Grantham Research Institute on Climate Change and the Environment

Reconciling conflicting evidence on the origins of comparative development: A finite mixture model approach

Thomas K.J. McDermott

September 2013

Grantham Research Institute on Climate Change and the Environment

Working Paper No. 130

The Grantham Foundation for the Protection of the Environment

Global Green Growth Institute

The London School of Economics and Political Science

The Grantham Research Institute on Climate Change and the Environment was established by the London School of Economics and Political Science in 2008 to bring together international expertise on economics, finance, geography, the environment, international development and political economy to create a world-leading centre for policy-relevant research and training in climate change and the environment. The Institute is funded by the Grantham Foundation for the Protection of the Environment and the Global Green Growth Institute, and has five research programmes:

1. Global response strategies
2. Green growth
3. Practical aspects of climate policy
4. Adaptation and development
5. Resource security

More information about the Grantham Research Institute on Climate Change and the Environment can be found at: http://www.lse.ac.uk/grantham.

Reconciling conflicting evidence on the origins of comparative development: A finite mixture model approach

Thomas K.J. McDermott

Grantham Research Institute on Climate Change and the Environment, London School of Economics, Houghton St., London WC2A 2AE, United Kingdom.

Email: T.McDermott@lse.ac.uk

September 16, 2013

Abstract

In this paper, I revisit the controversy over the fundamental sources of comparative development. In contrast to much of the previous literature, my focus is on the appropriate specification of the empirical strategy. Using a finite mixture model approach and Monte Carlo simulations, I demonstrate that the standard linear estimation strategy may be mis-specified and as a result is likely to obscure the true effects of the variables used to explain cross-country income differences. My findings could potentially reconcile apparently conflicting results from the existing literature on the role of geography and institutions in comparative development.

Keywords: *comparative development, institutions, geography, finite mixture models, Monte Carlo simulations*

JEL Classification: O11; O43; O44; O57; Q54; Q56

1 Introduction

In this paper I revisit the controversy over the origins of comparative economic development. For some time now, the debate within the macroeconomic literature on the *fundamental* sources of comparative development has been polarized into essentially two camps; on one side are advocates of geography- or endowments-based theories of development (e.g. Diamond, 1997; Gallup *et al.* , 1999; Masters & McMillan, 2001; Sachs *et al.* , 2001), and on the other, are advocates of institutions-based explanations (e.g. Acemoglu *et al.* , 2001, 2003; Rodrik, 1999; Easterly & Levine, 2003).

A particularly well known and frequently cited contribution to the latter camp is Acemoglu *et al.* (2001), henceforth AJR. That paper made the case that the income gap between rich and poor countries is primarily attributable to differences in their economic institutions (as proxied by security of property rights) and that geography (as proxied by latitude) has no direct effect on income, once institutions are 'properly controlled for'. Based on these results, Acemoglu (2008, p.162) goes as far as to claim that 'there appears to be no causal effect of geography on prosperity today (though geography may have been important historically in shaping economic institutions).'[1] In contrast, it has been shown that even within countries, the observed relationship between latitude and income persists (Parker, 2000), as does the observed correlation between temperature and income (Nordhaus, 2006; Dell *et al.* , 2009), and that disease environment has a direct impact on income, even when controlling for institutions (Sachs, 2003; Carstensen & Gundlach, 2006; Bhattacharyya, 2009b).

Much of the recent debate between these competing camps has concentrated on identification strategies, e.g. concerns over data, sample, choice of variables, instrument quality etc.[2] In contrast, my focus is on the appropriate specification of the empirical analysis (analagous to Cervellati & Sunde, 2011a,b). Observations of country level income per capita appear to be clustered around two distinct modes (see Figure 1), as predicted by modern theories of economic growth that include various forms of poverty traps or non-convexities.[3] With this in mind, it is likely that the determinants of income will have non-monotonic effects across the two growth regimes. Failing to take account of such bi-modality in the underlying distribution of the dependent variable can therefore obscure the true effects of the explanatory variables (as demonstrated by Conway & Deb, 2005; Cervellati & Sunde,

[1] Quote and page number refer to a preliminary draft of the book manuscript, Version 2.2, October 2007.

