A cross-country analysis of the roles of border openness, human capital and legal institutions in explaining economic development

Olayinka Oyekola

INTO University of Exeter, Stocker Rd, Exeter EX4 4PY

Abstract

This paper studies the income effect of three specific policy variables: border openness to migration, accumulation of human capital, represented by the education level of the adult population, and the strength of legal institutions, captured by the confidence of a country's citizens to abide by its laws. Using cross-country data covering all regions of the world, and employing instrumental variables for migration, education, and law, we establish that all three policy measures have a robust, positive, and strong association with income. This paper then considers whether there are any germane complementarities between migration, education, and law in explaining income per capita. Our findings show that the effect of improving the education outcomes of countries on their economic performance appears to be similar, regardless on their levels of immigration and law, but that the income effects of border openness and legal institutions can be substantially raised with appropriate institutional and educational policy reforms.

Key words: Migration, human capital, institutions, policy complementarity, economic development

JEL: E61, F22, I25, O43

1. Introduction

Policy matters for economic development. For instance, a country can experience economic development and growth by selecting an optimal level of socio-economic and political integration through openness to trade, migration and active membership in global and regional organisations, such as the World Trade Organisation and the European Union (Dollar 1992; Sachs and Warner 1995; Frankel and Romer 1999; Kaufmann et al. 2003; Alcalá and Ciccone 2004; Docquier et al. 2008). Additionally, countries can dictate how much investment goes into promoting human capital accumulation and, at the same time, shape the building, facilitation and protection of institutional environments that, amongst

Email address: o.oyekola@exeter.ac.uk (Olayinka Oyekola)

other factors, have been identified in the existing literature to be productivity-enhancing and crucial to the way an economy performs (Lipset 1959; North and Thomas 1973; Easterlin 1981; Acemoglu et al. 2001; Djankov et al. 2003; Glaeser et al. 2004; Rodrik et al. 2004).

This paper seeks to contribute to this line of research in international economics and international policy debates by exploring the following questions empirically: Does economic development derive from an increase in border openness to migration? Does the accumulation of more human capital lead to higher economic performance? Do institutions beget income? While these factors have been examined in the existing literature, either singularly or as dual combinations, as part of the crucial determinants of income levels (e.g., Gao 2004; Chang et al. 2009; Acemoglu et al. 2014; Ortega and Peri 2014), we observe that they are yet to be investigated as a triad. This paper fills this gap in the existing literature by investigating the relationships between border openness to migration, human capital measured by the education level of the adult population, legal institutions represented by the strengths of the law that citizens follow, and economic development captured by log GDP per capita. Thus, another focus of this paper is to provide an answer to the following question: Are there any germane complementarities between the three explanatory variables of interest in determining economic outcomes across countries? Our approach in dealing with the previous question borrows from Freund and Bolaky (2008) and Chang et al. (2009), who studied what complementary reforms were needed to reap the benefits of trade openness.

To answer the above questions, we use extensive cross-country data from across the globe and find that each of migration, education, and the law is strongly positively correlated with higher income per capita, even when we have controlled for standard additional regressors in the existing literature. In particular, the evidence marshalled in this paper shows that an increase of one standard deviation in migration, education, and law is, respectively, associated with 0.42, 0.41, and 0.57 standard deviation increases in log income. We then establish that these effects are robust by utilising predicted migration as an instrumental variable for migration, the number of Protestant missionaries for education, and the fraction of the population speaking a European language and legal origins for law. This two-stage least squares (2SLS) regression helps to overcome the omitted variables, endogeneity, and measurement error problems that are apparent in the ordinary least squares (OLS) regression. Finally, this paper considers whether there are any germane complementarities between migration, education, and law in explaining economic development. Our results show that the effect of improving the education

outcomes of countries on their economic performance appears to be similar, regardless on their levels of immigration and law, but that the income effects of border openness and legal institutions can be substantially raised with appropriate legal and education reforms.

2. Data

In this section, we describe our data sources, provide main variable definitions, explain the controls and outline the instruments utilised in the empirical work presented in section 3 below. All data are from Ortega and Peri (2014), except we state otherwise. Interested readers are, thus, encouraged to see their work (e.g., their Table 1, p. 236) for additional data sources. The dependent variable is the natural logarithm of real GDP per capita (PPP-adjusted) expressed in 2000 US Dollars, following the practice in the Penn World Tables. This variable is commonly employed in the existing literature because it captures more correctly comparable cross-country measure of economic development. The independent variables we are interested in are the proxies for border openness, human capital and legal institutions, which to be succinct, are henceforth referred to as migration, education, and law, respectively. Migration is the fraction of the population born in a country other than that in which they live; education is measured using the average number of years of schooling attained by the adult population (i.e., persons aged 25 and above); and law is a variable that captures the degree to which citizens have confidence in and abide by the rules governing a country.

