of schooling at the country-level by weighting the average years of schooling for each region by the fraction of the country's population 15 and older in that region. The correlation between this measure and the number of years of schooling for the population 15 years and older in Barro and Lee (2010) is 0.9. For the average (median) country in our sample, the number of years of schooling in Barro and Lee (2010) is 8.18 vs. 6.88 in ours (8.56 vs. 6.92 years). Two factors could potentially explain why the Barro-Lee dataset yields a higher level of educational attainment than ours: (1) Barro-Lee captures incomplete degrees while we do not; and (2) education levels have increased rapidly over time but some of our educational attainment data is stale (e.g. for 14 countries our educational attainment data is for the year 2000 or earlier). To make the Barro and Lee (2010) measure of educational attainment more comparable to ours, we make two adjustments to their data. First, we apply our methodology to the Barro-Lee dataset and compute the level of educational attainment in 2005. After this first adjustment, the level of educational attainment computed with the Barro-Lee dataset for the average (median) country in our sample drops to 7.07 (7.23). Second, we apply our methodology to the Barro-Lee dataset but –rather than use data for 2005 -- use figures for the year that best matches the year in our dataset. After this second adjustment, the level of educational attainment using the Barro-Lee dataset for the average (median) country in our dataset drop further to 6.95 (7.22).[14] Since most of our results are run with country-fixed effects, country-level biases in our measure of human capital do not affect our results.[15]

We gather data on seven measures of the quality of institutions from the Enterprise Survey and the Sub-national Doing Business Reports. The Enterprise Survey covers as many as 80 of the countries and 410 of the regions in our sample.[16] The Enterprise Survey asked business managers to quantify: (1)

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

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

[16] The main reason why we have fewer regions with measures of institutions than regions with productivity data is because we imposed a filter of a minimum of 10 establishments answering the particular institutions question. The

informal payments in the past year (percent of sales spent in informal payments by a typical firm in the respondent's industry), (2) the number of days spent in meeting with tax authorities in the past year, (3) the number of days without electricity in the previous year, and (4) security costs (cost of security equipment, personnel, or professional security service as a percentage of sales). The Enterprise Survey also asks managers to rate a variety of obstacles to doing business, including: (1) access to land, and (2) access to finance.[17] For each of these obstacles to doing business, we keep track of the percentage of the respondents that rate the item as a moderate, a major, or a very severe obstacle to business. The final Enterprise Survey variable that we examine is the perception of government predictability (measured as the percentage of respondents who tend to agree, agree in most cases, or fully agree that government officials' interpretations of regulations are consistent and predictable).

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

---

rest of the discrepancy in the number of regions is because some questions about institutions were not included in the survey for some countries.

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

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

Cultural variables are the last set of potential determinants of regional income that we examine. We gather two proxies for cultural values and attitudes from the World Value Survey for as many as 75 of our sample countries and 745 of our regions.[19,20] The first survey measure is the percentage of respondents in each region that answer that "most people can be trusted" when asked whether "Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?" The second measure is a proxy for civic values based on whether each of the following behaviors "can always be justified, never be justified or something in between.": (a) "claiming government benefits which you are not entitled to"; (b) "avoiding a fare on public transport"; (c) "cheating on taxes if you have the chance"; and (d) "someone taking a bribe" (Knack and Keefer, 1997).[21] In addition, we gather two measures of fractionalization for up to 1,568 regions and 110 of our sample

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

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

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

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

countries. The first one is simply the number of ethnic groups that inhabited each region in 1964. The second one is the probability that a randomly chosen person in a region shares the same mother language with a randomly chosen people from the rest of the country in 2004.

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

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

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

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

IV. Accounting for National and Regional Productivity.

In this section, we present cross-country and cross-region evidence on the determinants of productivity. We present national regressions only for comparison. These regressions are difficult to interpret in our model because it is not possible to express national output in closed form. More importantly, the problem of endogeneity of education is particularly severe in the national context, which of course turned some scholars to calibration. In Section V, we interpret the coefficients in the context of regional and firm level regressions. With respect to regional income, our benchmark is Equation (27). As already mentioned, we have measures of average education at the regional level, but we do not have either national or regional data on physical capital (except for public infrastructure) or other inputs, so these variables only appear in the firm-level regressions in Section V.

