Institutions, Geography, Regions, Countries and the Mobility Bias

Andrew Warner

CID Working Paper No. 91
April 2002

Working Papers
Center for International Development
at Harvard University

Institutions, Geography, Regions, Countries and the Mobility Bias

Andrew M. Warner[*]

June 1999
Revised April 2002

Abstract

Recent research on the ultimate causes of the large differences in economic development across countries has framed the issue as a competition between geography and institutions. Some papers claim to have evidence that geography explains nothing after controlling for institutions. This paper argues and shows evidence that geography correlates with economic activity across regions within countries where institutions are far more constant than across countries. The paper also argues that mobility of factors can in theory completely mask the impact of geography if one looks only at output per-capita. Comparing cross-region with cross-country data and comparing regressions of output density with regressions of output per-capita, the paper finds evidence of this mobility bias. Finally the paper tries to integrate recent lines of research with an earlier line that looked at regional data as a laboratory to test for economies of scale. Overall the paper finds evidence for three important determinants of spacial income levels: geography, institutions and agglomeration economies associated with very large urban areas. One needs to consider all three to fully explain the pattern of economic activity across space.

1 Introduction

This paper compares evidence on output density and population density across regions at the sub-national level and across nations in an attempt

[*]Center for International Development, Harvard University, 79 JFK Street, E403, Cambridge MA 02138, tel: 617-495-3628, E-mail: andrew_warner@harvard.edu. I thank John Gallup, Andrew Mellinger and Jeffrey Sachs for helpful discussions and Afsana Ahnkter for valuable research assistance.

to discriminate between a number of hypotheses about the determinants of the level of economic development. It is well known from casual observation or from Sachs and Warner (1997a) or Gallup, Sachs and Mellinger (1999) that geographic attributes of countries such as tropical climate or being landlocked correlate negatively with recent rates of economic growth. In a similar vein, but looking at levels rather than growth rates, Hall and Jones (1997) show that latitude (closely related to tropical climate) is negatively associated with the level of per-capita GDP across countries. More recently, Acemoglu Johnson and Robinson (2001) argue that geography works through the association with poor institutions since European migrants had little incentive to establish good institutions in locations with high settler mortality. Easterly and Levine (2002) show evidence that the geography mechanism works entirely through institutions and there is no independent effect of geography after controlling for institutions. Also relevant to this debate are earlier papers such as Ciccone and Hall (1996) that saw data on output across space as an opportunity to test for and to estimate the magnitude of increasing returns to scale.

The point of this paper is to argue that two contrasts in the data provide useful information. The first piece of useful information is the contrast between regional or sub-national data on the one hand and cross-country data on the other hand. To the extent that the institutions that matter are things like risk of appropriation from the Sovereign that are reasonably constant within countries, regional data provides a way to test for the effects of geography and climate holding institutions constant. The second piece of useful information is the contrast between output per land area and output per capita. The reason this second contrast can be useful is that there is an attenuation bias when looking at GDP per-capita. To the extent that factors of production can migrate across regions and countries, both the numerator and denominator in GDP per-capita should be influenced by the incentives provided by nature and geography. Indeed, if factors are perfectly mobile, a simple competitive model would predict that in equilibrium geography and climate would have no observed impact on GDP per-capita, even if there was an underlying effect in the production function. This provides a second reason for comparing regional and national data since presumably the extent of factor mobility is greater across regions than across countries. The testable predictions are that we should expect to see stronger effects of geography on output density than on output per capita in both kinds of data and that we should expect to see lower effects of geography on output per capita in regional data since regions have more factor mobility. More generally tests for the impact of geography based exclusively on output per capita are likely to be incomplete if the bias due to migration is important.

To deal with these kinds of issues, I have constructed a data set of identical GDP and geography variables at both the national level and at the sub-national (regional) level. For regional data, I examine data from nations that are large enough to have significant geographical variation. These large countries are Brazil, China, India, and the United States. Since the data are available and there is no harm in having more information, I also use regional data for several large economies in the European Union, including Germany, France, Italy and the United Kingdom. The paper compares identical regressions estimated with both kinds of data.

2 General observations about estimating the impact of geography on the economy

The research in this paper is motivated by the following general observations about estimating the impact of geography on economic activity.

1. Barro-style growth regressions are likely to underestimate the impact of geography on income, because they condition on initial income and, if geography and climate enter the production function, part of the geography effects are already controlled for in the initial income variable. This suggests that regressions of levels of GDP on geography variables provide better estimates of the full effect of geography. The usual problem with levels regressions, that of reverse causality, can be avoided by choosing regressors that are purely natural features of a region or country and as such independent of GDP. However, levels regressions of GDP per-capita are not sufficient.

2. Looking at output per person is also likely to underestimate the impact of geography effects. As long as mobile factors of production such as labor are able to move away from the 'bad-geography' areas, and there is some sort of diminishing returns, then factor mobility will vitiate geography effects. This consideration suggests, first, that we think about different dependent variables than Y/L, such as Y itself, or output per land area or population density. Second, it suggests that it would be helpful to compare estimates of the impact of geography in areas where the degree of factor mobility varies, hence the motivation for contrasting countries with regions of large countries.