We follow the existing literature in also conditioning on a vector of controls, which consist of regional dummies, geographical features, cultural values, and colonial history (Acemoglu et al. 2014; Ortega and Peri 2014). The variables in this vector have importantly been found to possess deep historical origins that are very unlikely to be affected by our three explanatory variables. Regional dummies are summarised in East Asia, Latin America, and Sub-Saharan Africa. Each of these variables takes a value of 1 if a country is located in it and 0 otherwise (Rodriguez and Rodrik 2001; Rodrik et al. 2004). We use a number of variables to indicate geographical features.[1] These include the logarithm of a country's land area, the latitude of country centroid measuring the distance from the Equator of its capital city, the share of tropical land, minerals measured as the first principal component of the dummy variables for the presence of gold, iron, silver, zinc, and oil reserves, a dummy variable for whether a country is landlocked, mean distance to the coast, mean yearly temperature, mean yearly

[1] The note attached to each table of results describes the relevant geographical endowment being used.

humidity, and indexes for the quality of soil and the incidences of malaria and yellow fever. Cultural values are captured by three fractionalization indexes computed for the existence and distribution of different ethnic tribes, languages spoken, and religious beliefs from Alesina and La Ferrara (2005). For colonial history, we include dummy variables that equal 1 if a country was ever a colony of England or France and 0 otherwise and the fraction of population of European origin at the end of the nineteenth century.

Next, we describe the plausibly exogenous sources of variation in migration, education, and law that are needed as we attempt to circumvent the potential identification problem apparent in treating the explanatory variables exogenously. We used the same instrument for migration as Ortega and Peri (2014), while the instruments for education and law are based on the works of Hall and Jones (1999) and Woodberry (2004, 2012). In particular, Ortega and Peri (2014) adapted the Frankel and Romer (1999) approach to build a gravity-based instrument for migration. The instrument is obtained in two steps. First, Ortega and Peri fit the data on bilateral migration shares for persons 25 years and above in the year 2000 from Docquier et al. (2010) to data on bilateral geographical and cultural characteristics of countries taken from the BACI dataset of CEPII. Second, the instrument–predicted migration–is obtained by aggregating the geography- and culture-based value of bilateral migration shares for each country i across other countries j, using $\widehat{\mathcal{M}_i} = \sum_{j \neq i} e^{\widehat{\gamma}_{\mathcal{M}} Z_{ij}}$, where $\mathcal{M}$ is the migration share predicted by bilateral costs and Z is a vector of independent variables, γ denoting the corresponding vector of coefficients.[2] The notion behind this instrument is that the independent variables (Z) are exogenous to the income of country i, such that if actual migration and predicted migration are significantly related, the latter may be employed as an instrument for the former.

For education, we follow Acemoglu et al. (2014) in using the number of Protestant missionaries per 10000 population in 1923 as an instrument. The source of this data is Woodberry (2004, 2012), with Acemoglu et al. augmenting Woodberry's data using information from the World Atlas of Christian Missions (Dennis et al. 1911) for Australia, Canada, Malta, New Zealand, and the United States. Further, these authors gave four reasons why using this measure to instrument for the level of educational attainment in a country may be challenging. They, meanwhile, concluded that once regional dummies, colonial history and institution are controlled for, protestant missions are largely a derivative of idiosyn-

[2] See Ortega and Peri (2014, pp. 233-237) for details. In all our regressions, we have used their predicted migration share that is based on the non-linear model specification.

cratic factors, qualifying it for a role as an instrument for education.[3] Our sample size ranges from one and a half to about two and a half more than Acemoglu et al.'s 62 countries. For this reason, we take their Dennis et al.'s data for the presence of Protestant missionaries and extracted data for all other countries from Woodberry directly. For the third and last explanatory variable, law, the instruments that we use are the fraction of the population speaking English and one of the other major languages from Western Europe: French, German, Portuguese, or Spanish in 1975. This variable was popularised as an instrument by Hall and Jones (1999) and has been adopted in several papers (e.g., Alcala and Ciccone 2004; Rodrik et al. 2004; Ortega and Peri 2014). As European language alone seems to be a weak instrument in our model, we add the legal origin of countries as an additional instrument.[4] The latter is measured by dummy variables that equal 1 if a country's company law or commercial code derives from English common law, French commercial code, German commercial code, Scandinavian commercial code, or Socialist laws, and 0 otherwise.

Table 1 contains an overview of the descriptive statistics of the variables used for analysis. We close this section by providing unconditional associations between the explanatory variables and outcome variable (figure 1). To do this, we split our sample into countries below and above the median values with regards to each of migration (left panel), education (middle panel), and law (right panel). Additionally, we rank the sample of countries according to their income levels into four groups. Then, for each above/below classification of the explanatory variables, we show the fraction of countries in each income quartile. According to this figure, which displays below-median observations in light bars and above-median observations in dark bars, the pattern that emerges is clear: migration, education, and law are all positively correlated with income, signifying their beneficial effect on economic performance. We investigate this more formally in the next section.

[3] See Acemoglu et al. (2014, p. 887) for details. They have also used primary school enrolment rates in 1900 relative to the population aged between 6 and 14 as an additional instrument. We find, however, that this variable lacks predictive power when the three explanatory variables are simultaneously treated as endogenous regressors.

[4] This approach was also used by Freund and Bolaky (2008).

3. Results

3.1. OLS regressions

Our estimation of income *GDP p.c.* is based on the following empirical specification:

$$\ln\ GDP\ p.c._i = \alpha_0 + \alpha_1 \text{migration}_i + \alpha_2 \text{education}_i + \alpha_3 \text{law}_i + \alpha_4 \text{control}_i + \epsilon_i \tag{1}$$

where, for each country indexed by i, equation (1) allows us to analyse the effects of border openness (migration), human capital (education), and legal institutions (law) on economic development. In this model, control represents a vector of additional country-specific characteristics (which, in some specifications, include interaction terms) discussed above, which may affect income, and ϵ is an error term.