[2] For example, a recent *Comment* (Albouy, 2012) has criticised the methods and data used in constructing the settler mortality variable in Acemoglu *et al.* (2001), while Auer (2013) argues that the identification strategy employed by these authors confounds the historical determinants of institutions with the direct effect of endowments on development.

[3] The bi-modality of global income was originally identified by Quah (1996, 1997) and since confirmed by, for example, Bloom *et al.* (2003). For a review of the poverty trap literature see Azariadis & Stachurski (2005).

2011a,b). However, to date, few if any authors have taken account of the bi-modality of the global income distribution in their analyses of comparative development.[4] In this short paper I use a flexible empirical strategy - a finite mixture model (FMM) approach - and Monte Carlo simulations to demonstrate how ignoring the bi-modal distribution of global income may have affected existing estimates of the effects of geography and institutions on income.

While both geography and institutions have a plausible causal relationship with development, it may be that their effects are apparent at different stages of development (as argued recently by Bhattacharyya, 2009a). In particular, we might expect that factors related to geography will be important in the early stages of development, but become less so over time as economies diversify to become less dependent on income derived directly from the environment. Similarly, it has been argued that institutional differences cannot explain differences in income amongst relatively poor countries.

The results of the FMM regressions seem to support the hypothesis that geography matters in the early stages of development, while for richer countries, institutions are important in sustaining development gains. The Monte Carlo simulations further demonstrate how the standard linear estimation approach could make geography appear unimportant, even when the true relationship between income and geography is large and significant, at least for a sub-set of (relatively poor) countries. These findings potentially reconcile the contrasting results of previous studies on the fundamental sources of comparative development.

2 Data and methods

With the intention of keeping the focus here on the methodology, I use an established set of data in my analysis, based on Acemoglu *et al.* (2001).[5]

[4]Bloom *et al.* (2003) use a finite mixture model to show that a poverty trap model is a better fit to the data than simple geographic determinism. Cervellati & Sunde (2011a,b) also apply these methods to the question of the influence of life expectancy on growth. Similarly Tol (2011), employs this empirical strategy to test the demo-economic model of Strulik (2008). Tol also finds evidence of a bi-modal income distribution, with hotter countries being poorer, with higher mortality and higher fertility. However, none of these studies has directly addressed the comparative development literature that I am interested in, and in particular the debate over the relative contributions of geography and institutions to the process of long-run growth.

[5]For simplicity of exposition, I focus on AJR's main specification, which involves regressions of log income per capita (in 1995, PPP) on the instrumented institutions variable and latitude, and concentrate on their 'base sample' of 64 countries. The data were ob-

The use of a finite mixture model (FMM) approach has the advantage of allowing the identification of heterogeneity that might otherwise be overlooked, without having to assign observations into groups a priori. The results therefore are entirely data driven, as regime sorting is endogenous in the model.

Based on the preceding discussion and the evidence of a bi-modal global income distribution, I assume two regimes in the data, which I label *rich* and *poor*, for convenience. The density function for global income (y) is then

$$\sum_{j=1}^{C} \pi_j(\boldsymbol{z}) f_j(y \mid \boldsymbol{x}; \boldsymbol{\beta_j}) \tag{1}$$

where $\boldsymbol{x}$ represents the explanatory variables (latitude and institutions) and π_j is the probability of membership in regime $j = rich, poor$. It is assumed that $0 < \pi_j < 1$ and $\sum_{j=1}^{C} \pi_j = 1$. The $\boldsymbol{\beta_j}$ are the parameters to be estimated that are expected to differ across regimes. One can also specify covariates $(\boldsymbol{z})$ that determine regime membership. I assume that each regime is normally distributed, i.e.

$$f_j(y \mid \boldsymbol{x}; \boldsymbol{\beta_j}) = \frac{1}{\sqrt{2\pi\sigma_j^2}} exp\left(-\frac{(y_i - x_i\beta_j)}{2\sigma_j^2}\right) \tag{2}$$

The model parameters, including coefficients, mixture probabilities and standard deviation for each regime, are estimated simultaneously by maximum likelihood. The log likelihood function is given by:

$$L = \sum_{i=1}^{N} ln\left(\pi_j f_j(y \mid \boldsymbol{x}; \boldsymbol{\beta_j}\right) \tag{3}$$