2.1 The bias from factor mobility

The easiest way to illustrate this bias is to think of the competitive Cobb-Douglass case. Suppose the production function in each region depends on one mobile factor (labor), one region-specific factor (capital) and land area. Capital is used broadly to include natural factors such as climate and geography as well as man-made additions to the capital stock. Land area is introduced to maintain a distinction between raw land area and improvements to the land (captured in capital). The notation is Y for output, L for labor, K for capital, and T for land area.

$$Y_i = L_i^{\alpha} K_i^{\beta} T_i^{\gamma} \tag{1}$$

We now reproduce a series of familiar results to set the stage for the empirical section. First, labor demand, conditional on wages (w) and the price of output (p) is.

$$L_i^d = \left(\frac{\alpha p_i}{w_i}\right)^{\frac{1}{1-\alpha}} (K_i)^{\frac{\beta}{1-\alpha}} (T_i)^{\frac{\gamma}{1-\alpha}} \tag{2}$$

If workers can move to regions that offer the best wages, and do not care about non-wage amenities, then wages will be equal across regions ($w_i = w$) , and employment and output in each region will be a function of p_i; K_i and T_i. The expression for regional employment under this case of full labor mobility is equation 2 above with ($w_i = w$). The expression for regional output with full labor mobility, obtained by substituting 2 into 1 is equation 3 below.

$$Y_i = \left(\frac{\alpha p_i}{w}\right)^{\frac{\alpha}{1-\alpha}} (K_i)^{\frac{\beta}{1-\alpha}} (T_i)^{\frac{\gamma}{1-\alpha}} \tag{3}$$

The essential point in this example is that output per given area of land and population for given area of land are jointly functions of 'capital' or geography. However, dividing equation 3 by equation 2, it is clear that output per worker will not be a function of geographical advantages, since the terms involving K and T cancel:

$$\frac{Y_i}{L_i} = \frac{w}{\alpha p_i} \tag{4}$$

This provides an extreme but nevertheless clear example in which geography does affect the distribution of employment and output across space but in which geography would have absolutely no observed impact on the distribution of output per worker. To fix ideas, it is helpful to think of equation 1 as an equation that holds across regions without labor mobility and equation 3 as an equation that would hold across regions with labor mobility. These are the two polar cases. Note that another prediction of this type of model is that the impact of geography on output per acre would be larger in regions with labor mobility, because labor moves to reinforce the natural advantages. This is shown by comparing the exponents on K in equations 1 and 3. Since $\frac{\beta}{1-\alpha} > \beta$, geography would always have a larger effect on output per land area in areas with high labor mobility.

2.2 The impact of distance on growth and steady state output

The model above posits a geography effect directly in the production function. This may be a reasonable simplification for circumstances in which climate directly reduces labor productivity or agricultural production or accelerates the depreciation of capital. This section shows an example in which poor geography can also reduce growth and the level of income in the steady state due purely to the costs associated with remoteness. Consider a representative enterprise with an investment technology that combines an imported machine with domestic capital to produce a composite capital good according to $k = \min[\theta k^f; k^d]$. The superscripts stand for foreign and domestic capital. With this technology, it is optimal to combine the two kinds of capital goods in fixed proportions: $\theta k^f = k^d$. To increase the composite capital stock by one unit, the enterprise must purchase θ units of the foreign capital good at a price of, say, p, and 1 unit of the domestic capital good at a price of p^d. Setting $p^d = 1$, the cost of an extra unit of the composite capital good is therefore $1 + \theta p$. We assume further that the domestic capital good and the domestic output are identical and that the cost of installing one unit of the capital good is $C(I)$ units of the domestic good. Specifically, let $C(I) = \frac{I^2}{2}$, and assume production takes place according to $y = Ak$. Then the representative firm solves the following Hamiltonian:

$$\max \int_0^\infty e^{-rs}\left[Ak - (\theta p + 1)\frac{I^2}{2}\right] ds + q(I - \delta k)$$

From which we obtain,

$$\dot{q} - (r + \delta)q = -A$$

and

$$I = \frac{q}{\theta p + 1}.$$

Integrating...we have

$$q = \frac{A}{r + \pm}$$

Therefore,

$$I = \frac{A}{(\mu p + 1)(r + \pm)}$$

Given the production function $y = Ak$, we have the familiar two sources of growth: productivity growth and capital accumulation ($g = \frac{\dot{A}}{A} + \frac{\dot{k}}{k}$). We suppress the productivity growth term to focus on the essential points in the discussion below, hence.

$$g = \frac{\dot{k}}{k}:$$

Growth in the capital stock comes from the accumulation equation. Given,

$$\dot{k} = I \;¡\; \pm k;$$

growth in the capital stock is

$$\frac{\dot{k}}{k} = \frac{I}{k} \;¡\; \pm:$$

The path of the capital stock over time is therefore

$$k(t) = k(t_0)e^{¡\pm(t\,¡\,t_0)} + \frac{I}{\pm}(1 \;¡\; e^{¡\pm(t\,¡\,t_0)}):$$