We begin to examine the relationship shown in equation (1) in table 2 by considering ordinary least square (OLS) regressions, choosing to display only the coefficient estimates for our three variables of interest and the corresponding robust standard errors in round brackets. Models 1, 2, and 3, which centre, respectively, on the contribution of each of migration, education, and law on log GDP per capita in 2000, do not include any control variables. As shown, all three explanatory variables are individually statistically significant at the 1 per cent level, with all having positive values. There are some disparities, however, in terms of sizes of effects and explanatory powers. While the coefficient estimate on migration is 7.179–with those on education and law being 0.292 and 1.034, respectively–legal institutions come out trumps as the factor that alone can account for the most substantial part of the variation in income–as much as 57.5% can be accounted for compared to 17.3% and 44.4% that border openness and human capital can explain, respectively. Further, model 4 includes all three explanatory variables simultaneously but still does not include any control variables. As before, all three explanatory variables enter positively and remain highly statistically significant. Interestingly, the ranking of the magnitude of their effects on income is preserved. We note that this happens to be the outcome in all the results reported in this paper.

In the rest of the models (5-16) assessed in table 2, we gradually introduce the vector of additional controls that the existing literature has identified to possess the potential to exert an influence on log GDP per capita in 2000 beyond our three explanatory variables. First, we add regional dummies to models 1-4 to obtain models 5-8. Next, we augment these last four models with country-specific

geographical features to realise models 9-12. Lastly, we insert colonial history to the previous four models to get models 13-16. In all models, the estimated coefficients of migration, education, and law remain positive and statistically significant. To further make sense of these coefficients, we examine their economic significance using models 13, 14, and 15, respectively, for migration, education, and law. It is observed that an increase of one standard deviation in each of the main explanatory variable in these models is associated with 0.42, 0.41, and 0.57 standard deviation increases in log income.

In table 3, we report the results from a variety of robustness tests, which we base on the model 16 from table 2. In particular, our approach is two-fold; these are to control for an even larger set of additional conditioning variable set and to exclude influential observations. In models 1-8, which falls in the first category, we sequentially include–in addition to the regional dummies, geographical features, and colonial history–minerals, landlocked dummy, distance to the coast, temperature, humidity, index of soil quality, the incidence of malaria, and incidence of yellow fever, respectively. In each case, we find that the estimated coefficients of migration, education, and law remain positive and statistically significant. In model 9, all the additional control variables in models 1-8 are included at the same time. Again, the estimated effects of migration, education, and law on income remain positive and statistically significant. Besides, model 10 introduces as additional determinants of income the fractionalisation indices for ethnicity, language, and religion, with the results remaining unaffected.

In the remainder of the models (11-18) examined in table 3, we deal with the issue that falls in the second category–excluding influential observations in our sample. In model 11, we alleviate any concerns that may arise regarding the results presented so far being driven mainly by the neo-European countries; hence, we dropped observations for Australia, Canada, New Zealand, and the United States. The estimated coefficients are still very close to our baseline estimates. In models 12-14, we remove observations by implementing more formal procedures: (i) excluding country-observations with absolute standardized residuals greater than 1.96 (model 12); (ii) excluding country-observations with a Cook's distance higher than the rule-of-thumb sill of 4 divided by the number of observations $\mathbb{N}$ (model 13); and (iii) implementing Li's robust regression that assigns influential observations, smaller weights (model 14). Finally, we omit all country-observation for which abs(dfbeta) $>$ (2/sqrt($\mathbb{N}$)) for migration in model 15, education in model 16, law in model 17, and all three together in model 18. In all these exercises, we find that the results remain robust.

3.2. 2SLS regressions

Despite all the several robustness tests carried out in the previous subsection, omitted variables, endogeneity, and measurement error still pose a threat to the reported results. Our identification strategy to resolve these concerns is to estimate the following three equations in the first stage of a two-stage least square (2SLS) regressions:

$$\text{migration}_i = \beta_0 + \beta_1 \text{predicted migration}_i + \beta_2 \text{control}_i + \varepsilon_i \tag{2}$$

$$\text{education}_i = \gamma_0 + \gamma_1 \text{protestant missions}_i + \gamma_2 \text{control}_i + \nu_i \tag{3}$$

$$\text{law}_i = \delta_0 + \delta_1 \text{European language}_i + \delta_2 \text{legal origin}_i + \delta_3 \text{control}_i + \upsilon_i \tag{4}$$

where the exclusion restrictions are that predicted migration, protestant missions, European language, and legal origin determine the extents of migration, education and law (respectively) but are not correlated with the factors in ϵ. We, therefore, present, alongside our 2SLS estimates, several first stage statistics, viz: Anderson-Rubin Wald test, Angrist-Pischke F test, Cragg-Donald Wald F statistic, Kleibergen-Paap Wald rk F statistic, Stock-Yogo weak ID test critical value, Hansen J statistic p-value, and Shea R-squared.

Table 4 presents the results of the second stage regressions along with the first stage test statistics based on models 13-16 of table 2 that condition on regional dummies, geographical features, and colonial history. Following Acemoglu et al. (2014), we have also added the dummy for the different sources of protestant missions in any model involving education. The results indicate that the explanatory variables of interest continue to have strong positive effects on economic performance. In particular, models 1, 2, and 3 of table 4 produce the second stage estimated coefficients of 8.73, 0.457, and 0.815, respectively, on migration, education, and law that are also statistically significant at the 1 per cent level. In model 4, when all the three explanatory variables entered synchronously, we observe that the positive and statistically significant impacts of these variables on the outcome variable are retained. Based on these estimated coefficients, we can perceive that the 2SLS estimates are consistent with the baseline OLS results. Moreover, the estimated coefficients from 2SLS regressions are quantitatively larger than those from OLS regressions, although the reported standard errors form the latter regression technique are generally smaller than in the former regression methodology.

