

Emigration and Democracy

Frédéric Docquier
Elisabetta Lodigiani
Hillel Rapoport
Maurice Schiff

The World Bank
Latin America and Caribbean Region
Office of the Chief Economist
February 2011



Abstract

Migration is an important yet neglected determinant of institutions. This paper documents the channels through which emigration affects home country institutions and considers dynamic-panel regressions for a large sample of developing countries. The authors find that emigration and human capital both increase democracy and economic freedom. This implies that unskilled

(skilled) emigration has a positive (ambiguous) impact on institutional quality. Simulations show an impact of skilled emigration that is generally positive, significant for a few countries and for many countries once incentive effects of emigration on human capital formation are accounted for

This paper is a product of the Chief Economist, in the Latin America and Caribbean Region. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The author may be contacted at mschiff@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Emigration and democracy*

Frédéric Docquier^a, Elisabetta Lodigiani^b,
Hillel Rapoport^c and Maurice Schiff^d

^a FNRS and IRES, Université Catholique de Louvain

^b CREA, Université du Luxembourg; and Centro Studi Luca d'Agliano

^c CID, Harvard University; Bar-Ilan University; and EQUIPPE

^d World Bank; University of Chile; and IZA-Bonn

JEL codes: O1, F22.

Keywords: Migration, institutions, democracy, diaspora effects, brain drain.

*Corresponding author: Hillel Rapoport, Center for International Development, Kennedy School of Government, Harvard University, 79 JFK Street, Cambridge, MA 02138. Email: hillel_rapoport@hks.harvard.edu. This paper is part of the World Bank Research Program on International Migration and Development. We thank Michel Beine, Eckhardt Bode, David McKenzie, Anna Maria Mayda, Caglar Ozden, Robert Vermeulen, Jeffrey Williamson, and participants at seminars and conferences at the World Bank, Louvain, Luxembourg, Paris I, Maastricht, Boston University, Kiel, Georgetown, Harvard, and the Final Conference of the TOM Marie-Curie Network, Venice, September 2010, for comments and suggestions. We are grateful to Pierre Yared and Cecily Defoort for sharing their data with us.

1 Introduction

Recent research has emphasized the importance of institutions for economic growth and development (Acemoglu, Johnson and Robinson, 2005, Rodrik, 2007) and explored the determinants of institutions.¹ This paper argues that migration is an important determinant of institutions, not considered so far in the economic growth literature.²

Migration first affects institutions by providing people with exit options, thereby changing their incentives to voice (as well as their voicing technology); the existence of an exit option and – for those who stay – the possibility of receiving remittance income tend to act as a safety net that can alleviate social, political and economic pressures to reform. For example, it is commonly argued that emigration to the U.-S. has contributed to delay political change in countries such as Mexico or Haiti.³ On the other hand, once abroad, migrants can engage in political activities (e.g., lobby the host-country government to encourage or block financial aid, or impose economic sanctions) that affect the institutional evolution of their home country, for good or bad. A well-known illustration of this strategy is the very active anti-Castro lobby in the United States which, under the leadership of the Cuban American Nation Foundation, has long succeeded in maintaining a total embargo on economic relations with Cuba. While it is unclear whether this has strengthened the radical or the moderate factions in Cuba, it seems the recent immigrants, who left Cuba more for economic than for political reasons, and the second generation of Cuban-Americans, are more supportive of a dialogue with the communist regime in Cuba and a softening of economic sanctions; and indeed, the Obama administration was able in 2009 to relax restrictions on travel and remittances to Cuba. A lesser known but maybe more effective illustration (in terms of influence on home country politics) is the Croatian diaspora in the United States and Western Europe, which strongly supported secession from the former Yugoslavia and the transition to a market-led economy, provided huge financial support to Tudjman's Croatian Democratic Union (CDU) party and, following the latter's victory in the first post-communist elections

¹For example, Rodrik et al. (2004) show that once institutions are controlled for, geography measures have a weak direct effect on income though they have a strong indirect effect through their impact on the quality of institutions.

²We use the terms "democracy" and "institutions" indifferently as three out of four of our institutional quality indicators are standard democracy indices.

³See for example Hansen (1988) on Mexico and Fergusson (2003) on Haiti.

in 1990, saw its efforts rewarded by the allocation of 12 out of 120 seats at the national assembly to diaspora Croats. Since then, the Croatian diaspora has remained very active, raising funds, organizing demonstrations, petitions, media campaigns and other lobbying activities that proved effective in obtaining official recognition of independence or in shaping European and American attitudes during the Yugoslavia war.⁴ Diasporas may also at times side with a specific group in a conflict that opposes various groups in their country of origin. For instance, Irish Catholics in the US have historically provided financial and other forms of support to the Catholic community in Northern Ireland. However, the continued support provided during the conflict opposing Protestants and Catholics in that country made it more difficult for these communities to reach a peace agreement.⁵

A second channel through which migration affects institutions has to do with the fact that migration is a selective process. Migrants are not randomly selected out of the country's population but tend to self-select along a variety of dimension. First and foremost, migrants are typically positively self-selected on education (migrants' positive self-selection on education is a rule that admits very few exceptions). Given that more educated individuals – and the middle class in general (Easterly, 2001) – tend to have a higher degree of political participation and generally contribute a greater deal to public policy debates, emigration is likely to hurt the quality of domestic institutions as well as the process through which sound policies are formulated and implemented. On the other hand, migration raises the expected return to human capital, thus inducing people to invest more (or more people to invest) in education (Mountford, 1997, Beine et al., 2001, Katz and Rapoport, 2005) and to reallocate talent toward productive and internationally transferable skills (Mariani, 2007);⁶ such effects on the skill distribution can mitigate or even reverse any adverse brain drain impact on political institutions. Another characteristic on which migrants are not randomly self-selected is ethnicity, with an over-representation of ethnic minorities among emigrants. This tends to recompose the home-country population towards

⁴See Eckstein (2009), Haney and Vanderbush (1999, 2005), and Vanderbush (2009) on Cuba, and Djuric (2003) or Ragazzi (2009) on Croatia.

⁵Similar analyses have been proposed notably in the cases of Lebanon and Sri Lanka. Studies providing detailed accounts and analysis of the role of the Irish diaspora include Holland (1999) and Wilson (1995).