Note that in the limit as $t \; ! \; 1$; $k(t) \; ! \; I{=}\pm$:

Since output is

$$y(t) = A ¤ k(t):$$

the path for output is:

$$y(t) = A ¤ k(t_0)e^{¡\pm(t\,¡\,t_0)} + \frac{A^2}{(\mu p + 1)\pm(r + \pm)}(1 \;¡\; e^{¡\pm(t\,¡\,t_0)}):$$

Di¤erentiating with respect to time, growth is given by:

$$g = \frac{A}{k(t)(\mu p + 1)(r + \pm)} \;¡\; \pm:$$

In this model any additional costs due to transportation over great distances or di¢cult terrain or di¢cult climate can raise p and reduce both growth and the level of income in the steady state. This is therefore an additional mechanism that can account for an inverse association between transport costs and income.

2.3 The interaction with increasing returns to scale

Another consideration in explaining the distribution of output across space is the possibility of increasing returns to scale. In an interesting recent paper, Antonio Ciccone and Robert Hall (1996) use regional data from the United States to estimate increasing returns. They present a model of local geographic externalities in which the basic proposition is that output in any acre of space depends positively on total output per acre of the county in which the acre is situated. Using their notation:

$$q = n^{\alpha}\left(\frac{q_c}{a_c}\right)^{\frac{\lambda-1}{\lambda}} \tag{5}$$

Where q is output of a given acre, n is employment in that acre, q_c is output of the county, a_c is total acreage of the county, α is a parameter for labor productivity and λ is a parameter indicating the degree of geographic spillovers or increasing returns to scale. The core idea is to try to estimate λ, which indicates the sensitivity of output of any small acre to the density of economic activity in the larger county in which it is situated (measured by the $\frac{q_c}{a_c}$ term). To solve this model, note that labor is equally productive in all acres within a county, so Ciccone and Hall assume reasonably that labor is evenly distributed across the county ($n = \frac{n_c}{a_c}$). Then, since total output of the county is q times the number of acres ($q_c = a_c q$), equation 5 solves to:

$$\frac{q_c}{a_c} = \left(\frac{n_c}{a_c}\right)^{\alpha\lambda} \tag{6}$$

This model suggests that regressions of (log) output density on population density would be a way to shed light on the magnitude of geographic increasing returns to scale. If the estimate of $\alpha\lambda$ is greater than 1, then the implication would be that increasing returns must exist in the sense that λ must exceed 1. Ciccone and Hall pursue this idea as well as other related ideas and generally find evidence of increasing returns to scale across U.S. counties.

It is natural to ask how the picture would change if we brought in geographic advantages to this model. Let us assume that each acre within a county has some geographical attribute summarized by the term k^{θ}. As soon as we introduce this, we are forced to deal with the issue of factor mobility and can no longer assume that labor would be evenly distributed across each acre in the county ($n = \frac{n_c}{a_c}$). If we instead assume that labor will be distributed to equate the marginal products across each acre, then the reduced form for output density of each county would not be equation 6 but rather:

$$q_c = \left(\frac{w}{\alpha p_c}\right)^{\frac{\lambda\alpha}{\alpha\lambda-1}} a_c^{\frac{\lambda-1}{\alpha\lambda-1}} \left(\sum k^{\frac{\mu}{1-\alpha}}\right)^{\frac{\lambda(\alpha-1)}{\alpha\lambda-1}} \tag{7}$$

Output at the county-level would depend on land area (with an elasticity - the exponent on a_c - that is not in general equal to one), and would depend on an index of the geographical attributes of each acre in the county (the $\sum k$ term). Note that the elasticity of output with respect to the geography index is $\epsilon = \frac{\lambda(\alpha-1)}{\alpha\lambda-1}$. Differentiating with respect to the increasing returns parameter, λ, we have:

$$\frac{d\epsilon}{d\lambda} = \frac{1-\alpha}{(\alpha\lambda-1)^2} > 0. \tag{8}$$

This shows that the elasticity on the geography index is increasing in λ; the higher are increasing returns, the greater is the sensitivity of output of the county to the county's geographical characteristics. Increasing returns serve to magnify the impact of geography on county output.

On the microeconomics of agglomeration, it is known that if industries have intermediate inputs with positive transport costs, and the intermediates are produced with declining average costs, then in equilibrium industries will tend to locate close to one another. A related idea, demonstrated formally with monopolistically-competitive intermediate producers, is that if local markets are active, generating a high level of demand, then more producers of the differentiated intermediates will break even, and there will be a larger number of intermediates and higher output of final goods. On a related point involving transport costs, it has been shown that if workers are very mobile, even very small differences in transport costs can generate agglomeration (Krugman 1991). Other studies point out that patterns of agglomeration can be reinforced by endogenous technology adoption (Ciccone 1992). On the empirical side, these issues are related to the studies that attempt to use spacial data to estimate the magnitude of increasing returns to scale. For example there are a number of studies that demonstrate that productivity in the US tends to rise with urbanization. Ciccone and Hall (1996) find a strong, positive relation between regional productivity in the US and population density.