3 Results

3.1 Estimation using a finite mixture model

Results of the FMM regressions are reported in Table 1, where I also present results from AJR's original 2SLS approach for comparison. The table shows the sharp contrast in results based on the two methodologies. Column 1 presents the results from AJR's 2SLS regression, which shows that in the second stage, which includes the instrumented institutions variable, latitude has no direct effect on income. However, in columns 2 and 3, using the

tained directly from Daron Acemoglu's website and are described in detail in Acemoglu *et al.* (2001).

FMM methodology, the effect of institutions on income is only significant in the rich country component, while latitude is a significant determinant of income for poor countries. In columns 4, 5 and 6 I repeat the analysis, using bootstrap methods to calculate standard errors.[6] Again, the institutions variable is significant only for the rich country component. Latitude is no longer significant in either component when using the bootstrap method to calculate standard errors.

3.2 Monte Carlo Simulations

In this section I use Monte Carlo simulations to examine directly the influence on estimation results of a finite mixture distribution.[7] I generate artificial income data, based on the estimated parameters from the finite mixture model regression reported in Table 2 (columns 2 and 3). In other words, I assume that the estimated parameters from the FMM regression are the 'true' population parameters. I then estimate income regressions by OLS for a range of assumptions about the underlying distribution from which the data were generated. I contrast estimation results for the baseline case where the data are generated from a single regime versus cases that involve a mixed distribution.[8] This approach has the advantage of allowing abstraction from issues of data or instrument quality, isolating the effects on regression results of estimating a linear model on data drawn from a bi-modal distribution.

The results of the Monte Carlo simulations are reported in Table 2. Not surprisingly, for the case with all observations assigned to the poor regime (rows 1 and 5), the coefficient on latitude is estimated very precisely, and the null of a coefficient on latitude equal to zero is rejected 100% of the time. However, as I increase the proportion of countries randomly assigned to the rich regime, the estimates of the coefficient on latitude become less precise and the test begins to lose power. In the first set of simulations (rows 1-4), where I assume that institutions are only significant in the rich regime, this process is relatively gradual. However, by the time I have assigned just 15% of observations to the poor regime, we see some negative values for the

[6]As noted by Conway & Deb (2005), the inclusion of a generated regressor–the instrumented institutions variable–invalidates the standard errors obtained using the ordinary least squares formula for the second stage in the 2SLS regression. This issue is corrected by the standard IV estimation software (e.g. the *ivreg* command in Stata). However, an analogous correction does not exist for finite mixture models. Standard errors are therefore calculated for the FMM regression using the bootstrap method. Bootstrapped standard errors are also reported for the 2SLS regression for comparison.

[7]The procedures I follow in generating the Monte Carlo simulations are based on Conway & Deb (2005).

[8]Details of the Monte Carlo procedure are included in the appendix.

coefficient on latitude and the null hypothesis is rejected just one in five times at the 10% level of significance.[9]

In the second set of simulations (rows 5-8), where it is assumed that institutions are significant within both regimes, this loss of precision and power occurs much more rapidly. Even with 85% of countries still assigned to the poor regime, we start to observe negative estimates of the coefficient on latitude, while the null is rejected less than half of the time, even at the 10% significance level.

These simulation results show how the standard linear estimation approach could make latitude appear unimportant, even when the true relationship between income and latitude is large and significant, at least for a sub-set of relatively poor countries.

4 Discussion and conclusions

The findings in this paper do not invalidate the AJR results (nor was that the aim). On the contrary, the FMM results presented above confirm that institutions are a robust and significant predictor of income, but only for a sub-set of countries (i.e. only within the 'rich' regime). On the other hand, the AJR methods and data do not appear to explain the variation in income across relatively poor countries (i.e. the poor regime in the FMM regressions).