⁶Other political economy analyses of the interaction between emigration and institutions in developing countries include Epstein et al. (1999), Docquier and Rapoport (2003) and Wilson (2011).

more homogeneity, again, for good or bad.⁷

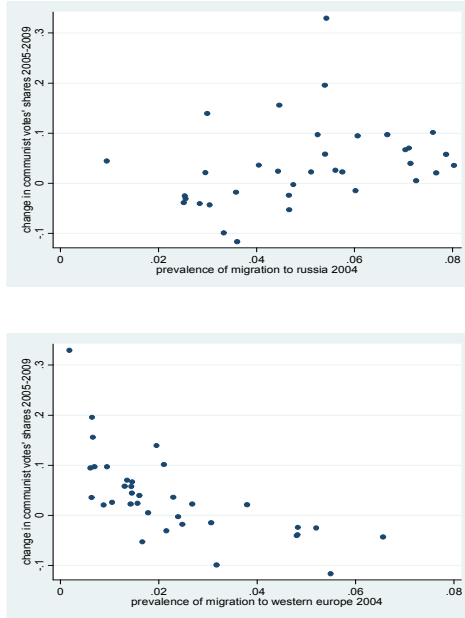
Finally, emigration increases the home country population's exposure to democratic values and norms, be it directly, through contacts with return migrants and relatives abroad, or indirectly, through the broader scope of migration and diaspora networks. Such networks have been shown to foster trade (Gould, 1994, Rauch and Trindade, 2002, Rauch and Casella, 2003, Iranzo and Peri, 2009) and FDI inflows (Kugler and Rapoport, 2007, Javorcik et al., 2011) and to contribute to the diffusion of technology (Kerr, 2008, Agrawal et al., 2011) as well as to the transfer of norms of low fertility (Fargues, 2007, Beine, Docquier and Schiff, 2008) and, in the case of foreign students, to the diffusion of democracy (Spilimbergo, 2009). In particular, Spilimbergo (2009) shows that foreign-trained individuals promote democracy at home, but only if foreign education is acquired in democratic countries. While he does not identify the exact mechanisms through which such an influence may materialize, he suggests a number of possibilities (e.g., access to foreign media, acquisition of norms and values while abroad that diffuse at home upon return, etc.) that can be generalized to other migration experiences.

Two recent micro studies come in support of this claim. The first context we report on is Cape Verde, a nine-island tropical country off the coast of West Africa with a population of half a million, good institutional scores by African standards, and a long tradition of migration (current migrants represent one-fifth of the population, and skilled emigration rates are extremely high).⁸ In this context, Batista and Vicente (2011) set up a "voting experiment" along the following lines: following a survey on perceived corruption in public services, respondents were asked to mail a pre-stamped postcard if they wanted the results of the survey to be made publicly available in the national media. Controlling for individual, household and locality characteristics, they regressed participation in the voting experiment, which they interpret as demand for accountability, on migration prevalence at the locality level. They show that current

⁷In the penultimate paragraph of their article on "artificial states", Alesina, Easterly and Matuzeski (2008) write: "probably the single most important issue that we have not addressed is that of migrations. One consequence of artificial borders is that people may want to move, if they can. ... In some cases, migrations that respond to artificial borders may be partly responsible for economic costs, wars, dislocation of people, refugee crises and a host of undesirable circumstances. ... But sometimes the movement of people may correct for the artificial nature of borders."

⁸Brain drain figures for Cape Verde are 67 percent in Docquier and Marfouk (2006) and remain very high (60 percent) even after excluding people who emigrated before age 18 and acquired their tertiary education abroad (Beine et al., 2007).

Figure 1:
 EMIGRATION TO THE EAST AND WEST AND CHANGE IN COMMUNIST PARTY
 VOTE SHARES BETWEEN 2005 AND 2009, BY DISTRICT
 (based on census data and official election results)



as well as return migrants significantly increase participation rates, and more so for the latter. Interestingly, in the spirit of Spilimbergo’s findings, they find that only migrants to the US seem to make an impact, while migrants to Portugal, the other main destination, do not. In contrast, they do not find evidence of additional effects for skilled migrants. The other context is that of Moldova, a former Soviet Republic with virtually no emigration before 1990 and which has seen a recent surge in migration outflows, estimated at half-a-million for a population of 3.6 million in 2008. The evidence we present for Moldova, which is purely descriptive, comes from the analysis of election outcomes in 2005 and 2009 (Omar Mahmoud et al., 2010). It shows that higher votes for the communist party are associated at the district level with migration to Russia while a negative correlation obtains for migration to the EU. Moreover, changes in the share of votes gained by the communist party between 2005 and 2009 follow the same pattern (see Figure 1) and there is evidence of spillover effects to non-migrant households as the same voting patterns are observed even after excluding households with a (current or past) migrant member.

At a macro level, the only paper attempting to assess the overall effect of emigration on institutions we are aware of is Li and McHale (2009), who use the World Bank

governance indicators (Kauffman, Kraay and Mastruzzi, 2005) (henceforth KKM) and the Docquier and Marfouk (2006) migration data set in their cross-sectional analysis. Focusing on skilled migration, they examine the impact of the brain drain on sending country’s institutional development and find a positive effect on “political” institutions (i.e., on “political stability” and “voice and accountability”) but a negative effect on “economic” institutions at home (i.e., on “government effectiveness”, “regulatory quality”, “rule of law”, and “control of corruption”). However, their results suffer from the limits of a cross-sectional analysis⁹ and, as they themselves acknowledge, from the weakness of their instrumentation strategy (they instrument skilled emigration rates using countries’ geographical characteristics). In this paper we look instead at migration in general, focus on democratic institutions, and consider dynamic-panel regressions. We find that the emigration rate and the level of human capital both positively affect democracy and economic freedom at home. This implies that unskilled migration has a positive impact while skilled migration has an ambiguous impact on institutional quality. Using the point estimates from our regressions, we simulate the marginal effect of skilled emigration on institutional quality. In general the simulations confirm the ambiguous effect of high-skill emigration. It is only when the incentive effects of emigration on human capital formation are taken into account that a significant institutional gain obtains for some countries in the short run, and for many countries in the long-run.

2 Empirical analysis

2.1 General Considerations

Empirical investigation of the effect of emigration on institutions in a cross-section or a panel setting raises a difficult trade-off. In a cross-sectional dimension, it is possible to use better data both for migration and institutional quality. In particular, for migration, it is possible to use the Docquier and Marfouk (2006) data set, which considers international migration by educational attainment. This data set describes the emigration of skilled workers to the OECD for 195 source countries in 1990 and 2000. For institutional quality, the World Bank Governance data by Kaufmann, Kray and

⁹The KKM (2005) data set starts in the late 1990s and is therefore not long enough to allow for panel data analysis. Similarly, the Docquier and Marfouk (2006) dataset offers estimates of emigration rates by skill levels for 1990 and 2000 only.

