Accepted Manuscript

Geography, human capital and urbanization: A regional analysis

Matthias Flückiger, Markus Ludwig

PII: S0165-1765(18)30118-6
DOI: https://doi.org/10.1016/j.econlet.2018.03.027
Reference: ECOLET 7988

To appear in: *Economics Letters*

Received date : 19 February 2018
Revised date : 22 March 2018
Accepted date : 25 March 2018

Please cite this article as: Flückiger M., Ludwig M., Geography, human capital and urbanization: A regional analysis. *Economics Letters* (2018), https://doi.org/10.1016/j.econlet.2018.03.027

*Highlights (for review)

Highlights

- We analyze the relationship between geography and human capital in developing countries.
- We find that geography explains a substantial share of within-country variation in the level of human capital.
- Urbanization is potentially an important intermediary factor explaining these findings.

Geography, Human Capital and Urbanization: A Regional Analysis

Matthias Flückiger* Markus Ludwig†

March 22, 2018

Abstract

Using subnationally representative data for 679 regions from 64 low- and middle-income countries, we document that geography, especially trade-related features, explains a substantial share of within-country variation in the level of human capital. We further show that a large part of the explanatory power can be attributed to the close interrelation between urbanization and concentration of human capital. These results bring together studies that identify human capital as an important determinant of regional development and the literature on economic geography that emphasizes geography's role in shaping the spatial distribution of urbanization and economic activity.

JEL classification: O11, R11, E24

Keywords: Geography, Human Capital, Urbanization

*Queen's University Belfast. Email: matthias.flueckiger@qub.ac.uk

†University of Bayreuth. Email: markus.ludwig@uni-bayreuth.de

1 Introduction

Geography's influence on national and subnational economic development is well documented. The relative importance of individual channels through which geography affects development, however, is much less clear. In this study we investigate if and how geography influences the regional level of human capital, a potentially important proximate determinant of economic prosperity (e.g., Acemoglu and Dell (2010); Gennaioli et al. (2013)).

For our empirical analysis, we draw on subnationally representative survey data from the Demographic and Health Surveys (DHS) Program and construct various proxies for human capital at the regional level, such as average years of schooling or prevalence of stunting. In total, our dataset encompasses 679 regions from 64 low-and middle income countries. To assess the explanatory power of geographic variables regarding within-country differences in the level of human capital, we employ a cross-sectional OLS regression approach in which we control for country-fixed effects. For parsimony, we include only nine geographic characteristics in our main regression setup. Following Henderson et al. (Forthcoming), this set of features encompasses five agriculture-related variables (caloric suitability, temperature, precipitation, elevation and absolute latitude) and four trade-associated characteristics (accessibility by inland waterways, share of a region's area that lies with 25 kilometer of an ocean, distance to coastline and terrain ruggedness). Overall, the nine geographic covariates explain a meaningful 23% of the within-country sample variation in average years of schooling. These findings suggest that an important channel through which geography influences regional development is via its effect on human capital.

A potentially important intermediary factor linking geography to human capital is urbanization. The location of these urban areas is strongly influenced by geography-related productivity advantages (Henderson et al., Forthcoming). Additionally, human capital is highly concentrated in the urban, skill-intensive, sectors in developing countries. As these sectors are primarily located in urban areas, this generates a rural-urban gap in the level of human capital (Young, 2013) and suggests that a large part of the explanatory power of geography may be accounted for by its effect on urbanization. Our regression results provide empirical support for this notion. Overall, 65% of the R-squared produced by the nine geographic characteristics is accounted for by the partial correlation of these covariates with urbanization.

Our study contributes to the large body of research that analyzes the effects of geography on economic development at the national (Mellinger et al., 2000; Nunn and Puga, 2012) and subnational level (Rappaport and Sachs, 2003; Dell et al., 2009; Mitton, 2016). Closely related to our paper is further the literature that documents the existence of a large rural-urban gap in productivity in developing countries (Gollin et al., 2014; Young, 2013). By documenting that geography's effects on the regional level of human capital is to a large part accounted for by its interrelation with urbanization, our paper provides a link between these studies and the branch of economic geography that identifies geography as a centrally important determinant of the spatial distribution of urbanization and economic activity (e.g., Davis and Weinstein (2002); Rappaport and Sachs (2003)). Within the latter field, the study of Henderson et al. (Forthcoming), showing that geography is highly predictive of economic activity at the local level, is most closely related to ours.

The remainder of this paper is organized as follows: In the next section, we present our empirical strategy; the data are in described in Section 3 and the results discussed in Section 4. A conclusion is offered in Section 5.