The FMM results also suggest that latitude may exert a significant influence on income within the poor country regime, although this finding is not robust, as demonstrated by the bootstrapped FMM results. Sample size is clearly an issue here, and could be responsible for the lack of robustness on the latitude variable in the FMM regressions. One way forward for the analysis of questions relating to the fundamental sources of comparative development may be to use sub-national data in order to increase sample size.[10]

The results presented in this paper demonstrate the importance for anal-

[9] The results of the FMM regression using AJR's data, reported above, show a relatively small proportion of their sample classified as poor (in this case, just 8% of countries). However, in previous studies that demonstrate the bi-modality of global income, as cited above, a much larger proportion of countries - often around 85% - tend to be classified as poor. This is also the pattern observed in Figure 1, which shows a majority of countries clustered around the lower mode of the income distribution. I therefore experiment with a wide range of values for π_1.

[10] Nordhaus (2006) makes the point that moving from country level data, which generally provide about 100 observations, to the G-Econ (sub-national) dataset, which covers over 25,000 grid cells, "is analogous to pictures from the Hubble telescope, which provide clear and crisp answers to many previously difficult and fuzzily answered questions" (p.3510).

yses of comparative development of accounting for the bi-modality of global income. My findings also suggest that some previous literature may have been premature in discounting the role of geography in comparative development. The analysis lends support to the idea that geography and institutions may both be relevant factors in explaining comparative development, but that their effects on income operate at different stages of development. This approach could potentially reconcile seemingly contradictory findings in the existing literature in relation to the relative merits of institutions and geography as the *fundamental* sources of comparative economic development.

Acknowledgements

This paper was largely written while I was an Ussher Fellow and a Research Student Associate at the Institute for International Integration Studies, Trinity College Dublin, and forms part of research towards a doctoral dissertation in the School of Business, Trinity College Dublin. My thanks to Frank Barry and Richard Tol for their advice and support throughout. I am grateful for support from the Grantham Foundation for the Protection of the Environment, and the Economic and Social Research Council (ESRC) through the Centre for Climate Change Economics and Policy. For helpful comments I thank participants at the 2013 Irish Economic Association conference in Maynooth and the 2013 Centre for the Study of African Economies conference in Oxford. I would also like to thank Stephen O'Neill, Partha Deb and Raphael Auer for their generous responses to my queries. All remaining errors are my own.

References

Acemoglu, Daron. 2008. *Introduction to modern economic growth.* Princeton University Press.

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

Acemoglu, Daron, Johnson, Simon, Robinson, James Alan, & Thaicharoen, Yunyong. 2003. Institutional causes, macroeconomic symptoms: volatility, crises and growth. *Journal of Monetary Economics*, **50**(1), 49–123.

Albouy, David Y. 2012. The Colonial Origins of Comparative Development: An Empirical Investigation: Comment. *American Economic Review*, **102**(6), 3059–3076.

Auer, Raphael A. 2013. Geography, institutions and the making of comparative development. *Journal of Economic Growth*, **18**, 179–215.

Azariadis, Costas, & Stachurski, John. 2005. Poverty Traps. *Chap. 5 of:* Aghion, Philippe, & Durlauf, Steven (eds), *Handbook of Economic Growth*, vol. 1. Elsevier.

Bhattacharyya, Sambit. 2009a. Institutions, diseases, and economic progress: a unified framework. *Journal of Institutional Economics*, **5**(1), 65–87.

Bhattacharyya, Sambit. 2009b. Root Causes of African Underdevelopment. *Journal of African Economies*, **18**(5), 745–80.

Bloom, David E., Canning, David, & Sevilla, Jaypee. 2003. Geography and Poverty Traps. *Journal of Economic Growth*, **8**(4), pp. 355–378.

Carstensen, Kai, & Gundlach, Erich. 2006. The Primacy of Institutions Reconsidered: Direct Income Effects of Malaria Prevalence. *The World Bank Economic Review*, **20**(3), 309–339.

Cervellati, Matteo, & Sunde, Uwe. 2011a. Disease and development: The role of life expectancy reconsidered. *Economics Letters*, **113**(3), 269–72.

Cervellati, Matteo, & Sunde, Uwe. 2011b. Life expectancy and economic growth: The role of the demographic transition. *Journal of Economic Growth*, **16**, 99–133.

Conway, Karen S., & Deb, Partha. 2005. Is prenatal care really ineffective? Or, is the 'devil' in the distribution? *Journal of Health Economics*, **24**, 489–513.