Mastruzzi (2005) measures six dimensions of governance from 1996 to 2005: voice and accountability, political stability and absence of violence, government effectiveness, regulatory quality, rule of law, and control of corruption. It covers 213 countries and territories for 1996, 1998, 2000, and annually for 2002-2005.

In unreported regressions, we consider OLS regressions using these data sets. We find a significant and positive correlation between the emigration rate and institutional quality indexes, but these regressions suffer from a lot of shortcomings. First, it is difficult to find an appropriate baseline specification, because different economic, political and cultural factors can be important in explaining the quality of institutions. As Alesina et al. (2003) noted, various explanatory variables have been used in the literature on the determinants of institutions, such as log of gdp per capita, legal origin dummies, religious variables, latitude, fractionalisation indices, etc. The main problem with these variables relates to the fact that the pattern of cross-correlations between explanatory variables cannot be ignored and that in many cases the results of cross-country regressions are sensitive to the econometric specification. For example, they point out that their index of ethnic fractionalization is highly correlated with latitude and with the log of gdp per capita (which, in addition, is very likely endogenous). Moreover, legal origin dummies are highly correlated with religious variables etc. In a panel dimension instead, it is possible to control for unobservable heterogeneity and for all time-invariant variables affecting institutional quality.

Another problem of cross-sectional analysis refers to endogeneity and reverse causality problems (i.e., bad institutions can cause migration). Attempting to confront the endogeneity issue directly requires finding a suitable instrument. This is not easy in our context. To properly instrument for migration we need a variable that is correlated with the emigration rate but not directly correlated with our endogenous variable, institutional quality. In the migration literature, country's geographical features are often used to instrument for emigration. However, in the institutions and growth literature, the very same geographical characteristics, such as latitude or country size, are also used as determinants of institutions, which would seem to question their theoretical validity as candidate instruments. Finally, an additional problem in cross-section analyses has to do with the fact that institutional quality is a quite persistent variable, therefore a dynamic model would seem to be more suited to study the relationship between emigration and institutions. Moreover, several papers discuss the influence of education on institutional quality, therefore it is worth to include

in our specifications a variable related to education or to human capital (of course, this would suffer from endogeneity). In the next section, therefore, we will study the impact of emigration on home-country institutions using dynamic-panel regressions. In particular, we will use the system-GMM estimator, and we will be able to control for unobservable heterogeneity and account for endogeneity and persistency of some of the variables, using internal instruments. As far as we know, this is the best suited technique available when it is difficult to find good external instruments, as in our case.

2.2 Panel analysis

We follow the literature on democracy and education (Acemoglu et al., 2005, Bobba and Coviello, 2007, Castello-Climent, 2008) and Spilimbergo’s (2009) study on democracy and foreign education and consider the impact of emigration on institutional quality using dynamic-panel regressions.

2.2.1 The econometric model

As in previous studies on democracy and education, we consider the level of democracy as our dependent variable and we estimate the following dynamic model:

$$\begin{aligned}
 Democracy_{i,t} = & \beta_0 Democracy_{i,t-5} + \beta_1 h_{i,t-5} + \beta_2 emrate_{i,t-5} + \\
 & + \beta_3 X_{i,t-5} + \eta_i + \alpha_t + \varepsilon_{i,t}
 \end{aligned}
 \tag{1}$$

where i is the country, t is the period. All explanatory variables are lagged five years. The lagged dependent enters the set of explanatory variables to account for persistence in democracy scores. Our coefficient of interest is β_2 , which reflects whether emigration (measured by the total emigration) affects democracy at home. The coefficient β_1 captures the effect of human capital on democracy. β_3 is a vector of coefficients reflecting the importance of other control variables such as population size and gdp per capita (both in logs), as in Acemoglu et al. (2005). We also control for time fixed effects, α_t , and country fixed effects, η_i . The advantage of a panel estimation is that it is possible to control for unobservable variables that are country-specific and whose omission in cross-sectional analyses can bias the estimated coefficients. Therefore, the results are robust to all country-specific time invariant explanatory

variables used in the cross-section literature on institutional quality, including ethnic fractionalisation, religions, legal origins, colonial ties, geographical variables etc.

A general approach to estimate such an equation is to use a transformation that removes unobserved effects and uses instrumental variables. The well-known Arellano-Bond (1991) method considers the first-difference of the explanatory variables which are instrumented by their lagged values in levels.¹⁰ Acemoglu et al. (2005) used this method to study the effect of education on democracy without finding any significant effect. One of the shortcomings of this method is that, as Bond, Hoeffler and Temple (2001) point out, the first-difference GMM estimator can behave poorly when time series are persistent and the lagged levels of the explanatory variables turn out to be weak instruments of the explanatory variables in first-difference. In small samples, this can cause serious estimation bias.¹¹ To overcome these problems, Bond et al. (2001) suggest to use a more informative set of instruments within the framework developed by Arellano and Bover (1995) and Blundell and Bond (1998). From our perspective, and given that democracy varies significantly across countries but is quite persistent over time, it is clear that the Blundell and Bond system GMM is most appropriate. New results on the relationship between democracy and education were found using the system GMM estimator.¹² Following this literature, we use the Blundell and Bond system GMM estimator that combines the regression in differences with the regression in levels in a single system. The instruments used in the first differentiated equation are the same as in Arellano-Bond (1991), but the instruments for the equation in level are the lagged differences of the corresponding variables. In order to use these additional instruments, a moment condition for the level equation, which implies that first differences of pre-determined explanatory

¹⁰Under the assumptions that the error term is not serially correlated and that the explanatory variables are weakly exogenous or predetermined (i.e. the explanatory variables are not correlated with future realizations of the error term), the following moment conditions are applied for the first difference equations:

$$E[W_{it-s}(\Delta\varepsilon_{it})] = 0 \text{ for } s \geq 2; t = 3, \dots, T$$

where W_{it-s} are the lagged dependent and all the pre-determined variables in the model.

¹¹Simulation results show that the Difference GMM may be subject to a large downward finite-sample bias when time series are persistent, particularly when T is small. The higher the persistence of the series used as instruments, the weaker the correlation between levels and differences (see Blundell and Bond (1998) for the weak instrumentation problem).