Increasing returns working through cities and agglomeration provide an additional channel for the observed association between geography and income. Cities are more likely to be established in good geography areas.

Geography can provide the initial impetus, and then agglomeration e¤ects can give the initial geographical advantage an extra kick, as in equation 7 above. Once established, increasing returns kick-in and the urban economy becomes more prosperous than the geographical advantage alone would suggest.

There are two main empirical tasks. One is to understand the extent to which the income density / population density association is explained by a general increasing returns to scale phenomena as in Ciccone Hall or whether it has to do with agglomeration and the presence of large cities. A second empirical task is to try to estimate any pure agglomeration e¤ect by exploiting the fact that some cities were formed not based on a geographical advantage but for political, administrative or military reasons. This provides scope for instrumental variables estimation.

3 Empirical Strategy

The general framework for thinking about geography, institutions, city formation, agglomeration economies and economic activity is summarized in …gure 1. Geography can a¤ect the economy directly but it may also work through city formation. The apparent e¤ect of geography may also be due to an association between geography and institutions due to selective European migration to areas with favorable geography. City formation can depend on geography as well as political or military reasons. The particular channel of impact of geography on the economy that runs through city formation may be magni…ed by the presence of agglomeration (as in equation 7).

Figure 1 also helps clarify the econometric strategy. First, I will estimate reduced form equations where only geographical variables appear as regressors. These are used for preliminary evidence on the existence of the factor mobility bias. Second, I will estimate regressions that condition on a large city variable in order to test for geography e¤ects through the channel labelled "B" (i.e. that do not work through city formation). The city variable will be the percent of the population in a region that lived in cities in 1990 with populations greater than one million. The city variable will be introduced in OLS regressions and also in IV regressions where the instrument will be a dummy variable to pick out cities that were formed for non-geography reasons. This IV regression will be used for evidence on channel "E". The regional data will be used also to test for channel B since institutions are more constant across regions than across countries.

To summarize, there are four issues to be dealt with in explaining economic activity across space: the impact of natural factors such as climate and geography; the impact of factor mobility; the extent of agglomeration economies associated with cities, and the impact of institutions. The evidence in this paper is designed to try to shed light on each of these.

The models in the previous section suggest that both employment and output are endogenous with respect to geography. Furthermore, a general result is that the quantitative estimates should differ depending whether one considers output per land or output per worker (or person). In line with this, the three dependent variables will be output density (output per square acre of land), population density, and output per capita. Remember also that there are two types of data: cross-country and cross-region. For the cross-country data, I generally estimate regressions in log levels. I estimate separate regressions when I am dealing with regional data, because of greater labor mobility. In these regressions I control for differences in national income levels across the regional data by using national country-dummies (fixed effects). For example there is a dummy for India and another for the U.S.A. in the regressions that use data for Indian states and US states.

4 Data

The data are taken from a number of sources: the Penn World Tables Mark 5.6, the geographical information system data sets on the world wide web (courtesy of Andrew Mellinger), various national sources for the regional data for the U.S., Brazil, India, China and Western Europe. With very few exceptions, the economic and demographic data are for the year 1990.

5 Results

The regression results are summarized in seven tables. Tables 1 and 2 have the simplest regressions, for cross-region data and cross-country data respectively. Tables 3 and 4 repeat these but control for large cities - the percent of the population that live in cities with population greater than one million. Table 5 shows a set of regressions where institutions (instrumented by settler mortality) is added to the basic specification. Table 6 then re-estimates the regression in table 3 with instruments for the urban variable. For comparison with the earlier literature, table 7 replicates the Ciccone-Hall regressions of output density on population density using the new data in this paper.

The main conclusions can be summarized with the following list of points.

1. On the question of which geography and climate variables are important the data suggest three: tropics, remoteness from the coast or a river, and mountainous terrain. The regressions that show this are not shown in this paper to conserve space. I was not able to …nd any e¤ect of polar zones, temperate zones or sub-tropical zones. The only climate variable that had a consistent negative e¤ect was the percent of land area in the extreme tropics. Average temperature or precipitation also did not have an e¤ect. On the remoteness variable, I measure the distance in kilometers from the "centroid" of a region/country to the nearest coast or navigable river. This proved to be a better variable that simply the distance to the coast. In addition, if a country or region bordered the sea, it was given no penalty for being large: i.e. the distance was set to zero. This adjustment is important for a country like Australia, which otherwise would be measured as being far from the sea (because the centroid is far from the sea). On the rough terrain variable, I measure the log of the standard deviation of elevation in the region/country. This proved to be a better variable that the average elevation or the coe¢cient of variation of elevation.

2. The …rst question to test from the theory is whether the degree of factor mobility a¤ects the impact of geography on GDP per-capita. By comparing equations (1) and (4), the theory suggests that if labor were not mobile, output per worker would be a function of geography (equation 1). But with high labor mobility output per worker would not depend on geography (equation 4). Is the observed impact of geography on per-capita GDP indeed higher across countries than across regions and is it in fact zero across regions?