Table 3 presents our basic regional results in perhaps the most transparent way. The table reports the results of univariate regressions of regional GDP per capita on its possible determinants, all with country fixed effects. Such specifications are loaded in favor of each variable seeming important since it does not need to compete with any other variable. We report both the within country and between countries R^2 of these regressions. The first row presents the main result: education explains 58% of between country variation of per capita income, and 38% of within country variation of per capita income. Although several other variables explain a significant share of between country variation, none comes close to education in explaining within country variation in income per capita.

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

Tables 4 through 6 show the multivariate regression results at the national and regional level. Table 4 presents the baseline regressions of national (Panel A) and regional (Panel B) per capita income on geography and education, controlling in some instances for population or employment, as suggested by our model. At the country level, temperature, inverse distance to coast, and oil endowment are all highly statistically significant in explaining cross-country variation in incomes, and explain an impressive 50% of the variance. Education is also statistically significant, with a coefficient of .25, raising the R^2 to 63%. Note that oil comes in positive and highly statistically significant.

As Panel B shows, the coefficients on geography and education continue to be significant at the regional level. However, the within country R^2 is much higher for education than for the geographic variables. The coefficient on the regional labour force is now positive and statistically significant, and ranges from .01 for population to .07 for employment. The coefficient on education is around .27.

Table 5 presents country-level regressions with measures of institutions (Panel A) and of infrastructure and culture (Panel B) added to the specification in Table 4. Education remains highly statistically significant in each specification, and its coefficient does not fall much. At the country level, only the logarithm of tax days is statistically significant. The last two rows of Table 5 show the adjusted R^2 of each regression if we omit the institutional (or infrastructure or cultural) variable, as well as the adjusted R^2 if we omit education. Dropping education sharply reduces explanatory power, while the only institutional variable that adds explanatory power is the logarithm of tax days.

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

As a robustness check, we have rerun but do not present here all the regressions breaking down our aggregated educational attainment measure into measures, for each level of educational attainment between 1 and 13 years, of the presence of individuals at that level of education in that region. Educational variables continue to be highly correlated with per capita income, and the coefficients increase monotonically with the level of attainment. That is, having more people at a higher level of education is associated with higher income.

What are some of the possible explanations of the low explanatory power of institutions, keeping in mind that endogeneity of institutions should if anything raise the coefficients? It is possible that we have inappropriate measures of institutions, although the measures we have are commonly considered to be relevant to economic outcomes. It is also possible that the measures from Enterprise Surveys are particularly noisy, although one should remember that these are surveys of managers who should be particularly focused on institutional constraints. In general, such subjective assessments correlate much better with measures of development than objective measures of institutions (Glaeser

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

et al. 2004). It is also likely that at least some institutions only matter at the national level, if, for example, the critical-to-development business activity is concentrated in the capital.

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

We have previously indicated that, 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. To address this problem, we next present the micro evidence based on Enterprise Surveys and combine it with calibration results to interpret the regional and firm-level coefficients in a unified framework.

V. Firm-Level Evidence.

In Tables 8-10, we turn to the micro evidence and estimate essentially Equation (28). We use the Enterprise Survey data described in Section III. In establishment-level regressions, we can try to disentangle entrepreneurial human capital, worker human capital, and human capital externalities, as well as to some extent control for regional effects. To address the concern that the sorting of entrepreneurs by unobservable skills into more productive regions may contaminate our firm-level estimates of the returns to schooling, in Table 8 we use an extremely flexible specification that controls for region-industry interactions by including region-industry fixed effects. This enables us to focus on the effect of managerial/entrepreneurial education on productivity without worrying about disentangling human capital externalities and regional productivity factors, since both are subsumed in the fixed effects. In Tables 9 and 10 we then turn to an examination of human capital externalities, first without regional controls and then adding those controls to the regressions.

We use three dependent variables to proxy for productivity. First, we look at the log of the establishment sales per worker, $y_{i,j}/l_{i,j}$. Second, we look at a rough measure of value added, namely the logarithm of sales net out raw material inputs, per worker. Third, we run regressions with the log of average wages paid by the establishment (which in our Cobb-Douglas production function correspond to a constant fraction of output) as a dependent variable.[23] We measure capital (which includes both land $t_{i,j}$ and physical capital $k_{i,j}$) by the log of property, plant and equipment per employee. As an alternative, we proxy for capital by the log of expenditure on energy per employee. We also use the log of the number of employees, which is a proxy of $l_{i,j}$, to control for the share of the entrepreneur's labour $l_{i,j}^{E}$

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

$/l_{i,j}$ and of the workers' labour $l_{i,j}^{W}/l_{i,j}$. Indeed, assuming that each firm has only one entrepreneur we have $l_{i,j}^{E}/l_{i,j} = 1/l_{i,j}$ and $l_{i,j}^{W}/l_{i,j} = (l_{i,j}-1)/l_{i,j}$. Unfortunately, the regression coefficient of the log of employees is not susceptible of a clean interpretation in terms of technological parameters.