¹²Bobba and Coviello (2007), and Castello-Climent (2008). Splimbergo (2009) also uses system GMM.

variables are orthogonal to the country fixed effects, must be satisfied.¹³

We test the validity of moments conditions by using the test of overidentifying restrictions proposed by Hansen and by testing the null hypothesis that the error term is not second order serially correlated. Furthermore, we test the validity of the additional moment conditions associated with the level equation using the Hansen difference test for all GMM instruments.¹⁴

A particular concern related to this method is the risk of instrument proliferation. In fact, if the use of the entire set of instruments in a GMM context gives significant efficiency gains, on the other hand, a large collection of instruments could overfit endogenous variables as well as weaken the Hansen test of the instruments' joint validity.¹⁵ The instrument proliferation problem is particular important in small samples, but unfortunately there is no formal test to detect it, even if a possible rule of thumb is to keep the number of instruments lower than or equal to the number of groups.¹⁶ In our analysis, we consider the lagged dependent and all the control variables of interest as predetermined, instrumented with "internal instruments", using their own one-period and further lags, according to the specification.

2.2.2 Data

Our data set is a five-year unbalanced panel spanning the period between 1980 and 2005, where the start of the date refers to the dependent variable (i.e., $t = 1980$, $t - 1 = 1975$). In our sample, we are considering only developing countries, and they enter the panel if they are independent at time $t - 1$. The data set employed in our analysis is an updated version of that used by Acemoglu et al. (2005) for the democracy indicators (except for economic freedom) and all the control variables. The migration data come from Defoort (2008).

Democracy

Data on democracy are taken from the Freedom House data set, from the POLITY IV data set, and from the Economic Freedom of the World project (Simon Fraser

¹³For the level equation the following moment conditions are to be satisfied:

$$E[(\Delta W_{i,t-1})(\eta_i + \varepsilon_{i,t})] = 0 \text{ for } t = 4, \dots, T.$$

¹⁴This test is not reported in the tables, but it is available upon request.

¹⁵See Roodman (2009)

¹⁶The `xtabond2` command, implemented in Stata, gives a warning when instruments exceed the number of groups.

Institute).

The Freedom House measures political rights (PR) and civil liberties (CL) using, respectively, an index which ranges from 1 to 7, with a higher score indicating more freedom. The ratings are determined by a list of questions. For the political rights index, for example, the questions are grouped into three sub-categories: electoral processes; political pluralism and participation; and functioning of the government. The civil liberties questions are grouped into four subcategories: freedom of expression and belief; association and organization rights; rule of law and personal autonomy; and individual rights. The sum of each country's sub-category scores translates to a rating from 1 to 7. Following Acemoglu et al. (2005) we transform the indexes so that they lie between 0 and 1, with 1 corresponding to the most-democratic set of institutions.

Another measure of democracy from the POLITY IV data set is considered. Indicators of democracy measure the general openness of political institutions and combines several aspects such as: the presence of institutions and procedures through which citizens can express effective preferences about alternative policies and leaders; the existence of institutionalized constraints on the exercise of power by the executive power; and the guarantee of civil liberties to all citizens in their daily lives and in acts of political participation. In our data set we consider a composite index (Polity2), that ranges from -10 to + 10. This index is also normalized from 0 to 1, with 1 corresponding to the most democratic set of institutions.

Finally, we also consider Economic Freedom of the World (EFW), an index which measures the degree to which countries' policies and institutions support economic freedom. Five broad areas are distinguished: (1) size of government; (2) legal structure and security of property rights; (3) access to sound money; (4) freedom to trade internationally; and (5) regulation of credit, labor and business. This index is also normalized between 0-1.

Migration

For emigration data, we use the estimates provided in Defoort (2008). Focusing on the six major destination countries (USA, Canada, Australia, Germany, UK and France), she computed skilled emigration stocks and rates by educational attainment from 1975 to 2000 (one observation every 5 years). On the whole, the six destination countries represent about 75 percent of the OECD total immigration stock.¹⁷

¹⁷However, for some sending countries, the coverage by the Defoort dataset may be quite low. For

Other data

Data on human capital are based on Barro and Lee (2001). Data on GDP per capita and population data are taken from the PWT and from the World Development Indicators. Data on legal origins are taken from La Porta et al. (1999).

2.2.3 Regression results

Tables 1, 2, 3, 4 present our main general results from estimating equation 1 and using the Freedom House PR and CL indicators, the Polity2 measure from the Polity IV data set, and the Economic Freedom Indicator (EFW). We start by considering as variables of interest the lagged dependent, the total emigration rate, the share of tertiary educated workers over the total resident labor force, and the log of population size.

Column 1 of each table shows the pooled OLS relationship between the total emigration rate and democracy by estimating equation 1. The results show a positive correlation between openness to migration and democracy, statistically significant, however, only when considering the Polity2 and EFW indexes (all standard errors are robust and clustered by country group). In column 2, when we control for fixed effect, the coefficient related to the total emigration rate becomes negative (except for EFW), and statistically not significant. We know that in a dynamic panel data model, the standard fixed effect estimator is biased and inconsistent in panels with a short time dimension (the so called Nickell bias (Nickell, 1981)). Moreover, both in our fixed effect and pooled OLS estimations, explanatory variables are considered as exogenous. To deal with these problems we use the system GMM estimator that is consistent in dynamic panel estimations and rely on "internal instruments" to control for a weak form of exogeneity of all explanatory variables. We consider the explanatory variables of interest as predetermined, i.e. instrumented using their own one-period and further lags, in order to use a relevant number of instruments for efficiency reasons and at the same time keeping the number of instruments lower than or equal to the number of country groups in all specifications.¹⁸ In column (3)

example, Surinamese emigrants mainly live in the Netherlands, with just 3 percent of Surinamese emigrants living in the six receiving countries in Defoort's sample. We will therefore conduct a sensitivity analysis to check the robustness of the results to the exclusion of low-coverage countries in the Defoort dataset.