Dell, Melissa, Jones, Benjamin F., & Olken, Benjamin A. 2009. Temperature and Income: Reconciling New Cross-Sectional and Panel Estimates. *American Economic Review: Papers and Proceedings*, **99**(2), 198–204.

Diamond, J.M. 1997. *Guns, germs, and steel: the fates of human societies.* W.W. Norton.

Easterly, William, & Levine, Ross. 2003. Tropics, germs, and crops: how endowments influence economic development. *Journal of Monetary Economics*, **50**(1), 3 – 39.

Gallup, John Luke, Sachs, Jeffrey D., & Mellinger, Andrew D. 1999. Geography and Economic Development. *International Regional Science Review*, **22**(2), 179–232.

Masters, William A., & McMillan, Margaret S. 2001. Climate and Scale in Economic Growth. *Journal of Economic Growth*, **6**(3), pp. 167–186.

Nordhaus, William D. 2006. Geography and macroeconomics: New data and new findings. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(10), 3510–3517.

Parker, Philip M. 2000. *Physioeconomics: the basis for long-run economic growth*. MIT Press.

Quah, Danny. 1996. Twin Peaks: Growth and Convergence in Models of Distribution Dynamics. *Economic Journal*, **106**, 1045–1055.

Quah, Danny. 1997. Empirics for Growth and Distribution: Stratification, Polarization, and Convergence Clubs,. *Journal of Economic Growth*, **2**, 27–59.

Rodrik, Dani. 1999. Where Did All the Growth Go? External Shocks, Social Conflict, and Growth Collapses. *Journal of Economic Growth*, **4**, 385–412.

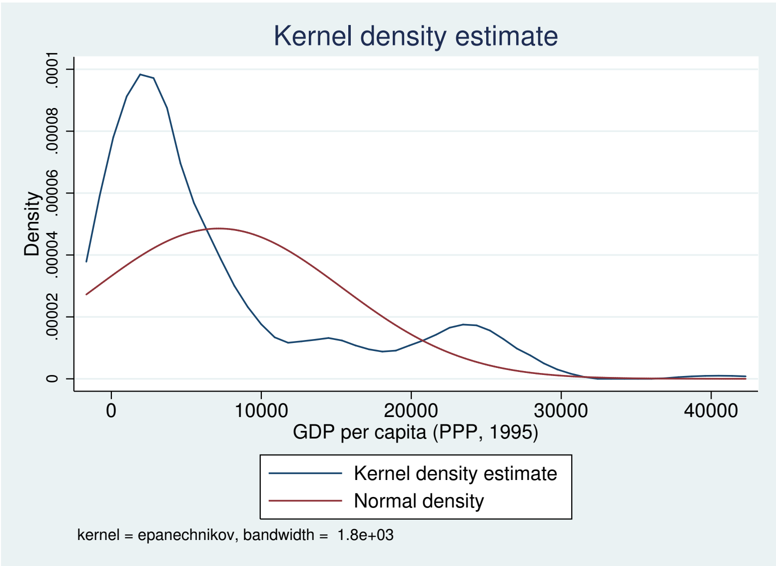
Sachs, Jeffrey D. 2003 (February). *Institutions Don't Rule: Direct Effects of Geography on Per Capita Income*. Working Paper 9490. National Bureau of Economic Research.

Sachs, Jeffrey D., Mellinger, Andrew D., & Gallup, John Luke. 2001. The Geography of Poverty and Wealth. *Scientific American*, **284**, pp. 70–75.

Strulik, Holger. 2008. Geography, health, and the pace of demo-economic development. *Journal of Development Economics*, **86**(1), 61 – 75.

Tol, Richard S.J. 2011 (November). *Poverty traps and climate change*. Working Paper 413. ESRI, Dublin.

Figure 1: Kernel density plot of country-level income per capita, 1995 PPP



The graph is based on income per capita data (1995 US$, PPP) from the Penn World Tables (PWT 7.0) and includes all countries for which these data were available (188 observations). The AJR data represent a sub-sample from this distribution. The graph displays a classic bi-modal pattern, with a large proportion of observations clustered at income levels below $10,000 per capita (PPP), and a significant minority clustered at a higher income level between 20,000 and 30,000.