Most important, to trace out the effects of human capital, we include the years of schooling of the manager S_E and the years of schooling of workers S_W in Table 8, and subsequently the average years of schooling in the region S_i in Tables 9 and 10. As we explained in Section 2, the Mincer model of the relationship between education and human capital implies that schooling should enter the specification in levels, rather than in logs. Accordingly, the regression coefficients of the schooling variables should respectively capture parameters $(1-\alpha-\beta-\delta)\overline{\mu}_{E,i}$, $\alpha\overline{\mu}_{W,i}$ and $\gamma\psi\overline{\mu}_{i}$ in Equation (28). To capture scale effects in regional externalities, in Tables 9 and 10 we control for the log of the region's population L_i. The regression coefficient on this variable should capture γ in Equation (28). After presenting the basic estimation results, we compute values for these coefficients by combining estimation and calibration.

Estimation Results

While in Tables 8 we control for region-industry fixed effects, in Tables 9 and 10 we directly control for proxies of regional productivity A_i by including regional schooling and measures of geography, as well as country and industry fixed effects, using dummies for 16 industries. All standard errors are clustered at the regional level. We have experimented with some indicators of regional institutions and infrastructure as independent variables, but consistent with the findings for regional data, they are usually insignificant, and hence we do not focus on these results.

Begin with the results in Table 8, which includes region-industry fixed effects. In the (log) sales per employee specification, the coefficient on energy per employee is .34, while that on capital per employee is .30. These coefficients, however, are closer to .3 in the value added specification, and to .2 in the average wage specifications. When both variables are included in the specification, their sum is higher. Based on these estimates, we will use .35 as capital share when we calibrate the model and assess its ability to account for variation in productivity across space.

The coefficient on worker schooling averages to about .03 in the sales per employee specifications, roughly the same in the value added specifications, but is closer to .015 in the wage specifications. The coefficient on management schooling is also about .03 in the sales per employee specification, slightly lower in the valued added specification, but falls to about .02 in the wage specification. The coefficient on the log employment (firm size) is about .1 across specifications.

In Table 9, we include country and industry fixed effects, but add regional years of education to the regressions. There are some changes in parameter estimates, but the coefficients on worker education remain around .03, those on manager education likewise remain around .03, and capital shares stay around .35, like in the specifications of Table 8. In addition, we find consistent evidence of large effects on productivity from regional factors. The coefficient on regional schooling is amazingly consistent and statistically significant across specifications, and varies between .05 and .1. The coefficient on regional population varies across specifications, but we will take it to be around .09 based on the results for sales per employee and value added per employee.

In our analysis of determinants of regional productivity, geographic variables, but not measures of culture or institutions, have been consistently statistically significant. Accordingly, in Table 10 we examine the robustness of the results in Table 9 by controlling for the important geographic factors. Such controls might also go some way toward enabling us to attribute the coefficient on regional human capital to externalities rather than omitted regional productivity factors. The coefficients on geography variables are quite unstable in these specifications, with inverse distance to coast exerting a large positive influence on productivity in four specifications (but not the other eleven), and oil endowment exerting now a large negative effect in some specifications. The most obvious measures of omitted regional productivity thus do not appear to be important. Critically, the coefficients on years of education of managers and years of education of workers do not fall much relative to the specifications in Table 8, indicating that returns to education of entrepreneurs remain high even with controls. The coefficient on years of education in the region falls a bit in some specifications relative to its value in Table 9. We will use the average estimate of .05 in our calculations.

We added additional controls to these regressions, and obtained similar results. This evidence needs to be explored further, but most of the specifications confirm both the general findings, and parameter estimates, computed from Tables 8 and 9. There does not appear to be much evidence of significant omitted regional effects, although since we do not have a complete set of determinants of regional productivity, our assessment of external effects might be exaggerated.