2 Empirical Strategy

We assess the explanatory power of geographic characteristics using the following cross-sectional OLS regression:

$$y_{i,c} = \psi' \mathbf{G}_{i,c} + \mu_i + \varepsilon_{i,c}, \tag{1}$$

where the dependent variable $y_{i,c}$ represents different proxies for the level of human capital in region i and country c. Vector $\mathbf{G}$ encompasses geographic characteristics. Following Henderson et al. (Forthcoming), we classify these into two groups: Characteristics that primarily influence agricultural productivity and characteristics particularly relevant for trade. In our main analysis, the former category includes precipitation, temperature, agricultural suitability, elevation and absolute latitude; the latter encompasses distance to coastline, the percentage of a region's surface area that lies within 25 kilometer of the coastline, accessibility by inland waterways (captured by the Strahler index), as well as terrain ruggedness.

All regressions control for the landmass encompassed by the individual regions as well as country fixed effects (μ_i). Our estimates therefore capture reduced-form effects of geography within countries. The idiosyncratic error term is symbolized by $\varepsilon_{i,c}$; the standard errors are clustered at the country level.

3 Data and Descriptive Analysis

Human Capital

The proxies for human capital are drawn from the Demographic and Health Surveys (DHS) Program. The surveys are representative at sub-national reporting areas which typically correspond to first-level administrative country subdivisions or groups thereof. To construct regional-level variables we average survey responses over regions using the sample weights provided by the DHS. Specifically, we compute average years of schooling, the percentage of households with access to improved drinking water sources and toilet facilities as well as prevalence of stunting in children under five. The representativeness of the DHS data further allows us to determine the urbanization rate for each region, i.e., the share of total population residing in areas that are defined as urban. In total, our dataset encompasses 679 regions from 64 low-and middle-income countries.[1] The spatial distribution of these regions is depicted in Figure 1.

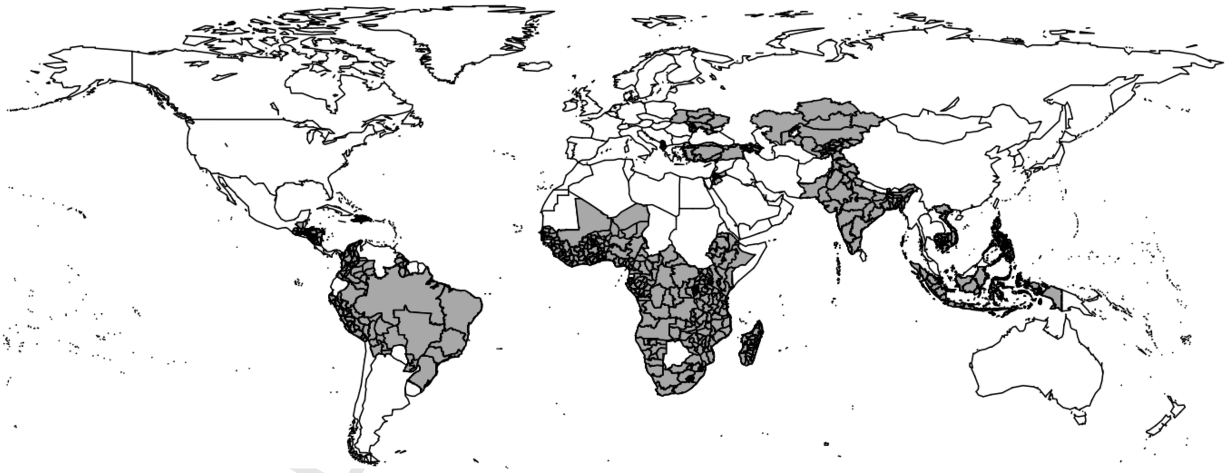

Figure 1: Sample coverage. Regions shaded gray are included in our dataset.

Geography

Long-run (1960–1990) temperature, precipitation and elevation are drawn from Hijmans et al. (2005), terrain ruggedness is taken from Nunn and Puga (2012), and the caloric suitability index from Galor and Ozak (2015). Based on information provided by `www.naturalearthdata.com`, we compute a region's average distance to the coastline and its

[1]When multiple survey waves exist, we use the most recent one. Tables A.1–A.2 report details on countries and survey waves included in our analysis as well as data construction processes.

share of total area that lies within 25 kilometers of an ocean. Accessibility by inland waterways is measured by the Strahler order of streams. For the empirical analysis, we standardize all variables to a mean of zero and a standard deviation of one. The summary statistics are depicted in Appendix A, Table A.3.