3. The evidence on these questions comes from comparing the geography coe¢cients in table 1 regression 3 (cross-regional data, therefore high mobility) versus table 2 regression 3 (cross-country data, therefore low mobility). We expect the absolute value of the estimated coe¢cients to be higher in table 2 than in table 1. This is indeed the case for the tropics variable and the remoteness variable. For the surface roughness variable, neither estimated coe¢cient is signi…cant. Furthermore, after controlling for the e¤ect of large cities we get if anything stronger results. Comparing regression 3 in tables 3 and 4, the absolute value of the estimated coe¢cients are still higher in cross-country data with low factor mobility than in cross-region data with high factor mobility. In addition, these di¤erences are statistically signi…cant. In all cases

where the estimated coefficient is statistically significant, its confidence interval does not contain the estimate of the sister coefficient from the other table. For example, in table 2 regression 3 the estimated coefficient for the tropics variable is -1.58. The 95 percent confidence interval for this is (-1.02,-2.14). The estimated coefficient in table 1 regression 3 of -0.5 is not contained in this confidence interval. In this sense the estimated coefficients are statistically different from one another. The final question is whether the impact of geography on per-capita GDP is actually zero in areas with high factor mobility - the extreme prediction with high factor mobility. For this we look at regression 3 in table 1 and table 3. Note that all the geography coefficients are statistically zero in table 3 in which the regression controls for large cities, but not in the same regression in table 1 which does not control for large cities. This result suggests that in areas of high factor mobility, some of the observed effect of geography on per-capita GDP comes about because major cities have tended to form in geographically favored areas. We conclude that labor mobility does seem to matter for understanding the magnitude of the impact of geography on GDP per-capita. In regions with labor mobility, tropical status and distance from the sea do not help explain per-capita output, but have stronger measured effects across countries where there is less labor mobility.

4. The previous results were all about per-capita GDP, and suggest a downward bias associated with factor mobility. A second general prediction of the theory is that the estimated impact of geography will be stronger on output density than on output per-capita. The reason is that labor should systematically move away from the bad geography areas, reinforcing the effect on GDP per acre but attenuating the effect on GDP per capita. One theoretical motivation for such a result is in equation 3 and equation 4. In equation 3, holding constant land area, output is an increasing function of K. If we re-write equation (3) so that output density is on the left, we get

$$Y_i/T_i = \left(\frac{\alpha p_i}{w}\right)^{\frac{\alpha}{1-\alpha}} (K_i/T_i)^{\frac{\beta}{1-\alpha}}$$

so that output density is a function of geographic capital of the region per acre of land. But in equation 4 output per-capita is not a function of K. For evidence, compare regression 1 with regression 3 in any of the following tables: 1, 2, 3, or 4. With a few minor exceptions, the absolute values of the coefficients on the geography variables in the first regressions are larger than those in the third regressions. Often the coefficients are smaller in a statistical sense as well. The conclusion is that the magnitude of geography effects are underestimated by analyses that focus solely on output per capita.

5. Acemoglu, Johnson and Robinson (2001) argue that the association we observe today between geography and income is a by-product of the fact that European migrants with traditions of good institutions selectively migrated to temperate climates with low mortality risk. They invested in establishing better or more durable institutions in areas with large European migrant populations or where the European settlers were expected to survive. This paper and related empirical results in Easterly and Levine (2002), which use mortality risk in the 18th century as an instrument for institutions suggest that there is no direct geography effect on per capita income after controlling for institutions. This paper has two kinds of evidence on this issue. The first is that there is a significant effect of geography on output density in the regressions using regional or sub-national data (table 1 or table 3). A large class of institutional variables that operate at the national level are held constant by the national dummy variables in these regressions. Some may argue that there remain institutional differences across regions of large countries. For example, it may be argued that the north of Brazil has different institutions than the south of Brazil because the North has a legacy of slavery, and thus even these cross-region regressions do not fully control for institutional differences. The response to this is basically that these differences need to be proven and established on a systematic basis and in any case would be far lower than the differences in institutions across countries. When the regressions in table 1 are estimated without Brazil, the elevation measure remains highly significant in the regression of output per-capita and distance and elevation remain highly significant in the regression on output density. The second response to the institutions-only school is that the results that support this are based on regressions using output per-capita, neglecting the factor mobility bias. In the cross country regressions here in table 5, we replicate this particular result in the regressions using per-capita GDP, but not in the regressions using output density. When output density is regressed on risk of appropriation (instrumented by settler mortality) and the geography variables, we find that although all variables have the anticipated sign,

neither risk of appropriation nor the geography variables are significant at the 5 or 10 percent level. In other words, the cross-country data are too highly correlated for the regression to discriminate effectively between these two explanations.

6. We now turn to urbanization. One reason for looking at urbanization is that a casual glance at the data shows that the regions with unusually high per-capita incomes all contain large cities. From a purely empirical point of view, this strongly suggests that one needs to deal with urbanization for a complete explanation of economic activity across space. The regression evidence showing the large city result is in tables 3 and 4. Large cities are measured by the percentage of the population in a regions/country that lives in cities with populations in excess of one million. Note that this variable is strongly associated with higher incomes in both across countries and across regions.