Combining Estimation with Calibration

So what do these coefficients mean in light of our model, and how do they fit with the work in development accounting? Can the effects estimated from firm level regressions account for the important role of schooling in the regional regressions?

We address these questions by starting with the roles of managers' and workers' schooling. The coefficient on workers' average schooling in the firm level regressions is about 0.03, which in our model implies $\alpha \bar{\mu}_W = 0.03$. If we take the standard calibration for the U.S. labour share $\alpha = 0.6$, we back out an average Mincerian $\overline{\mu}_W$ return of workers equal to $\overline{\mu}_W$ = 0.05, which is slightly lower than the ballpark 0.06-0.1 of micro evidence on workers' Mincerian returns (Psacharopoulos 1994). If we calibrate α = 0.55 to capture the fact that in developing countries the labour share tends to be lower than in the U.S. (also because part of labour income goes to self employment, Gollin 2002), we obtain $\overline{\mu}_W$ = 0.06, which is the value we stick to. The fact that Mincerian returns to education implied by our empirical evidence are consistent with existing research suggests that our firm level productivity regressions help reduce identification problems, at least as far as firm-level variables are concerned.

The regressions also point to an overall capital share (considering energy or equipment) roughly equal to 0.35. In our model, this captures the income share $\delta + \beta$ going to K and T which leaves – under constant returns – a share of 0.15 going to entrepreneurial rents. That is, $(1–\alpha–\beta–\delta)$ = (1 – 0.55 – 0.35) = 0.1. Since the estimated coefficient on managerial education roughly implies $(1– \alpha–\beta–\delta)\,\overline{\mu}_E$ = 0.03, our results are consistent with a Mincerian return $\overline{\mu}_E$ equal to 30% for entrepreneurs. This preliminary assessment suggests that a neglected but critical channel through which schooling and human capital affect productivity is via entrepreneurial inputs. Individuals selected into entrepreneurship appear to be vastly more talented than workers, driving up productivity. Of course, entrepreneurial talent may be more important than entrepreneurial schooling in explaining this finding. Our analysis cannot adequately address this issue (which would require better data and an endogenous determination of the connection between schooling and talent). Our analysis is, nevertheless, sufficient to identify a critical role of management and entrepreneurship in determining productivity.

The large returns to entrepreneurial education, compared to the modest returns to worker education, might explain the problem that the previous literature on development accounting has experienced with the Mincer regressions (Caselli 2005, Hsieh and Klenow 2010): the returns to the education of labor are indeed low unless a worker becomes an entrepreneur. Entrepreneurial returns might not be measured in surveys seeking to capture returns to education.

We can then use the estimates of Tables 9 and 10 to assess the magnitude of human capital externalities. The coefficient on population in Table 9, roughly equal to 0.09, suggests that γ is also about 0.09. This assessment is consistent with the regional regressions in Table 4, where the coefficient on population is positive and roughly equal to 0.01, which implies that $\gamma - \beta/(1 - \delta) = 0.01$ in Equation (27). Given that $\gamma = 0.09$ and $\beta + \delta = 0.35$, this condition yields β to be roughly 0.06, which is in the ballpark of the land share estimated from income accounts (Valentiny and Herrendorf 2008). In sum, $\beta = 0.06$, $\delta = 0.29$ and $\gamma = 0.09$ are roughly consistent with both firm level and regional regressions.

The coefficient on regional schooling in the firm-level regressions of Table 9 is about .065. This implies (given $\gamma = 0.09$) that $\psi \overline{\mu}$ is about .72. To separate the effect of the population-wide Mincerian return $\overline{\mu}$ from the strength of the externalities ψ, we exploit the regional regressions. According to Equation (27) describing regional output, in these regressions the coefficient on schooling is equal to $[1+ \gamma\psi - \beta/(1 - \delta)]\overline{\mu}$. In Table 4, this coefficient is about 0.26. Since we have already established that $\gamma = 0.09$ and $\beta/(1 - \delta) = 0.08$ are reasonable estimates, we are left with two equations with two unknowns, namely $(1+ 0.09\psi - 0.08)\overline{\mu} = 0.26$ and $\psi \overline{\mu} = 0.72$. These equations imply an average population-wide Mincerian return $\overline{\mu}$ of about 0.21 (which is in between our estimates of workers' and entrepreneurs' values) and that the social return to schooling ψ of about 3.45.[24] These estimates point to a large effect of schooling for productivity via social interactions, consistent with Lucas (1985, 2008) as well as with the literature in urban economics cited in the introduction. Finally, note that at the above parameter values and at a reasonable share of housing consumption of θ = 0.4, the spatial equilibrium is stable, since $(\beta - \psi\gamma)(1 - \theta) + \theta(1 - \delta) = -(0.25)(0.6) + (0.4)(0.74) > 0$.