In the bottom half of table 4, we have also reported several first stage statistics, as mentioned

above. Our objective in reporting these various test statistics is to evaluate the quality/validity of the instruments used for analyses. As can be seen, the generated statistics are all largely in support of the strength of the instruments. For example, the reported Kleibergen-Paap rk F statistics, which allows testing the null of jointly weak instruments, ranges from 10.22 to 18.46, suggesting that we can reject the weak instrument null hypothesis. In particular, the tabulated critical values for weak instruments in Stock and Yogo (2005) that are derived under homoskedasticity for Cragg-Donald Wald F statistic is, as is now commonly adopted in the applied literature, is used for the Kleibergen-Paap rk F statistics, which are based on heteroskedasticity. Thus, for an exactly identified 2SLS specification as in models 1 and 2, the reported F statistics imply that we can reject the hypothesis that the IV size bias is greater than 5 per cent at the 5 per cent level of significance.[5] Further, a similar conclusion holds for models 3 and 4 of table 4, where the models are overidentified such that we can also report the Hansen J statistic.

Our goal in the rest of this subsection is to test whether any meaningful relationships exist between migration, education, and law in determining economic development. Hence, we ask: What complementarities matter? To answer this question, we next extend the linear econometric models employed so far by including interaction terms of the main independent variables along the categorisations established in figure 1 for the explanatory variables. In particular, we examine: (i) how the income effect of migration depends on country differences in education and law; (ii) how the income effect of education depends on country differences in migration and law; and (iii) how the income effect of law depends on country differences in migration and education. Since these specifications now include interaction terms, we need to provide additional instruments for which we use interactions of our primary instruments with the new variables. Table 5, which documents the 2SLS regression results from this exercise, gives more details on the instruments.

In table 5, models 1-2, 4-5, and 7-8 mainly involve interacting one endogenous regressor with dummy variables for high (above the median) and low (below the median) levels of another endogenous regressor. For example, model 1 of table 5 is an augmented version that permits interaction of migration with dummy variables for above-median education and below-median education. In contrast, model 2 involves the interactions of migration with dummy variables for above-median law and below-median law. On

[5] Besides, the values of the F statistics mean that we can also reject the hypothesis that the relative IV bias is greater than 5 per cent at the 5 per cent level of significance. Moreover, these F statistics of exclusion restriction exceed the rule of thumb threshold of 10 suggested by Staiger and Stock (1997)–this is particularly appropriate for models 1 and 2 as the rule of thumb was conceived in the context of models with one endogenous regressor, given that instruments can be weak in models with higher endogenous regressions because of collinearity.

the other hand, models 3, 6, and 9 take the form of a horse race, where we interact one endogenous regressor with dummy variables for high and low levels of the other two endogenous regressors. As an example, model 3 of table 5 is an augmented one that allows the interaction of migration with dummy variables for above-median migration, below-median migration, above median law, and below-median law. Due to collinearity, however, the following interaction terms drop out of the estimated models 3, 6, and 9 (respectively)–the interactions of migration and below-median law, education and below-median migration, and law and above-median education.

The following four observations are worth making. First, the beneficial influence of border openness as measured by migration does not depend on the levels of education in a country. However, the estimated coefficient for migration interacted with above-median education is higher (11.21 *vs* 7.52); they are both statistically significant at the 1 per cent level (model 1). On the contrary, the beneficial effect of migration is affected by the strengths of the legal institutions in a country as represented by law (model 2). In essence, we find that the estimated coefficient on migration interacted with above-median law is positive and highly significant, while the corresponding estimate on migration interacted with below-median law is negative albeit insignificant. This result indicates that poor law countries may find it challenging to make the most of the gains brought by immigrants.

Second, human capital as captured by our education measure seems to have positive impacts that are statistically significant on all countries, with these effects appearing not to depend on the extents of border openness to foreigners (model 4) or the strengths of the law (model 5). Thus, having a more educated country appears to be universally better for economic improvements. Third, the beneficial effects of legal institutions as measured by law are not conditioned by the volume of migration in a country (model 7). As shown, however, the role played by law varies with the levels of education in a country, such that the estimated coefficient on law interacted with above-median education is positive and highly significant, while that of law interacted with below-median education though still positive is not significant.

Fourth and turning to the models with one endogenous regressor interacted with the dummy variables for above the median and below the median levels of the other two endogenous regressors (i.e., models 3, 6, and 9). It is revealed in model 3 that only migration interacted with above-median law with an estimated coefficient of 11.04 is statistically significant at the 1 per cent level, with migration interacted with above-median and below-median education yielding, respectively, positive and negative

point estimates that are not significant. For model 6, although all the estimated coefficients are positive, only the ones for education interacted with the law dummy variables are statistically significant (at least at the 5 per cent level). Lastly, the results in model 9 show that all the estimated coefficients of law interacted with dummy variables for above-median and below-median of both migration and education are statistically significant. In particular, law's beneficial impact is upheld regardless of the levels of migration; however, it is shown that law interacted with below-median education results in the lowering of economic performance. A possible explanation for this is that citizen engagement in political and productive activities are inhibited when a country is consisted of low-level educated individuals, thereby reducing the overall outcome.