Table 1: 2SLS vs FMM

	2SLS	FMM		2SLS	FMM	
		Comp. 1 'poor'	Comp. 2 'rich'		Comp. 1 'poor'	Comp. 2 'rich'
	(1)	(2)	(3)	(4)	(5)	(6)
				Bootstrap s.e.	Bootstrap s.e.	Bootstrap s.e.
Avg. protection against exprop. risk 1985-1995	1.00***	-0.10	1.08***	1.00	-0.10	1.08***
	(0.22)	(0.07)	(0.14)	(3.50)	(0.67)	(0.29)
Latitude	-0.65	4.73***	-1.07	-0.65	4.73	-1.07
	(1.34)	(0.45)	(0.83)	(4.83)	(3.19)	(2.10)
Prob. of comp.1 (π_1)		0.08			0.08	
		(0.04)			(0.04)	
Obs.	64			64		

Note: The dependent variable in each case is log GDP per capita in 1995. All data are from AJR.
Models 1, 2 and 3 are standard 2SLS and FMM regressions.
Standard errors in models 4, 5 and 6 are bootstrapped with 1,000 replications.
FMM bootstrapped standard errors based on 927 replications.
(The model did not converge in 73 out of 1,000 attempted replications).
Standard errors in parenthesis. $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

Table 2: Monte Carlo Simulations

β_{11}	β_{12}	β_{21}	β_{22}	π_1	Percentiles of estimated coefficient 5th	50th	95th	Power of test 10%	5%
0.00	1.08	4.73	0.00	1.00	4.56	4.73	4.91	1.00	1.00
0.00	1.08	4.73	0.00	0.85	2.62	4.05	5.17	0.99	0.99
0.00	1.08	4.73	0.00	0.50	0.47	2.39	4.19	0.73	0.63
0.00	1.08	4.73	0.00	0.15	-0.99	0.75	2.44	0.22	0.13
1.08	1.08	4.73	0.00	1.00	4.56	4.73	4.91	1.00	1.00
1.08	1.08	4.73	0.00	0.85	-0.94	4.11	8.74	0.41	0.29
1.08	1.08	4.73	0.00	0.50	-4.27	2.35	9.10	0.17	0.09
1.08	1.08	4.73	0.00	0.15	-4.16	0.49	5.98	0.12	0.06

Note: β_{11} and β_{12} are the parameters on the institutions variable for regimes 1 and 2.
β_{21} and β_{22} are the parameters on latitude for each regime.
π_1 is the probability of a country being assigned to regime 1 (the poor regime).

Appendix - Monte Carlo simulation details

The Monte Carlo simulations are carried out as follows: First, I take the results of the FMM regression in Table 1 (columns 2 and 3) and use these as the chosen 'true' parameter values in generating the artificial income data. I also choose values for π_1, the probability of a country being in regime 1 (the poor regime), starting with the baseline case of $\pi_1 = 1$ (the non-mixture case).

I then take the actual latitude data and the predicted values of the institutions variable and randomly assign each observation to one of the two regimes (according to the chosen probability π_1). Using these data for the explanatory variables, the artificial income data are then generated for each observation, based on the chosen parameters for each regime, plus a normally distributed random error term (ϵ_j) with the finite mixture estimated variance (σ_j^2). The artificial income data are then used to estimate a regression of income on institutions and latitude using OLS. This estimation procedure assumes that all observations come from the same distribution. However, this is only 'true' for our generated data for the baseline case of $\pi_1 = 1$.

This process of generating artificial income data and using these to estimate coefficients on latitude and institutions is repeated 2000 times, in order to generate a large sample of results. I report the 5th, 50th and 95th percentile values from our sample of OLS results, for the estimated latitude coefficient. I also report the proportion of times (out of our sample of 2000 replications) that the estimates lead the hypothesis of no effect of latitude on income (i.e. the latitude coefficient is 0) to be rejected at the 10 and 5% levels of significance (power of the test). This shows how likely one is to find a statistically significant effect of latitude on income. This entire process is repeated for a range of values of π_1.