¹⁸A problem of the GMM estimator is that too many instruments can over fit the endogenous variable. As rule of thumb, the number of instruments should be less or at least equal to the number

of tables 1, 2, 3, 4 the estimates for the total emigration rate are now positive and highly significant at the one percent level for all four indicators. Column (4) shows the same specification, but now reducing the number of instruments for robustness check. Our previous results are confirmed.¹⁹

The share of tertiary educated workers over the total resident labor force, as a proxy for resident human capital, is another variable of interest in our model. As for the total emigration rate, the results show a statistically significant and positive correlation between the share of total educated workers and democracy in pooled OLS regressions. The coefficients turn out to be negative (except for EFW) and not statistically significant in fixed effect regressions. Column (3) of tables 1, 2, 3, 4 shows the SYS GMM estimates. The estimated coefficients of the share of tertiary educated workers are now positive and statistically significant at usual significance levels, except for the Polity2 indicators. The results are confirmed when reducing the number of instruments in column (4).

In our basic specification, we add also as a regressor the logarithm of population size (lagged), which is positive and statistically significant for two indicators out of four when using the SYS GMM estimator. Including population size in our model is important to avoid omitted variable bias. Indeed, population size can affect institutional quality and is often considered as an explanatory variable in the relevant literature (see for example Acemoglu et al., 2005, and Bobba and Coviello, 2007). At the same time, population size is negatively correlated with the emigration rate (big countries have small emigration rates); therefore, including population size is important to make sure the emigration rate is not simply capturing a country-size effect.

Column (5) controls for GDP per capita (in logs). The estimated coefficient of the emigration rate is again positive and statistically significant at 10 and 1 percent when considering the Civil Liberties and Polity2 indicators, but loses its significance when

of groups. We follow this rule even if sometimes, given few data observations and specifications with additional controls, in the reported regressions the number of instruments is slightly higher than the number of groups. For comparative reasons, we show regressions where explanatory variables are instrumented using their one-period to their second or third lags. In unreported regressions, when instruments outnumber the number of groups, for robustness check, we further reduce the number of instruments using only their one period lag. We find that results do not substantially change.

¹⁹In column 3, all the explanatory variables are considered as predetermined and instrumented using their own first to third lags. In column 4, all the variables are instrumented using their own first to second lags.

using the Political Rights indicator and Economic Freedom. The share of tertiary educated workers over the total residence labor force is not significant anymore, probably due to the two variables being highly correlated (0.7145). Finally, the coefficient on the GDP per capita is in general positive but not always significant.²⁰

The estimations confirm that democracy is very persistent. Moreover, considering the first 3 columns in each table, the coefficient on past democracy ranges between the estimated coefficient in pooled OLS, which is usually biased upwards, and the estimated coefficient for the fixed effect, which usually displays a downward bias.

The AR(2) test which tests the null hypothesis that the error term is not second order serially correlated, and the Hansen J test of overidentifying restrictions, indicate that the moment conditions are satisfied and the instruments are valid.

In general, the results appear quite robust across specifications and indices.²¹ To evaluate whether the skill composition of migration, and not just its size, affects institutional quality at home, we introduce in column 6 the share of tertiary educated amongs migrants. The coefficient of the share of tertiary educated migrants is negative but not statistically significant for 3 indicators out of 4 and is only positive and significant at the 10 percent when considering the EFW indicator. In spite or, rather, because of this inconclusive result, we will further investigate this issue in the next section using numerical simulations.

Finally, one may be concerned, as Acemoglu et al. (2005) were about their own study, that the presence of socialist countries in our sample may largely affect the estimation results. Indeed, most socialist countries had high levels of education in the 1980s and did not experience any particularly increase in educational attainments during or immediately after the transition. In addition, prior to the transition, legal emigration was strongly restricted, while after the transition most socialist countries

²⁰In unreported regressions, we also introduce as control variables, the median age of the population, and urbanization rate. While human capital loses its significance, probably because of multicollinearity, the total emigration rate remains significant when considering these additional control variables as exogenous. If they are considered as pre-determined, then the emigration rate also loses its significance too, which may be due either to collinearity or instruments proliferation.

²¹To further assess the robustness of our results, in unreported regressions we considered the total emigration rate divided by a coverage measure in the Defoort (2008) dataset. Recall that the Defoort figures are based on the six major destination countries (USA, Canada, Australia, Germany, UK and France). Comparing the emigration stocks in 2000 in the Defoort data set with those in the Docquier and Marfouk (2006) data set (which is based to 30 OECD destination countries) yields a variable indicating the percentage of coverage of the Defoort data set. Dividing the total emigration rate by this coverage measure does not affect the quality of the results.

experienced a strong increase in emigration. To control for the specific characteristics of these economies, in column (7) of each tables we interact human capital and emigration with legal origin socialist dummies, finding in general a statistically significant effect for the interacted terms, in particular for emigration.²² The interaction term on emigration is negative and significant for all three "political" indicators of democracy, and positive for the "economic" indicator. This suggests that emigration caused socialist regimes to become politically more repressive, an interpretation which fits well with the popular historical accounts of the former Communist bloc. If it is correct, however, it should be relevant only prior to the transition. In column (8) of each tables, we therefore consider the same interaction, but now introduce a dummy variable which takes a value equal to 1 in years before (or equal to) 1990. The magnitude and significance levels of the coefficient are thereby increased, which supports our interpretation of these results.

2.2.4 Robustness

The evidence found in the previous section reveals that human capital and emigration may improve institutional quality. To control for the robustness of these results, for each indicator we consider in table 5 our benchmark specification in a balanced sample. This allows for checking whether the entry and exit of countries from the unbalanced sample may affect our estimates. The results for PR, CL, Polity2 indicators are very similar to those in previous tables. Moreover, now the estimated coefficient for human capital is also statistically significant at 10 percent for the Polity2 indicator. In the case of the Economic Freedom Indicator, the estimates are not reliable due to the fact that too many observations are lost.

Table 6 provides additional robustness checks in a balanced sample when considering non-linear effects for socialist countries as in columns (7) and (8) of tables 1, 2, 3, 4. Again, the estimates are very similar to the previous ones in an unbalanced sample, with more significant results for interacted terms with human capital. As before, in the case of the Economic Freedom Indicator estimates are not reliable, because too many observations are lost.

Finally, in tables 7, 8, 9, 10, socialist countries are excluded from the sample. The results show that our findings are not driven by socialist countries.

²²In the regressions, the legal origin dummy is not introduced by itself, because in SYS-GMM fixed effects are already taken into accounts.