4 Results

We evaluate the explanatory power of geography by looking at the within-country R-squared, i.e., the share of variation our model can account for once country fixed effects and surface areas of regions have been partialled out. Below, we first look at reduced-form relationships and then investigate the importance of urbanization as an intermediary factor.

4.1 Geography and Human Capital

The results of Table 1 column (1) show that the five agricultural covariates explain about 9% of the within-country variation in average years of schooling. Of the individual variables, temperature exerts the strongest and statistically most significant effect. This is consistent with studies that document the negative influence of temperature on regional income per capita (e.g., Dell et al. (2009)). The point estimate of –0.881 implies that a one-standard deviation increase in temperature reduces average years of schooling by 0.881 standard deviations. In column (2) we replace the agriculture variables with trade-associated covariates. Compared to the former, the explanatory power of the latter is substantially higher. More than 15% of subnational variation in average years of schooling is explained by the four trade-related geographic characteristics. Among these, access to coast, measured as the percentage of a region's area that lies within 25 kilometers of an ocean, is economically the most significant. A one-standard deviation increase in this covariate raises average years of schooling by 0.2 standard deviations. This results ties in with the literature that identifies coastal access as an important determinant of the spatial distribution of income per capita (e.g., Mellinger et al. (2000); Rappaport and Sachs (2003)). In column (3) we include both agriculture- and trade-related variables in the regression setup. Combined, the nine characteristics explain a remarkable 23.2% of the relevant sample variation.

We next investigate geography's influence on proxies for health, a second important dimen-

Table 1: Geography and Human Capital

Dependent Variable	Average Years of Schooling			Share Improved Drinking Water	Share Improved Sanitation Facilities	Share Children Stunted
	(1)	(2)	(3)	(4)	(5)	(6)
Caloric suitability	0.068		0.176***	0.065	0.101	-0.130
	(0.051)		(0.058)	(0.106)	(0.071)	(0.080)
Temperature	-0.881***		-0.783***	-0.097	-0.440**	0.855***
	(0.180)		(0.177)	(0.209)	(0.186)	(0.275)
Precipitation	0.012		0.012	-0.135*	-0.006	0.108**
	(0.055)		(0.045)	(0.082)	(0.069)	(0.049)
Elevation	-0.571***		-0.333***	0.070	-0.187	0.557***
	(0.111)		(0.118)	(0.132)	(0.117)	(0.175)
Absolute latitude	-0.356*		-0.299**	0.134	-0.029	0.375*
	(0.183)		(0.145)	(0.181)	(0.144)	(0.196)
Strahler index		0.056***	0.063***	0.086**	0.036	-0.089***
		(0.018)	(0.019)	(0.043)	(0.031)	(0.019)
% Area within 25km of coastline		0.200***	0.200***	0.221***	0.214***	-0.274***
		(0.033)	(0.028)	(0.058)	(0.042)	(0.042)
Average distance to coastline		-0.099	-0.127**	-0.061	-0.155***	-0.017
		(0.062)	(0.056)	(0.083)	(0.048)	(0.078)
Ruggedness		0.014	-0.113**	-0.116	-0.023	0.058
		(0.036)	(0.050)	(0.074)	(0.051)	(0.061)
Observations	679	679	679	671	662	600
Within-country R^2	0.090	0.154	0.232	0.081	0.143	0.183

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The table reports standardized beta coefficients. Standard errors clustered at the country level. All regressions control for country fixed effects and surface area of regions.

sion of human capital. For brevity, we only report full specifications in which we control for both agriculture- and trade-related covariates. As depicted in columns (4)–(5), geography explains a meaningful fraction of the relevant sample variation in access to drinking water (8.1%) and sanitation (14%), both of which are important determinants of health (Duflo et al., 2015; Alsan and Goldin, Forthcoming). The results are similar when we look at the share of stunted children in population aged five and below (column (6)). Stunting is an indicator for health during childhood which, in turn, strongly influences the level of human capital accumulated as an adult (e.g., Bleakley (2007); Shah and Steinberg (2017)).

Taken together, the results of Table 1 document that variation in geography—particularly trade-associated geography—is of great importance in accounting for regional differences in the level of human capital.[2] This finding consolidates studies that link human capital to regional income differences (Gennaioli et al., 2013) with papers that identify geography

[2]The explanatory power increases slightly if we include further control variables (Appendix C.1).

as central for regional development (e.g., Dell et al. (2009)). Next,we investigate the importance of urbanization as an intermediary factor underlying the reduced-form effects of geography.