4. Conclusion

We have used cross-country regressions to provide answers to the questions introduced in Section 1, and, in line with the existing literature, we estimate that increases in migration, education, and law are strongly associated with higher income, both statistically and economically. In particular, the evidence in this paper reveals that an increase of one standard deviation in migration, education, and law is, respectively, associated with 0.42, 0.41, and 0.57 standard deviation increases in log income per capita.

However, we go beyond the existing literature in two ways. First, to our best knowledge, there has been no research undertaken so far that studies our key variables as a triad policy for determining economic development. Therefore, no systematic attempt has been made to address the concerns regarding omitted variables, endogeneity, and measurement errors. We overcome these problems by utilising predicted migration as an instrumental variable for migration, the number of Protestant missionaries for education, and the fraction of the population speaking a European language and legal origins for law, based on the existing literature (Ortega and Peri 2014; Acemoglu et al. 2014). We establish that the estimated effects from the OLS regressions are robust to using 2SLS regressions. Second, this paper considers whether there are any germane complementarities between migration, education, and law in accounting for economic development. The results show that the impact of improved education outcomes of countries on their economic performance appears to be similar regardless of their levels of immigration and law, but that the income effects of border openness and legal institutions can be substantially raised with appropriate institutional and educational reforms.

Our findings thus imply that income level will rise when governments design and implement good

complementary economic policy reforms that favour freer movement of people, increase the pool of skills, ideas and aggregate human capital, and foster and encourage strong institutional qualities. We conclude, not only that policy matters for economic development, but the interactions between them– border openness to migration, human capital, measured by the education level of the adult population, legal institutions represented by the confidence of a country's citizens to abide by its laws, in our case–have a notable influence on economic development captured by log GDP per capita.

References

Acemoglu, D., Gallego, F. A., & Robinson, J. A. (2014). Institutions, human capital, and development. Annu. Rev. Econ., 6(1), 875-912.

Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. American economic review, 91(5), 1369-1401.

Alcalá, F., & Ciccone, A. (2004). Trade and productivity. The Quarterly journal of economics, 119(2), 613-646.

Alesina, A., & Ferrara, E. L. (2005). Ethnic diversity and economic performance. Journal of economic literature, 43(3), 762-800.

Chang, R., Kaltani, L., & Loayza, N. V. (2009). Openness can be good for growth: The role of policy complementarities. Journal of development economics, 90(1), 33-49.

Dennis, J. S., Beach, H. P., & Fahs, C. H. (1911). World Atlas of Christian Missions. Containing a Directory of Missionary Societies, a Classified Summary of Statistics, an Index of Mission Stations, and Maps Showing the Location of Mission Stations Throughout the World. Edited by James S. Dennis... Harlan P. Beach... Charles H. Fahs... Maps by John G. Bartholomew. Student volunteer movement for foreign missions.

Djankov, S., Glaeser, E., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2003). The new comparative economics. Journal of comparative economics, 31(4), 595-619.

Docquier, F., Faye, O., & Pestieau, P. (2008). Is migration a good substitute for education subsidies?. Journal of Development Economics, 86, 263-276.

Docquier, F., Özden, Ç., & Peri, G. (2011). The wage effects of immigration and emigration. The World Bank.

Dollar, D. (1992). Outward-oriented developing economies really do grow more rapidly: evidence from 95 LDCs, 1976-1985. Economic development and cultural change, 40(3), 523-544.

Easterlin, R. A. (1981). Why isn't the whole world developed?. The Journal of Economic History, 41(1), 1-17.

Frankel, J. A., & Romer, D. H. (1999). Does trade cause growth?. American economic review, 89(3), 379-399.

Freund, C., & Bolaky, B. (2008). Trade, regulations, and income. Journal of development economics, 87(2), 309-321.

Gao, T. (2004). FDI, openness and income. The Journal of International Trade & Economic Development, 13(3), 305-323.

Glaeser, E. L., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2004). Do institutions cause growth?. Journal of economic Growth, 9(3), 271-303.

Hall, R. E., & Jones, C. I. (1999). Why do some countries produce so much more output per worker than others?. The quarterly journal of economics, 114(1), 83-116.

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2003). Governance matters III: Governance indicators for 1996–2002. The World Bank.

Lipset, S. M. (1959). Some social requisites of democracy: Economic development and political legitimacy. American political science review, 53(1), 69-105.

North, D. C., & Thomas, R. P. (1973). The rise of the western world: A new economic history. Cambridge University Press.

Ortega, F., & Peri, G. (2014). Openness and income: The roles of trade and migration. Journal of International Economics, 92(2), 231-251.

Rodrik, D., Subramanian, A., & Trebbi, F. (2004). Institutions rule: the primacy of institutions over geography and integration in economic development. Journal of economic growth, 9(2), 131-165.

Rodriguez, F., & Rodrik, D. (2001). Trade policy and economic growth: a skeptic's guide to the cross-national evidence. NBER macroeconomics annual, 15, 261-325.

Sachs, J. (1995). A. Warner (1995): Economic Reform and the Process of Global Integration. Brookings papers on economic activity, 1, 1-118.

Staiger, D., & Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments. Econometrica, 65(3), 557-586.

Stock, J. H., & Yogo, M. (2005). Testing for Weak Instruments in Linear IV Regression. Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg, 80.

Woodberry, R. D. (2004). The shadow of empire: Christian missions, colonial policy, and democracy in postcolonial societies (Doctoral dissertation, University of North Carolina at Chapel Hill).

Woodberry, R. D. (2012). The missionary roots of liberal democracy. American political science review, 106(2), 244-274.