Another concern refers to the presence of oil-exporting countries. Several studies have pointed out a negative correlation between oil export dependence and democracy, with oil endowment appearing as a cause for lower democracy (e.g., Ross, 2001, Tsui, 2010). To control for the specific characteristics of these economies, in table 11 we consider interaction terms with human capital, total emigration rate and a dummy for oil-exporting countries, both in an unbalanced and balanced sample. The estimated coefficients of human capital and the total emigration rate are in general positive and statistically significant across indicators, as in the baseline regressions. Interaction terms with human capital and a dummy for oil-exporting countries are generally negative and statistically significant (with higher coefficients than the estimated coefficient of human capital). This means that, in the case of oil-exporting countries, human capital has a negative impact on institutional quality. Interaction terms with total emigration rate, instead, are positive, but in general not statistically significant (except for the CL indicator).

Finally, another concern is whether Sub-Saharan African countries, which have sometimes unstable political dynamics, may affect our results. Table 12 shows the estimated results when we include interaction terms with a dummy for Sub-Saharan African countries. Again, the estimated coefficients of human capital and total emigration rate are positive and statistically significant across the various specifications and different institutional quality indicators, confirming our results. The interaction terms with human capital are in general not statistically significant while those with emigration are generally positive and statistically significant. This would seem to suggest that African countries tend to benefit more from the institutional gains emigration brings about.²³

²³See the appendix for Tables 7 to 12.

Table 1: Dependent Variable: Freedom House Political Rights Index (PR)

	Pooled		F.E.		SYS		SYS		SYS		SYS	
	OLS (1)	OLS (2)	GMM (3)	GMM (4)	GMM (5)	GMM (6)	GMM (7)	GMM (8)	GMM (9)	GMM (10)	GMM (11)	GMM (12)
PR _{t-5}	0.721*** (0.0381)	0.355*** (0.0520)	0.640*** (0.0623)	0.609*** (0.0658)	0.626*** (0.0605)	0.647*** (0.0576)	0.680*** (0.0556)	0.695*** (0.0628)				
Human capital _{t-5}	0.678*** (0.243)	-0.650 (0.700)	0.642* (0.335)	0.796*** (0.361)	-0.125 (0.426)	0.670*** (0.321)	0.662*** (0.290)	0.659*** (0.294)				
Total emigration rate _{t-5}	0.183 (0.112)	-0.647 (0.661)	0.885*** (0.338)	0.914*** (0.349)	0.304 (0.413)	0.673*** (0.256)	0.513** (0.237)	0.513* (0.277)				
Log population _{t-5}	-0.00865 (0.00663)	-0.443*** (0.158)	0.0478* (0.0273)	0.0485* (0.0284)	-0.00304 (0.0201)	0.0329 (0.0204)	0.00390 (0.0166)	0.00753 (0.0209)				
Log GDP per capita _{t-5}					0.0773** (0.0319)							
Share tertiary ed. migrants _{t-5}						-0.108 (0.126)						
Human capital _{t-5} *Soc. dummy							0.744 (0.611)					
Total emigration rate _{t-5} *Soc. dummy												
Human capital _{t-5} *Soc. dummy*d90												1.393** (0.655)
Total emigration rate _{t-5} *Soc. dummy*d90												-2.351*** (0.591)
Time dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
R-squared	0.611	0.247										
AR(1) test			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AR(2) test			0.542	0.553	0.427	0.571	0.555	0.646	0.555	0.646	0.646	0.646
Hansen J test			0.351	0.302	0.471	0.620	0.836	0.788	0.836	0.788	0.788	0.788
Observations	476	476	476	476	423	476	476	476	476	476	476	476
N. countries		91	91	91	85	91	91	91	91	91	91	91
N. instr.			74	62	76	91	90	86	90	86	86	86

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors clustered by country in parentheses. One step system GMM estimator. The sample is an unbalanced sample comprising data at five year interval between 1980 and 2005. AR(1) and AR(2) are the p-values of Arellano-Bond test for serial correlations. The values reported for the Hansen J test are the p-values for the null hypothesis of instrument validity. All the variables are treated as pre-determined. They are instrumented for using their own first to third lags in columns 3 and 6. They are instrumented for their own first to second lags in all the other columns. In particular, column (4) shows the same specification as column (3), but now reducing the number of instruments till the second lags for robustness check. In addition to these instruments, the system GMM also uses as instruments for the level equations the explanatory variables in the first differences lagged one period.

Table 2: Dependent Variable: Freedom House Civil Liberties Index (CL_t)

	Pooled		F.E.		SYS		SYS		SYS		SYS	
	OLS (1)	OLS (2)	GMM (3)	GMM (4)	GMM (5)	GMM (6)	GMM (7)	GMM (8)	GMM (9)	GMM (10)	GMM (11)	GMM (12)
CL _{t-5}	0.770*** (0.0375)	0.352*** (0.0550)	0.621*** (0.0637)	0.577*** (0.0695)	0.593*** (0.0698)	0.648*** (0.0571)	0.678*** (0.0540)	0.721*** (0.0628)				
Human capital _{t-5}	0.536*** (0.183)	-0.326 (0.492)	0.596*** (0.263)	0.676*** (0.276)	0.0112 (0.310)	0.582*** (0.238)	0.497*** (0.197)	0.432*** (0.204)				
Total emigration rate _{t-5}	0.144 (0.0938)	-0.268 (0.454)	0.682*** (0.260)	0.754*** (0.280)	0.498* (0.274)	0.508*** (0.215)	0.473*** (0.215)	0.434*** (0.216)				
Log population _{t-5}	-0.00656 (0.00506)	-0.224* (0.119)	0.0271 (0.0168)	0.0291 (0.0185)	0.00452 (0.0134)	0.0197 (0.0140)	-0.00002 (0.0125)	0.00443 (0.0135)				
Log GDP per capita _{t-5}					0.0467*** (0.0227)							
Share tertiary ed. migrants _{t-5}						-0.126 (0.0911)						
Human capital _{t-5} *Soc. dummy									0.858* (0.489)			
Total emigration rate _{t-5} *Soc. dummy									-1.702* (0.999)			
Human capital _{t-5} *Soc. dummy*d90										1.783*** (0.374)		
Total emigration rate _{t-5} *Soc. dummy*d90											-2.144*** (0.482)	
Time dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
R-squared	0.692	0.327										
AR(1) test			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AR(2) test			0.571	0.604	0.836	0.521	0.544	0.374	0.544	0.374	0.374	0.374
Hansen J test			0.241	0.0662	0.247	0.291	0.588	0.467	0.588	0.467	0.467	0.467
Observations	476	476	476	476	423	476	476	476	476	476	476	476
N. countries		91	91	91	85	91	91	91	91	91	91	91
N. instr.			74	62	76	91	90	86	90	86	86	86

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors clustered by country in parentheses. One step system GMM estimator. The sample is an unbalanced sample comprising data at five year interval between 1980 and 2005. AR(1) and AR(2) are the p-values of Arellano-Bond test for serial correlations. The values reported for the Hansen J test are the p-values for the null hypothesis of instrument validity. All the variables are treated as pre-determined. They are instrumented for using their own first to third lags in columns 3 and 6. They are instrumented for their own first to second lags in all the other columns. In particular, column (4) shows the same specification as column (3), but now reducing the number of instruments till the second lags for robustness check. In addition to these instruments, the system GMM also uses as instruments for the level equations the explanatory variables in the first differences lagged one period.