7. The first question is the extent to which urbanization is affected by geography. I have gone over the history of all of the large cities in the world and have identified several cities that were founded for clearly military or political purposes. Madrid owes its location to the fact that the King of Spain wanted his capital in the center of his kingdom. Otherwise there is no geographical reason for Madrid to be where it is. It is not even near a large river. The Manzanares was quite small and not serviceable for major commerce. Saint Petersburg was founded by Peter the Great to have a capital away from Moscow with access to the sea to fight naval battles with the Swedish Empire. Its harbor is frozen for six months out of every year, so it is not a particularly valuable assets from an economic point of view. Certainly it did not arise in that location for any economic reason because when Peter arrived it was just a swamp. Delhi and Mexico City are also cities that exist in their present location for non-economic reasons. To account for these cases I have created a dummy variable that takes the value 1 if a region contains a city that was founded for non-economic reasons, 0 otherwise. For purposes of estimation, I presume that urbanization is potentially a function of natural and geographic attributes of the region but also these non-economic, exogenous reasons. In line with this, the regressions to explain urbanization in table 5 contains both kinds of variables: the exogenous city variable and the geography variables. The results show that both geography and the explanations summarized by the dummy variable help explain urbanization. The tropical

variable and the surface roughness variable are the most powerful. And the regression has higher explanatory power for the regional data than for the cross-country data. The exogenous city variable works with the regional data but not the cross country data. Overall, the regressions in table 5 support the idea that our geography variables work through city formation to some extent. Because the empirical estimates are more precise, the regional data set seems best for examining this issue further.

8. Continuing now with the idea that urbanization is partly caused by geography and partly by other factors we now check the previous results by re-estimating the regressions in table 3 with instrumental variables. The instrument for the urban variable is the exogenous city variable. Essentially, we are using variation in urbanization generated only by the exogenous cities to estimate a pure urban effect. The coefficients in table six are better estimates of the urban effect than are the coefficients in tables 3 or 4. A second reason for doing this is to check whether the estimated geography effects are sensitive to instrumenting the urban variable. A comparison of the non-urban coefficients in table 6 and table 3 shows that the instrumenting does not change the estimates substantially..

9. Previous research has used data across regions (specifically population density or urbanization) to test for economies of scale and to estimate the magnitude of economies of scale. Here we estimate similar equations using our data and compare the results with this research. In table 7 we show estimates of Ciccone and Hall (1996) regressions with and without controls for large cities. The equation we are estimating is equation (6) above. This corresponds to equation (2) in Ciccone and Hall. As they mention (page 56) only the product $\alpha\beta$ is identified by these regressions. In their interpretation α measures the effect of congestion and β measures the effect of agglomeration. They consistently obtain estimates in the range 1.06, indicating, in their words, "that the net effect favors agglomeration". The comparable results with the present data are estimated coefficients of 1.07 (regional data) and 1.10 (cross-country data). These estimates are very close to Ciccone and Hall using entirely different data. However, Table 7 also shows that the control for very large cities reduces both estimates to one. Therefore, the fact that the estimated coefficient without any controls is greater than one is entirely due to the regions and countries with

very large cities. It is not a general continuous result that holds at all levels of population density. This suggests that the increasing returns behind these estimates are more specifically agglomeration economies associated with very large cities. Outside these large cities output density and population density correlate with an elasticity that is not significantly different from one.

6 Conclusions

This paper has presented evidence designed to shed light on the importance of three competing explanations for the large differences in economic activity and prosperity across the world: climate/geography, institutions and agglomeration economies associated with urbanization. The paper agrees with the view, shared by other authors, that the best way to estimate such effects is with data on levels of economic activity rather than growth rates. However, the paper extends the evidence in two ways: first, by comparing data across regions in addition to data across countries; second, by examining output density and population density in addition to output per-capita.

The results suggest that the geographical and climate variables that have the highest explanatory power over income levels are distance to rivers or coastlines, extreme tropical climates, and mountainous terrains. More remote regions, extremely tropical regions and highly mountainous regions have lower output densities and population densities. These are the simple reduced-form relationships, presented in regressions in tables 1 and 2, for cross regional data and cross country data respectively.

The paper argues further that we need to look at output density rather than output per capita to fully test for and estimate the impact of geography because factor mobility can mask the underlying effect of geography on production. The paper showed the polar case of a competitive model with high factor mobility in which geography would have absolutely no observed effect on per-capita income, even though there was indeed an underlying influence of geography in the production function. The regional data showed that this bias due to factor mobility is important, since the geography variables exhibited larger effects on output density than on output per-capita, especially when controls were introduced for very large cities. The same result shows up in the cross country regressions, although the mobility bias is presumably lower due to lower factor mobility across countries.

The results from both cross-region and cross-country data actually support the prediction of the competitive model with high factor mobility, namely that in equilibrium there will be zero association between geography variables and output per-capita even though there may be a real e¤ect of geography in the production function. This result is found in the regional data after controlling for very large cities (table 3 or table 6), and in the cross-country data after controlling for instrumented institutions (table 5).