4.2 Geography, Urbanization and Human Capital

In Table 2, we re-run the regressions of Table 1 but additionally include the urbanization rate as an explanatory variable. We then analyze how much of geography's explanatory power can be accounted for by urbanization. That is, we compute the share of geography's total explanatory power regarding variation in the outcome variables that is absorbed by the partial correlation with urbanization.[3] The results document the close relationship between urbanization and proximate determinants of development in low- and middle-income countries. With urbanization included, the predictive power of the regression model grows substantially compared to Table 1. Averaged across all columns, the within-country R-squared increases from 16% to 48%. A likely driver underlying this result is the structural urban-rural gap in human capital (e.g., Young (2013)).[4]

The last row in Table 2 reports the share of explanatory power of geographic variables accounted for by urbanization. With average years of schooling as dependent variable, 38% of the explanatory power produced by agriculture-related covariates is absorbed by the partial correlation with urbanization rates (column (1)). For the set of trade-associated variables, the share accounted for by urbanization increases to a staggering 81%. This implies that the bulk of explanatory power of the trade-related geographic characteristics can be explained by the close relationship between urbanization and human capital: Trade geography strongly influences the extent to which a region is urbanized. This, in turn, influences its average level of human capital.[5] When incorporating both sets of geographic characteristics into the regression setup, 64.9% of their joint explanatory power is absorbed

[3]Formally, this share is given by:

$$\text{Within R}^2 \text{ accounted for by urbanization} = \frac{R_g^2 - \left(R_{u+g}^2 - R_u^2\right)}{R_g^2} = 1 - \frac{\left(R_{u+g}^2 - R_u^2\right)}{R_g^2},$$

where R_g^2 represents the within-country R-squared produced by the geographic characteristics, R_u^2 the within-country R-squared produced by urbanization and R_{u+g}^2 the joint explanatory power of urbanization and geographic features.

[4]Table B.1 documents that this gap is also observable at the individual level in our survey data.

[5]Urbanization and human capital are two reinforcing processes. Disentangling directions of causality between the two variables is not possible within the framework of our analysis. The point estimates for urbanization should therefore not be interpreted as causal effects.

Table 2: Geography, Urbanization and Human Capital

Dependent Variable	Average Years of Schooling			Share Improved Drinking Water	Share Improved Sanitation Facilities	Share Children Stunted
	(1)	(2)	(3)	(4)	(5)	(6)
Share urban population	0.374***	0.353***	0.345***	0.446***	0.387***	-0.312***
	(0.021)	(0.022)	(0.022)	(0.041)	(0.035)	(0.032)
Caloric suitability	0.070		0.119**	-0.009	0.045	-0.067
	(0.043)		(0.047)	(0.087)	(0.058)	(0.072)
Temperature	-0.558***		-0.525***	0.242	-0.166	0.611**
	(0.129)		(0.123)	(0.214)	(0.184)	(0.252)
Precipitation	0.059		0.056	-0.079	0.044	0.053
	(0.041)		(0.036)	(0.068)	(0.069)	(0.040)
Elevation	-0.277***		-0.189***	0.258*	-0.042	0.419***
	(0.078)		(0.073)	(0.145)	(0.140)	(0.150)
Absolute latitude	-0.304**		-0.277**	0.162	-0.013	0.312*
	(0.134)		(0.118)	(0.149)	(0.121)	(0.189)
Strahler index		0.020**	0.029***	0.047*	0.002	-0.060***
		(0.009)	(0.008)	(0.028)	(0.022)	(0.016)
% Area within 25km of coastline		0.074***	0.079***	0.065	0.077*	-0.149***
		(0.020)	(0.019)	(0.058)	(0.041)	(0.047)
Average distance to coastline		-0.060	-0.084**	-0.004	-0.102**	-0.050
		(0.048)	(0.042)	(0.073)	(0.047)	(0.072)
Ruggedess		0.062**	-0.039	-0.020	0.062	0.006
		(0.027)	(0.035)	(0.066)	(0.041)	(0.052)
Observations	679	679	679	671	662	600
Within-country R^2	0.560	0.534	0.586	0.323	0.445	0.337
% Within R^2 accounted for by urbanization	0.380	0.811	0.649	0.704	0.716	0.674

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The table reports standardized beta coefficients. Standard errors clustered at the country level. All regressions control for country fixed effects and surface area of regions. *% Within R^2 accounted for by urbanization* represents the share of the total explanatory power of the geographic variables that is absorb by the partial correlation with the urban population share (see the main text for more details).

by the partial correlation with urbanization. The pattern of results is very similar when we look at health-related outcomes. Averaged across columns (4)–(6), the share of explanatory power accounted for by the correlation with urbanization is 70%.