Table 3: Dependent Variable: Polity2 Index

	Pooled		F.E.		SYS		SYS		SYS		SYS	
	OLS	OLS	OLS	OLS	GMM	GMM	GMM	GMM	GMM	GMM	GMM	GMM
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Polity2 _{t-5}	0.723*** (0.0415)	0.365*** (0.0558)	0.577*** (0.0675)	0.560*** (0.0661)	0.554*** (0.0664)	0.594*** (0.0606)	0.639*** (0.0615)	0.638*** (0.0682)				
Human capital _{t-5}	0.662*** (0.213)	-0.700 (0.693)	0.520 (0.349)	0.565 (0.360)	0.173 (0.557)	0.643*** (0.309)	0.516* (0.268)	0.547* (0.293)				
Total emigration rate _{t-5}	0.219* (0.127)	-0.470 (0.745)	1.389*** (0.420)	1.486*** (0.450)	1.141*** (0.400)	0.955*** (0.309)	0.980*** (0.334)	1.120*** (0.370)				
Log population _{t-5}	-0.00306 (0.00745)	-0.312** (0.150)	0.0894*** (0.0307)	0.0928*** (0.0337)	0.0568** (0.0265)	0.0471** (0.0226)	0.0337 (0.0215)	0.0531* (0.0274)				
Log GDP per capita _{t-5}					0.0420 (0.0418)							
Share tertiary ed. migrants _{t-5}						-0.163 (0.143)						
Human capital _{t-5} *Soc. dummy							0.997** (0.448)					
Total emigration rate _{t-5} *Soc. dummy							-2.527** (1.023)					
Human capital _{t-5} *Soc. dummy*d90								0.833 (0.544)				
Total emigration rate _{t-5} *Soc. dummy*d90								-3.032*** (0.541)				
Time dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
R-squared	0.660	0.432										
AR(1) test			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AR(2) test			0.664	0.661	0.663	0.711	0.742	0.754	0.742	0.742	0.754	0.754
Hansen J test			0.279	0.200	0.243	0.284	0.605	0.627	0.605	0.605	0.627	0.627
Observations	459	459	459	459	412	459	459	459	459	459	459	459
N. countries		85	85	85	79	85	85	85	85	85	85	85
N. instr.			74	62	76	76	90	86	90	90	86	86

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors clustered by country in parentheses. One step system GMM estimator. The sample is an unbalanced sample comprising data at five year interval between 1980 and 2005. AR(1) and AR(2) are the p-values of Arellano-Bond test for serial correlations. The values reported for the Hansen J test are the p-values for the null hypothesis of instrument validity. All the variables are treated as pre-determined. They are instrumented for using their own first to third lags in column 3. They are instrumented for their own first to second lags in all the other columns. In particular, column (4) shows the same specification as column (3), but now reducing the number of instruments till the second lags for robustness check. In addition to these instruments, the system GMM also uses as instruments for the level equations the explanatory variables in the first differences lagged one period.

Table 4: Dependent Variable: Economic Freedom of the World Index (EFW)

	Pooled		F.E.		SYS		SYS		SYS		SYS	
	OLS	OLS	OLS	OLS	GMM	GMM	GMM	GMM	GMM	GMM	GMM	GMM
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
EFW _{t-5}	0.760*** (0.0323)	0.456*** (0.0559)	0.759*** (0.0615)	0.760*** (0.0610)	0.788*** (0.0415)	0.741*** (0.0585)	0.744*** (0.0542)	0.713*** (0.0561)				
Human capital _{t-5}	0.201*** (0.0604)	0.0997 (0.203)	0.167** (0.0767)	0.173** (0.0805)	-0.00642 (0.134)	0.158** (0.0762)	0.175* (0.0901)	0.214** (0.0900)				
Total emigration rate _{t-5}	0.158*** (0.0432)	0.382 (0.325)	0.203*** (0.0779)	0.201** (0.0825)	0.0996 (0.0989)	0.238*** (0.0878)	0.169** (0.0754)	0.215*** (0.0770)				
Log population _{t-5}	0.00215 (0.00175)	-0.0238 (0.0590)	0.00221 (0.00494)	0.000484 (0.00527)	-0.00381 (0.00541)	0.000695 (0.00507)	-0.000487 (0.00476)	0.00234 (0.00413)				
Log GDP per capita _{t-5}					0.0182* (0.0103)							
Share tertiary ed. migrants _{t-5}						0.0666* (0.0365)						
Human capital _{t-5} *Soc. dummy							-0.178 (0.176)					
Total emigration rate _{t-5} *Soc. dummy							1.491* (0.851)					
Human capital _{t-5} *Soc. dummy*d90								-0.861*** (0.190)				
Total emigration rate _{t-5} *Soc. dummy*d90								1.982*** (0.566)				
Time dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
R-squared	0.708	0.579										
AR(1) test			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AR(2) test			0.0711	0.0754	0.0472	0.0841	0.0765	0.0997	0.0765	0.0765	0.0997	0.0997
Hansen J test			0.391	0.308	0.716	0.543	0.896	0.846	0.896	0.896	0.846	0.846
Observations	372	372	372	372	357	372	372	372	372	372	372	372
N. countries		74	74	74	73	74	74	74	74	74	74	74
N. instr.			74	62	76	76	87	83	87	87	83	83

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors clustered by country in parentheses. One step system GMM estimator. The sample is an unbalanced sample comprising data at five year interval between 1980 and 2005. AR(1) and AR(2) are the p-values of Arellano-Bond test for serial correlations. The values reported for the Hansen J test are the p-values for the null hypothesis of instrument validity. All the variables are treated as pre-determined. They are instrumented for using their own first to third lags in columns 3. They are instrumented for their own first to second lags in all the other columns. In particular, column (4) shows the same specification as column (3), but now reducing the number of instruments till the second lags for robustness check. In addition to these instruments, the system GMM also uses as instruments for the level equations the explanatory variables in the first differences lagged one period.