We can now use our results to assess the magnitude of the effect of schooling. Begin with the role of workers' and entrepreneurs' inputs. In regional regressions, the population-wide Mincerian return of 0.21 is needed to make sense of the data, while the firm level regressions suggest that the Mincerian return is 6% for workers and 30% for entrepreneurs. Although we lack direct data on the number of entrepreneurs in the economy, it is useful to make a back-of-the-envelope calculation to assess whether our firm level evidence is consistent with a 21% Mincerian return population-wide. It turns out that if: (1) an average entrepreneur is as educated as the entrepreneurs in the enterprise survey on average, i.e. entrepreneurs have 14 years of schooling; and that (2) an average worker in the economy is as educated as the average person in the sample, i.e. workers have roughly 7 years of schooling, to obtain an average population-wise Mincerian return of 21% entrepreneurs need to account for only 4.8% of the workforce.[25] Our estimates thus suggest that private returns to schooling may be much larger than what previously thought due to the neglected role of entrepreneurial inputs.

Consider now the role of externalities. Using our estimated parameters, raising the educational level from the sample mean of 6.58 years by one year can be calculated to increase regional TFP by about 6.7%. Rauch (1993) estimates a comparable magnitude of 3-5%. Acemoglu and Angrist (2000) estimate that a one year increase in average schooling is associated with a 7% increase in average wages. Moretti (2004) examines the impact of spillovers associated with the share of college graduates living in a city. If we run the regressions in Table 9 using the fraction of the population with college degrees instead of our measure of years of schooling, our estimates imply that a one percentage point increase in the share of region's population with a college degree increases output per capita by 7.9%.

[25] The population-wise average Mincerian return is computed as the return $\overline{\mu}$ that solves the equation exp($\overline{\mu}$ [$f \cdot 14 + (1-f)\cdot 7$]) = $f\exp(0.3\cdot 14) + (1-f)\exp(0.06\cdot 7)$ where f is the fraction of entrepreneurs.

Iranzo and Peri (2009) estimate that one extra year of college per worker increase the state's TFP by a very significant 6-9%, whereas the effect of an extra year of high school is closer to 0-1%. The agreement among the various estimates is quite striking. Even if we use the coefficients obtained in Table 10 when controlling for factors potentially affecting regional productivity, the change is very modest, increasing the required population-wide Mincerian return $\overline{\mu}$ to .23 and reducing the externality parameter ψ to 2.4. Notwithstanding the difficulties involved in the identification of externalities, the quantitative role of the latter seems to be quite robust.

As a final step, we assess the importance of our evidence on the returns to entrepreneurial inputs and human capital externalities in explaining cross-country differences in income per capita in the context of standard macro development accounting exercises. 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 level predicted by our model when: i) all regions in a country are identical and all countries are equally productive, and ii) where – in line with standard development accounting - we consider only physical and human capital, thereby attributing land rents to physical capital (deducting these rents would not change much our results). This simplified 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. Using Caselli's (2005) dataset, the observed variance of (log) GDP per worker is 1.32. Ignoring human capital externalities (i.e., assuming ψ=γ=0) and using the standard 8% average

Mincerian return on human capital for both workers and entrepreneurs (i.e. setting $\overline{\mu}$ = 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 development accounting 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 $\hat{Y}$ by assuming no human capital externalities (i.e., $\psi = \gamma = 0$) while still keeping an population-wide Mincerian return $\overline{\mu}$ of 21%, which is consistent with our firm-level estimates. Under these assumptions, *success* rises to 83%. This improvement is solely due to accounting for managerial schooling. Finally, to assess the incremental explanatory power of human capital externalities, we compute $\hat{Y}$ assuming our estimated values (i.e., ψ = 3.45 and γ = 0.09), while retaining the assumption that the average Mincerian return equals 21%. Under these new assumptions, s*uccess* rises to 99%. Of course, these results need to be interpreted cautiously since there is considerable uncertainty regarding the true values of the underlying coefficients. Nevertheless, these calculations illustrate the large role that entrepreneurial inputs may play in increasing the explanatory power of the factor-based model.