Table 1 Descriptive statistics

	Mean	Standard deviation	Minimum	Maximum	Observations
Ln GDP p.c.	8.5063	1.3537	4.7641	11.214	184
Migration	0.04341	0.07777	0	0.5165	188
Education	5.7232	3.0593	0.5657	12.490	138
Law	-0.005704	0.9693	-2.2158	1.9635	177
Area	11.337	2.6767	3.2189	16.653	186
Latitude	0.2896	0.1888	0	0.7222	162
Tropics	0.4862	0.4774	0	1	153
British colonial origin	0.3161	0.4665	0	1	155
French colonial origin	0.1548	0.3629	0	1	155
European settlers	28.378	40.971	0	100	153
Minerals	6.846e-09	1.0000	-0.3881	6.7591	148
Landlock	0.1987	0.4003	0	1	156
Coast	452.01	552.12	1.0433	3,418.5	153
Temperature	19.065	8.0962	-4	32	153
Humidity	67.477	16.413	18	97	153
Soil quality	0.2157	0.4126	0	1	153
Malaria	0.2975	0.4028	0	1	150
Yellow fever	0.4808	0.5012	0	1	156
East Asia	0.1011	0.3022	0	1	188
Latin America	0.1543	0.3622	0	1	188
Sub-Saharan Africa	0.2340	0.4245	0	1	188
Predicted migration	0.03781	0.02601	0	0.1604	188
Protestant missions	0.9218	1.6403	0	9.9130	142
European language	31.009	43.013	0	100	149
English legal origin	0.3314	0.4721	0	1	175
French legal origin	0.4400	0.4978	0	1	175
German legal origin	0.03429	0.1825	0	1	175
Scandinavian legal origin	0.02857	0.1671	0	1	175
Socialist legal origin	0.1657	0.3729	0	1	175

Table 2 OLS regression results

Dependent variable is log real GDP per capita (PPP) in 2000, *Ln GDP p.c.*. All regressions include a constant. Robust standard errors in round brackets. Abbreviations: N is no controls; R is regional dummies (East Asia, Latin America, and Sub-Saharan Africa); G is geographical features (Area, Latitude and Tropics); C is colonial history (British colonial origin, French colonial origin, and European settlers). Readers are referred to Section 2 for variable definitions and sources. *** p<0.01, ** p<0.05, * p<0.1.

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Migration	7.179***			3.613***	5.773***			4.180***
	(1.270)			(1.120)	(1.099)			(0.896)
Education		0.292***		0.137***		0.189***		0.0499*
		(0.0281)		(0.0282)		(0.0349)		(0.0283)
Law			1.034***	0.710***			0.822***	0.722***
			(0.0542)	(0.0816)			(0.0560)	(0.0796)
Controls	N	N	N	N	R	R	R	R
Observations	184	138	176	138	184	138	176	138
R-squared	0.173	0.444	0.575	0.691	0.515	0.558	0.746	0.803
Model	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Migration	7.376***			4.168***	7.226***			4.668***
	(0.972)			(0.943)	(0.961)			(0.922)
Education		0.177***		0.0614*		0.182***		0.0770**
		(0.0442)		(0.0324)		(0.0559)		(0.0373)
Law			0.835***	0.717***			0.795***	0.658***
			(0.0666)	(0.0880)			(0.0854)	(0.0994)
Control	R, G	R, G	R, G	R, G	R, G, C	R, G, C	R, G, C	R, G, C
Observations	148	128	147	128	136	118	135	118
R-squared	0.647	0.574	0.765	0.807	0.730	0.647	0.775	0.829

Table 3 Establishing robust relationships

Dependent variable is log real GDP per capita (PPP) in 2000, *Ln GDP p.c.*. All regressions include a constant. Robust standard errors in round brackets. Abbreviations: N is no controls; R is regional dummies (East Asia, Latin America, and Sub-Saharan Africa); G is geographical features (Area, Latitude and Tropics); C is colonial history (British colonial origin, French colonial origin, and European settlers). Readers are referred to Section 2 for variable definitions and sources. *** p<0.01, ** p<0.05, * p<0.1.

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Migration	5.058***	4.612***	4.855***	4.850***	4.899***	4.915***	4.858***	4.829***	4.700***
	(1.041)	(0.908)	(0.920)	(1.006)	(1.034)	(1.022)	(0.886)	(0.907)	(1.029)
Education	0.0894**	0.0853**	0.0878**	0.0982**	0.0845**	0.0878**	0.0742**	0.0867**	0.0757*
	(0.0433)	(0.0369)	(0.0377)	(0.0389)	(0.0387)	(0.0372)	(0.0372)	(0.0376)	(0.0384)
Law	0.670***	0.661***	0.653***	0.662***	0.679***	0.686***	0.641***	0.672***	0.626***
	(0.144)	(0.100)	(0.113)	(0.103)	(0.101)	(0.0998)	(0.102)	(0.102)	(0.111)
Control	R, G, C + first principal component: gold, iron, silver, zinc and oil reserve dummies	R, G, C + dummy for landlockedness	R, G, C + distance to the coast	R, G, C + temperature	R, G, C + humidity	R, G, C + index of soil quality	R, G, C + index of incidence of malaria	R, G, C + index of incidence of yellow fever	R, G, C + all geography together
Observations	111	118	118	116	116	116	118	118	116
R-squared	0.839	0.837	0.832	0.838	0.837	0.836	0.838	0.832	0.856