We conduct a number of robustness checks in order to document the stability of our results. Appendix C shows that the results remain qualitatively unaltered if we drop regions that are home to capital cities or drop the five most and five least urbanized countries.

5 Conclusion

In this study, we document that both agriculture- and trade-associated geographic characteristics are important in accounting for regional differences in the level of human capital. A large part of the explanatory power, particularly for the trade-related variables, is absorbed by the correlation with urbanization. This implies that the effects of geography on the concentration of human capital and urbanization are closely interlinked and is consistent with the existence of a urban-rural gap in human capital.

References

Acemoglu, Daron, and Melissa Dell (2010) 'Productivity Differences between and within Countries.' *American Economic Journal: Macroeconomics* 2(1), 169–188

Alsan, Marcella, and Claudia Goldin (Forthcoming) 'Watersheds in infant mortality: The role of effective water and sewerage infrastructure, 1880 to 1915.' *Journal of Political Economy*

Bleakley, Hoyt (2007) 'Disease and development: Evidence from Hookworm eradication in the American South.' *The Quarterly Journal of Economics* 122(1), 73–117

Davis, Donald R., and David E. Weinstein (2002) 'Bones, bombs, and break points: The geography of economic activity.' *American Economic Review* 92(5), 1269–1289

Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken (2009) 'Temperature and income: Reconciling new cross-sectional and panel estimates.' *American Economic Review* 99(2), 198–204

Duflo, Esther, Michael Greenstone, Raymond Guiteras, and Thomas Clasen (2015) 'Toilets can work: Short and medium run health impacts of addressing complementarities and externalities in water and sanitation.' Working Paper 21521, National Bureau of Economic Research, September

Galor, Oded, and Omer Ozak (2015) 'Land productivity and economic development: Caloric suitability vs. agricultural suitability.' Technical Report, Brown University, Department of Economics

Gennaioli, Nicola, Rafael La Porta, Florencio Lopez de Silanes, and Andrei Shleifer (2013) 'Human Capital and Regional Development.' *The Quarterly Journal of Economics* 128(1), 105–164

Gollin, Douglas, David Lagakos, and Michael E. Waugh (2014) 'The agricultural productivity gap.' *The Quarterly Journal of Economics* 129(2), 939–993

Henderson, J. Vernon, Tim L. Squires, Adam Storeygard, and David N. Weil (Forthcoming) 'The global spatial distribution of economic activity: Nature, history, and the role of trade.' *Quarterly Journal of Economics*

Hijmans, Robert J., Susan E. Cameron, Juan L. Parra, Peter G. Jones, and Andy Jarvis (2005) 'Very high resolution interpolated climate surfaces for global land areas.' *International Journal of Climatology* 25(15), 1965–1978

Mellinger, Andrew, Jeffrey D. Sachs, and John Luke Gallup (2000) 'Climate, coastal proximity, and development.' In *The Oxford Handbook of Economic Geography,* ed. Gordon L. Clark, Maryann P. Feldman, and Meric S. Gertler pp. 169–194

Mitton, Todd (2016) 'The wealth of subnations: Geography, institutions, and within-country development.' *Journal of Development Economics* 118, 88 – 111

Nunn, Nathan, and Diego Puga (2012) 'Ruggedness: The blessing of bad geography in Africa.' *The Review of Economics and Statistics* 94(1), 20–36

Organization, World Health, and UNICEF 2006 'Core questions on drinking-water and sanitation for household surveys.' Geneva, Switzerland

Rappaport, Jordan, and Jeffrey D Sachs (2003) 'The United States as a Coastal Nation.' *Journal of Economic Growth* 8(1), 5–46

Rutstein, Shea Oscar, and Guillermo Rojas (2006) 'Demographic and health surveys methodology.' In 'Guide to DHS Statistics' (Calverton, Maryland: Demographic and Health Surveys)

Shah, Manisha, and Bryce Millett Steinberg (2017) 'Drought of opportunities: Contemporaneous and long term impacts of rainfall shocks on human capital.' *Journal of Political Economy* 152(2), 527–561

Young, Alwyn (2013) 'Inequality, the urban-rural gap, and migration.' *The Quarterly Journal of Economics* 128(4), 1727–1785