The comparison between Mozambique and the US illustrates the importance of entrepreneurial inputs to understand cross-country income differences. 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 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 11.6 higher ($H_{US}/H_{MOZ}=e^{.21*(12.69-1.01)}$) in the US than in Mozambique if the average Mincerian return is 21%. 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 worker should be 29 times higher in the US than in Mozambique ($29=11.6^{2/3}x185^{1/3}$), which is much closer to the actual value of 33 times than the 10.6 multiple implied by 8% Mincerian return ($10.6=2.5^{2/3}x185^{1/3}$).

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 are a fundamental and relatively neglected channel for understanding the role of schooling in shaping productivity differences, and iii) human capital externalities might also be extremely important determinants of regional productivity differences. Our parameter estimates point to very large returns to entrepreneurial schooling (perhaps due to entrepreneurs' general talent) and to large social returns at the regional level arising from education.

VI. Additional Implications.

The model has a number of additional implications. Specifically, it predicts that higher human capital regions within a country should have larger establishments, as well as a higher share of employment in population. As described in Section III, we have collected data from official censuses of establishments and population for 1, 068 regions from 69 countries in our sample. Before looking at the data, note that in our model these additional predictions are influenced by regional variation in the average firm size l_i/f_i. Since by Proposition 3 productive regions have larger firms on average and since they have a larger number of workers (i.e., $l_G > l_B$), they should also have a higher share of the workforce in total population $l_i/(f_i + l_i + r)$, where r is the measure of rentiers in productive and unproductive regions (recall we assumed $r = 1$, but here for clarity we keep it general). Accordingly, the number of establishments relative to the population would also be higher in productive regions provided these regions have a sufficiently higher number of firms than unproductive ones, i.e., if $(f_G - f_B) > (f_G f_B/r)(l_G/f_G - l_B/f_B)$. This latter condition however is not satisfied in the parameter region of Proposition 3.

Table 11 presents the results on the effects of human capital in the region (holding country fixed effect constant) on the number of official establishments relative to the population, two indicators of establishment size, and the number of formal employees relative to the population. It shows that higher human capital in the region is strongly associated with larger establishments, higher labor force participation rate, a larger share of employees working in large establishments, but also a higher number of establishments per person. Figure 7 presents the graphs of these relationships.

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

More broadly, these results suggest the possibility that one channel through which higher human capital raises regional income is by drawing more workers out of the unofficial sector (or agriculture) into the productive formal sector. The externality from social interactions might be larger in higher human capital regions because more of the firms and workers in such regions benefit from informational spillovers associated with human capital. These results are consistent with the predictions of the model, but also suggest that the analysis of human capital externalities, particularly in the developing countries, should take into account the massive role of informality.

VII. Conclusion.

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

A simple Cobb-Douglas production function specification used in development accounting would have difficulty accounting for all this evidence. Instead, we presented what we called a Lucas-Lucas model of an economy, which combines the allocation of talent between work and entrepreneurship, human capital externalities, and migration of labour across regions within a country. Although many issues remain to be resolved, the empirical findings we presented are both consistent with the general predictions of this model, and provide plausible values for the model's parameters. In addition, we follow Caselli (2005) in assessing the ability of the model to account for variation of output per worker across countries. When we use our Lucas-Lucas model, we can roughly double the ability of the model to account for cross-country income differences relative to the traditional specification. Our parameterization can explain 99% of income differences 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, but private returns to entrepreneurial education (in the form of profits) and possibly also social returns to education through external spillovers, are large. This evidence suggests that earlier estimates of return to education have perhaps underestimated one of its important benefits – the externalities, and largely missed the other – entrepreneurship. This final observation has significant implications for economic development.

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

Bibliography

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

Acemoglu, Daron, Simon Johnson, 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.

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

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

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

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

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

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

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

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

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.

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

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

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

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

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

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

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

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

Cohen, Daniel, and Marcelo Soto (2007). “Growth and Human Capital: Good Data, Good Results.” *Journal of Economics Growth*, 12 (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, Hedi Kallal, Jose Scheinkman, and Andrei Shleifer (1992). “Growth in Cities.” *Journal of Political Economy* 100(6):1126-1152.

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

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.

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." NBER Working Paper No. 15199.

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

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.

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

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.

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

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

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

Lucas, 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.

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

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

Murphy Kevin, 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.

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

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

Rauch, James (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.

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