Model	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Migration	5.289***	4.813***	4.710***	5.034***	4.787***	3.827***	4.850***	4.241***	3.456***
	(0.857)	(0.894)	(0.867)	(0.828)	(0.824)	(1.148)	(0.751)	(0.875)	(0.986)
Education	0.0850**	0.0881**	0.0787**	0.0625**	0.0646**	0.0983***	0.0690**	0.0876**	0.0538*
	(0.0389)	(0.0382)	(0.0353)	(0.0283)	(0.0321)	(0.0338)	(0.0292)	(0.0374)	(0.0272)
Law	0.688***	0.675***	0.638***	0.630***	0.641***	0.654***	0.686***	0.690***	0.738***
	(0.0989)	(0.101)	(0.0914)	(0.0768)	(0.0819)	(0.0966)	(0.0911)	(0.0764)	(0.0666)
Control	R, G, C + fractionalization indices for ethnicity, language and religion	R, G, C + excluding neo-European countries of Australia, Canada New Zealand and the United States	R, G, C + excluding $\mathbb{N}$ when estimated standardized residual > abs(1.96)	R, G, C + excluding $\mathbb{N}$ when Cook's distance > 4/$\mathbb{N}$	R, G, C + Li's robust regression	R, G, C + excluding $\mathbb{N}$ when abs(dfbeta) > (2/sqrt($\mathbb{N}$)) for migration	R, G, C + excluding $\mathbb{N}$ when abs(dfbeta) > (2/sqrt($\mathbb{N}$)) for education	R, G, C + excluding $\mathbb{N}$ when abs(dfbeta) > (2/sqrt($\mathbb{N}$)) for law	R, G, C + excluding $\mathbb{N}$ when abs(dfbeta) > (2/sqrt($\mathbb{N}$)) for migration, education and law
Observations	113	114	113	104	118	107	110	108	97
R-squared	0.845	0.823	0.871	0.902	0.862	0.864	0.866	0.882	0.911

Table 4 2SLS regression results

Dependent variable in the second stage is log real GDP per capita (PPP) in 2000, *Ln GDP p.c.*. All regressions include a constant. Robust standard errors in round brackets. Abbreviations: R is regional dummies (East Asia, Latin America, and Sub-Saharan Africa); G is geographical features (Area, Latitude and Tropics); C is colonial history (British colonial origin, French colonial origin, and European settlers); D is dummy variables for different sources of Protestant missions. Endogenous regressors (instruments) are migration (predicted migration) in model (1), education (protestant missions) in model (2), law (European language and legal origin–English, French, German, Scandinavian, and Socialist–in model (3), and migration, education, and law (predicted migration, protestant missions, European language and legal origin–English, French, German, Scandinavian, and Socialist) in model (4). Readers are referred to Section 2 for variable definitions and sources. †Angrist-Pischke F test in the curly brackets are individual tests of weak identification for migration, education and law, respectively. ††Critical values in the square brackets are for the 10% maximal IV size and 25% maximal IV size, respectively; critical values in the triangular brackets are for the 5% maximal IV relative bias and 30% maximal IV relative bias, respectively. †††Shea R-squared in corner brackets are for migration, education and law, respectively. *** p<0.01, ** p<0.05, * p<0.1.

Model	(1)	(2)	(3)	(4)
Second stage regressions, with first stage test statistics				
Migration	8.730***			4.910***
	(2.146)			(1.836)
Education		0.457***		0.160**
		(0.128)		(0.0765)
Law			0.815***	0.733***
			(0.0816)	(0.274)
Control	R, G, C	R, G, C, D	R, G, C	R, G, C, D
Anderson-Rubin Wald test	7.77	9.89	20.26	6.66
Angrist-Pischke F test†	12.62	11.22	18.46	{2.65, 13.90, 7.14}
Cragg-Donald Wald F statistic	35.84	14.38	16.675	3.959
Hansen J statistic p value			0.6652	0.0097
Kleibergen-Paap Wald rk F statistic	12.62	11.22	18.46	10.215
Stock-Yogo weak ID test critical value††	[16.38, 5.53]	[16.38, 5.53]	[26.87, 8.84]	
			⟨18.37, 5.25⟩	⟨12.20, 4.40⟩
Shea R-squared†††	0.2228	0.1574	0.4140	⌜0.3083, 0.3885, 0.2520⌟
Observations	136	89	133	89

Table 5 Interactions between migration, education, and law and log real GDP per capita