The paper also contains results to help interpret what lies behind the geography-income association. One recent view is that this association is a coincidence due to the fact that new world migrants in past centuries migrated to regions with low mortality risk, creating an association today between geography and income that really is about institutions rather than geography. The evidence here does not support the view that the association of poor climate and geography with lower income is completely explained by the association of poor geography with poor institutions. The …rst reason is that there is a signi…cant e¤ect of geography on output density in the regressions using regional or sub-national data (table 1 or table 3). A large class of institutional variables that operate at the national level are held constant by the national dummy variables in these regressions. A second reason is that previous authors have made the claim that there is no independent geography e¤ect after controlling for institutions based on cross country regressions of output per-capita, neglecting the factor mobility bias. In the cross country regressions here, we replicate this result in the regressions using per-capita GDP, but not in the regressions using output density (table 5). In the latter regression, where we regress output density on risk of appropriation (instrumented by settler mortality) and the geography variables, we …nd that although all variables have the anticipated sign, neither risk of appropriation nor the geography variables are signi…cant at the 5 or 10 percent level. In other words, the cross-country data are too highly correlated for the regression to discriminate e¤ectively between these two explanations. Therefore, while the association between poor geography and poor institutions can accounts for some of the association between poor geography and low income, the results from regional data and output density suggests that it cannot account for all of the association.

Another possible channel through which poor geography and climate affect income is by the establishment of cities in favorable geographic areas. If cities tend to be formed in areas with good geography and agglomeration economies are quantitatively important, then we will observe higher income in these areas, due to the location of cities. We look at the regional data for evidence on this issue since the …ner level of disaggregation allows us to isolate regions dominated by large cities. We …rst …nd that the (instrumented) large city variable does explain higher income, supporting the view that there are indeed agglomeration economies (table 6). But we also …nd that even after accounting for this, the geography variables have a signi…-cant impact on output density, supporting the existence of an independent geography channel (also table 6).

Using our data across regions and across countries, we …nd that output density correlates with population density with an elasticity of approximately 1.07. This is the same result found in Ciccone and Hall (1997) using data exclusively from the United States. This result is sometimes interpreted as proof for the importance of general increasing returns to scale. Here we …nd that the elasticity greater than one can be accounted for entirely by very large cities. This suggests that the result is closely connected to the existence of agglomeration economies associated with cities rather than a general increasing returns result. Variation in population density outside very large cities correlates only one for one with output density.

In summary, the results here run against the view that there is no independent geography e¤ect after controlling for institutions. The evidence supports a role for geography working through institutions, but also an independent e¤ect of geography associated with climate and distance. There is also a channel of causality that runs from geography through urban formation to output. There also appears to be an independent large city e¤ect or agglomeration e¤ect since cities founded for non-economic reasons also exhibit higher incomes.

References

[1] Acemoglu, Daron, Simon Johnson and James A. Robinson, The Colonial Origins of Comparative Development, American Economic Review, December 2001.

[2] Ciccone, Antonio; Hall, Robert E. Productivity and the Density of Economic Activity. American Economic Review. Vol. 86 (1). p 54-70. March 1996.

[3] Ciccone, Antonio Fausto. Cities and the Economy: Aggregate Production, Urban Productivity and the Distribution of Economic Activity Mimeo, Stanford University. 1992.

[4] Easterly, William and Ross Levine, Tropics, Germs and Crops: How Endowments In‡uence Economic Development, March 2002.

[5] Hall, Robert E; Jones, Charles I. Levels of Economic Activity across Countries. American Economic Review. Vol. 87 (2). p 173-77. May 1997.

[6] Sachs, Je¤rey D; Warner, Andrew M. Sources of Slow Growth in African Economies. Journal of African Economies. Vol. 6 (3). p 335-76. October 1997.

[7] Sachs, Je¤rey D; Warner, Andrew M. Fundamental Sources of Long-Run Growth. American Economic Review. Vol. 87 (2). p 184-88. May 1997.

[8] Krugman, Paul. Geography and trade. Gaston Eyskens Lecture Series Cambridge, Mass. and London: MIT Press; Louvain, Belgium: Louvain University Press. p xi, 142. 1991.

[9] Glaeser, Edward L; Mare, David C. Cities and Skills. National Bureau of Economic Research Working Paper: 4728. p 22. May 1994.

[10] Gallup, John Luke; Sachs, Je¤rey D; Mellinger, Andrew D. Geography and Economic Development. International Regional Science Review. Vol. 22 (2). p 179-232. August 1999.