Dependent variable in the second stage is log real GDP per capita (PPP) in 2000, *Ln GDP p.c.*. All regressions include a constant. Robust standard errors in round brackets. Abbreviations: R is regional dummies (East Asia, Latin America, and Sub-Saharan Africa); G is geographical features (Area, Latitude and Tropics); C is colonial history (British colonial origin, French colonial origin, and European settlers); D is dummy variables for different sources of Protestant missions. Endogenous regressors (instruments) are migration*high education and migration*low education (predicted migration*high education and predicted migration*low education) in model (1), migration*high law and migration*low law (predicted migration*high law and predicted migration*low law) in model (2), migration*high education, migration*low education, and migration*high law (predicted migration*high education, predicted migration*low education, and predicted migration*high law) in model (3), education*high migration and education*low migration (protestant mission*high migration and protestant mission*low migration) in model (4), education*high law and education*low law (protestant mission*high law and protestant mission*low law) in model (5), education*high migration, education*high law, and education*low law (protestant mission*high migration, protestant mission*low migration, and protestant mission*high law) in model (6), law*high migration and law*low migration (European language*high migration, European language*low migration, English legal origin*high migration, English legal origin*low migration, French legal origin*high migration, French legal origin*low migration, Socialist legal origin*high migration, Socialist legal origin*low migration, Scandinavian legal origin*high migration, and German legal origin*high migration) in model 7, law*high education and law*low education (European language*high education, European language*low education, English legal origin*high education, English legal origin*low education, French legal origin*high education, French legal origin*low education, Socialist legal origin*high education, Socialist legal origin*low education, and Scandinavian legal origin*high education) in model (8), and law*high migration, law*low migration, and law*low education (European language*high migration, European language*low migration, European language*high education, English legal origin*high migration, English legal origin*low migration, English legal origin*high education, French legal origin*high migration, French legal origin*low migration, French legal origin*low education, Socialist legal origin*high migration, Socialist legal origin*low migration, Socialist legal origin*high education, Scandinavian legal origin*high migration, and German legal origin*high migration) in model (9). Readers are referred to Section 2 for variable definitions and sources. †Angrist-Pischke F test in the curly brackets are individual tests of weak identification for migration, education and law, respectively. ††Critical values in the square brackets are for the 10% maximal IV size and 25% maximal IV size, respectively; critical values in the triangular brackets are for the 5% maximal IV relative bias and 30% maximal IV relative bias, respectively. †††Shea R-squared in corner brackets are for migration, education and law, respectively. *** p<0.01, ** p<0.05, * p<0.1.

Model		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Interacted with:									
Migration	> median education	11.21***		1.339						
		(3.557)		(5.965)						
Migration	< median education	7.515***		-0.798						
		(2.095)		(4.962)						
Migration	> median law		11.41***	11.04***						
			(2.191)	(4.201)						
Migration	< median law		-1.638	d.c						
			(5.474)							
Education	> median migration				0.446***		0.0899			
					(0.124)		(0.0681)			
Education	< median migration				0.349***		d.c			
					(0.132)					
Education	> median law					0.467***	0.362***			
						(0.124)	(0.117)			
Education	< median law					0.419**	0.339**			
						(0.178)	(0.157)			
Law	> median migration							0.828***		0.922***
								(0.0936)		(0.118)
Law	< median migration							1.057***		1.561***
								(0.144)		(0.260)

Table 5 Continued

Model		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Law	> median education								0.992***	d.c
									(0.110)	
Law	< median education								0.405	-0.828**
									(0.291)	(0.347)
Control		R, G, C	R, G, C	R, G, C	R, G, C, D	R, G, C, D	R, G, C, D	R, G, C	R, G, C	R, G, C
Anderson-Rubin Wald test		4.34	12.56	6.16	6.35	6.88	4.56	19.22	12.21	12.52
Angrist-Pischke F test†		4.36, 6.70	19.78, 5.19	2.94, 4.61, 10.58	18.92, 13.57	29.57, 18.37	15.47, 16.79, 11.27	66.69, 18.21	10.50, 8.76	28.06, 12.71, 7.90
Cragg-Donald Wald F statistic		16.729	10.10	7.841	6.054	5.433	3.634	11.805	4.168	1.941
Hansen J statistic p value								0.5109	0.6850	0.2465
Kleibergen-Paap Wald rk F statistic		5.107	3.926	2.981	3.185	4.812	2.946	15.863	7.533	3.498
Stock-Yogo weak ID test critical value††		7.03, 3.63	7.03, 3.63	n.a	7.03, 3.63	7.03, 3.63	n.a	29.32, 9.31	27.51, 8.85	
								18.76, 4.66	18.30, 4.69	18.47, 4.39
Shea R-squared†††		0.25, 0.33	0.35, 0.15	0.23, 0.19, 0.28	0.17, 0.14	0.19, 0.13	0.31, 0.20, 0.14	0.55, 0.53	0.63, 0.28	0.66, 0.39, 0.26
Observations		118	135	118	89	89	89	133	117	117

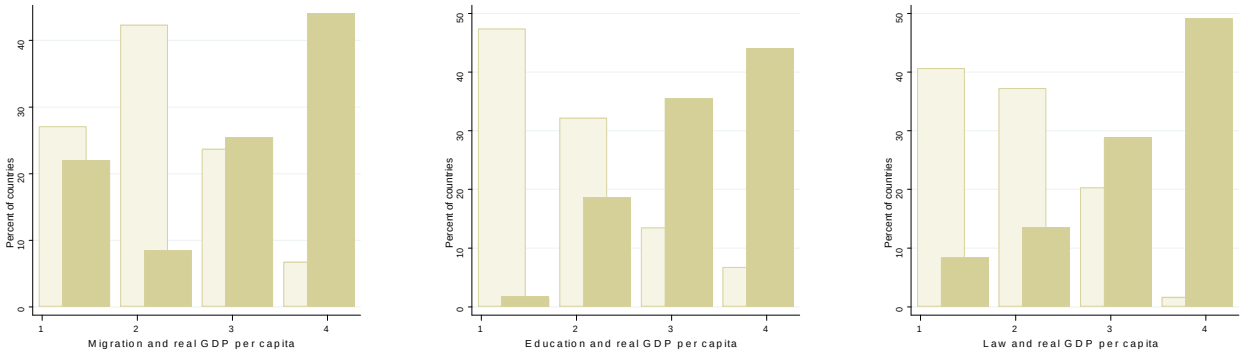

Figure 1 Migration, education, law and log real GDP per capita

Log real GDP per capita (PPP) in 2000 below and above the median of migration, education, and law. Readers are referred to Section 2 for variable definitions and sources. The light bars are for below the median observations, while the dark bars are for those above the median.