Table 1. Regressions of output density, population density and GDP per capita on geographic variables (data are for regions of large countries)

	(1) output density	(2) population density	(3) log GDP pc 90
percent tropical	-1.74	-0.81	-0.50
	(3.02)**	(1.21)	(2.94)**
km from river/coast	-1.08	-0.93	0.04
	(3.04)**	(8.39)**	(0.39)
rough terrain	-0.66	-0.31	-0.11
	(7.35)**	(4.12)**	(3.96)**
Constant	16.80	5.76	7.79
	(31.45)**	(14.65)**	(49.31)**
Observations	179	260	180
R-squared	0.63	0.47	0.93

Absolute value of t-statistics in parentheses

* significant at 5% level; ** significant at 1% level

(Dummy variables for countries are included in the regression but not shown)

Table 2. Regressions of output density, population density and GDP per capita on geographic variables (Cross-country data)

	(1) output density	(2) population density	(3) log GDP pc 90
percent tropical	-2.26	-0.98	-1.58
	(4.01)**	(2.66)**	(5.67)**
km from river/coast	-1.82	-0.54	-1.47
	(3.11)**	(2.14)*	(5.08)**
rough terrain	-0.43	-0.12	-0.10
	(2.70)**	(1.31)	(1.21)
Constant	14.70	4.67	8.90
	(16.16)**	(9.47)**	(19.75)**
Observations	94	153	94
R-squared	0.26	0.08	0.40

Absolute value of t-statistics in parentheses

* significant at 5% level; ** significant at 1% level

Table 3. Regressions of output density, population density and GDP per capita on geographic variables and urbanization (data are for regions of large countries)

	(1) output density	(2) population density	(3) log GDP pc 90
percent tropical	-1.07	-0.89	-0.27
	(2.02)*	(1.68)	(1.72)
km from river/coast	-1.18	-1.41	0.04
	(3.38)**	(4.09)**	(0.41)
rough terrain	-0.47	-0.43	-0.04
	(5.37)**	(4.95)**	(1.65)
percent in cities>1mill	2.57	1.81	0.83
	(6.20)**	(4.35)**	(6.80)**
Constant	15.59	7.92	7.37
	(29.41)**	(14.89)**	(47.34)**
Observations	173	174	174
R-squared	0.71	0.64	0.95

Absolute value of t-statistics in parentheses

* significant at 5% level; ** significant at 1% level

(Dummy variables for countries are included in the regression but not shown)

Table 4. Regressions of output density, population density and GDP per capita on geographic variables and urbanization (Cross-country data)

	(1) output density	(2) population density	(3) log GDP pc 90
percent tropical	-1.71	-0.70	-1.32
	(3.15)**	(1.71)	(4.87)**
km from river/coast	-1.47	-0.42	-1.31
	(2.64)**	(1.10)	(4.74)**
rough terrain	-0.35	-0.15	-0.05
	(2.33)*	(1.47)	(0.72)
percent in cities>1 mill	3.98	2.23	1.80
	(3.70)**	(2.58)*	(3.36)**
Constant	13.80	4.61	8.46
	(15.61)**	(7.78)**	(19.20)**
Observations	93	131	93
R-squared	0.35	0.11	0.47

Absolute value of t-statistics in parentheses

* significant at 5% level; ** significant at 1% level

Table 5. Cross-country regressions with risk of appropriation instrumented by settler mortality.

	(1) output density	(2) population density	(3) log GDP pc 90
Average.protection Against appropriation	0.68	-0.40	0.85
	(1.55)	(0.87)	(4.02)**
percent tropical	-0.89	-0.55	-0.54
	(1.02)	(0.71)	(1.30)
km from river/coast	-0.52	-1.61	0.71
	(0.41)	(1.58)	(1.16)
rough terrain	-0.27	-0.29	0.16
	(1.36)	(1.75)	(1.66)
Constant	8.64	7.96	1.36
	(2.36)*	(2.18)*	(0.78)
Observations	52	60	52
R-squared	0.15	0.05	0.34

Absolute value of t-statistics in parentheses

* significant at 5%; ** significant at 1%

Table 6. Instrumental variable estimates of the impact of geography and large cities on economic activity (data are for regions of large countries)

	(1) output density	(2) population density	(3) log GDP pc 90
percent in cities> 1 mill	4.02	3.43	0.71
	(3.43)**	(2.90)**	(2.14)*
percent tropical	-0.73	-0.50	-0.30
	(1.19)	(0.81)	(1.72)
km from river/coast	-1.30	-1.53	0.05
	(3.48)**	(4.14)**	(0.48)
rough terrain	-0.36	-0.31	-0.05
	(2.93)**	(2.50)*	(1.47)
Constant	14.84	7.08	7.43
	(18.77)**	(8.88)**	(33.23)**
Observations	173	174	174
R-squared	0.68	0.61	0.95

Absolute value of t-statistics in parentheses
* significant at 5% level; ** significant at 1% level

The exogenous city variable is used as an instrument for percent urban.
(Dummy variables for countries are included in the regression but not shown)

Table 7. Ciccone-Hall regressions of output density on population density, with and without controls for large cities.

	(1) output density (regions)	(2) output density (regions)	(3) output density (countries)	(4) output density (countries)
population density	1.07	1.01	1.10	1.01
	(49.86)**	(47.43)**	(15.80)**	(15.26)**
percent in cities> 1 mill		0.91		2.76
		(7.38)**		(4.31)**
Constant	7.03	7.30	7.50	7.61
	(53.19)**	(55.80)**	(26.47)**	(29.08)**
Observations	180	174	103	100
R-squared	0.97	0.97	0.71	0.76

Absolute value of t-statistics in parentheses

* significant at 5% level; ** significant at 1% level
(the regressions with regional data have separate intercepts for each country)