Table 5: Balanced sample

	PR		CL		POL2		EFW
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PR _{t-5}	0.651*** (0.0600)	0.619*** (0.0650)					
CL _{t-5}			0.619*** (0.0592)	0.565*** (0.0662)			
Polity2 _{t-5}					0.577*** (0.0634)	0.560*** (0.0645)	
EFW _{t-5}							0.726*** (0.0609)
Human capital _{t-5}	0.708** (0.317)	0.870** (0.348)	0.631*** (0.240)	0.737*** (0.260)	0.542* (0.318)	0.557* (0.331)	0.0342 (0.0958)
Total emigration rate _{t-5}	0.819*** (0.317)	0.848*** (0.329)	0.518** (0.260)	0.589** (0.274)	1.127*** (0.379)	1.203*** (0.399)	0.0616 (0.116)
Log population _{t-5}	0.0456* (0.0246)	0.0452* (0.0257)	0.00971 (0.0153)	0.0129 (0.0171)	0.0644** (0.0281)	0.0673** (0.0306)	-0.00437 (0.00657)
Time dummies	yes	yes	yes	yes	yes	yes	yes
AR(1) test	0.000	0.000	0.000	0.000	0.000	0.000	0.001
AR(2) test	0.551	0.562	0.579	0.621	0.856	0.852	0.210
Hansen J test	0.414	0.429	0.252	0.113	0.391	0.181	0.997
Observations	456	456	456	456	432	432	216
N. countries	76	76	76	76	72	72	36
N. instr.	74	62	74	62	74	62	62

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors clustered by country in parentheses. One step system GMM estimator. The sample is a balanced sample comprising data at five year interval between 1980 and 2005. AR(1) and AR(2) are the p-values of Arellano-Bond test for serial correlations. The values reported for the Hansen J test are the p-values for the null hypothesis of instrument validity. All the variables are treated as pre-determined. They are instrumented for using their own first to third lags in columns 1, 3 , 5. They are instrumented for using their own first to second lags in columns 2, 4, 6, 7. In addition to these instruments, the system GMM also uses as instruments for the level equations the explanatory variables in the first differences lagged one period.

Table 6: Robustness for legal origin socialist countries (Balanced sample)

	(1) PR	(3) CL	(4) Polity2	(5) EFW	(6) PR	(7) CL	(8) Polity2	(9) EFW
PR _{t-5}	0.683*** (0.0546)				0.709*** (0.0600)			
CL _{t-5}		0.664*** (0.0511)				0.708*** (0.0582)		
Polity2 _{t-5}			0.627*** (0.0627)				0.625*** (0.0663)	
EFW _{t-5}				0.726*** (0.0609)				0.726*** (0.0609)
Human capital _{t-5}	0.667** (0.289)	0.516*** (0.198)	0.505* (0.274)	0.0342 (0.0958)	0.655** (0.286)	0.448** (0.204)	0.514* (0.286)	0.0342 (0.0958)
Total emigration rate _{t-5}	0.736*** (0.282)	0.518** (0.240)	0.994*** (0.350)	0.0616 (0.116)	0.553** (0.269)	0.357 (0.220)	1.008*** (0.342)	0.0616 (0.116)
Human capital _{t-5} *Soc. dummy	1.168*** (0.405)	1.106** (0.446)	0.834* (0.429)					
Total emigration rate _{t-5} *Soc. dummy	-2.804*** (0.589)	-2.275*** (0.804)	-3.042*** (0.697)					
Human capital _{t-5} *Soc. dummy*d90					1.419** (0.632)	1.805*** (0.401)	0.884* (0.523)	
Total emigration rate _{t-5} *Soc. dummy*d90					-2.297*** (0.611)	-2.073*** (0.553)	-2.889*** (0.550)	
Log population _{t-5}	0.0238 (0.0219)	0.00389 (0.0157)	0.0389 (0.0252)	-0.00437 (0.00657)	0.0105 (0.0202)	-0.00480 (0.0133)	0.0459* (0.0261)	-0.00437 (0.00657)
Time dummies	yes	yes	yes	yes	yes	yes	yes	yes
AR(1) test	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001
AR(2) test	0.564	0.551	0.904	0.210	0.650	0.378	0.937	0.210
Hansen J test	0.945	0.803	0.795	0.997	0.856	0.652	0.721	0.997
Observations	456	456	432	216	456	456	432	216
N. countries	76	76	72	36	76	76	72	36
N. Instruments	90	90	90	62	86	86	86	62

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors clustered by country in parentheses. One step system GMM estimator. The sample is a balanced sample comprising data at five year interval between 1980 and 2005. AR(1) and AR(2) are the p-values of Arellano-Bond test for serial correlations. The values reported for the Hansen J test are the p-values for the null hypothesis of instrument validity. All the variables are treated as pre-determined. They are instrumented for using their own first to second lags. In addition to these instruments, the system GMM also uses as instruments for the level equations the explanatory variables in the first differences lagged one period.

3 Country specific results for skilled emigration

The estimation results above suggest that both openness to migration (as measured by the total emigration rate) and human capital (as measured by the share of tertiary educated in the resident labor force) have a positive impact on domestic institutional quality. What happens, then, when an educated individual emigrates? As we have seen, the share of skilled among emigrants is never significant in our regressions. Furthermore, interacting this variable with either human capital or emigration creates multicollinearity problems that further prevent us from testing empirically whether the institutional gain from migration is higher or lower when the composition of emigration becomes more high-skill. In addition, independently of the average effect we would also like to know which countries would gain or lose from having more skilled migration. For these reasons, in what follows we will rely on numerical simulations to address the following questions:

i) Which countries would lose/gain from a marginal increase in high-skilled emigration?

ii) And which countries would lose/gain from having a non-selective emigration, that is, from having their high-skilled emigration rates counterfactually reduced so as to equal their low-skilled emigration rate?

To answer these questions we start from the empirical model (1) and then proceed with numerical experiments. To do so, we need to use additional notations and supplement equation (1) with some identities. Let us drop country and time indexes to simplify the notations. Each country at each period is characterized by N_k natives of skill k (with $k = s$ for the highly skilled and $k = u$ for the low skilled) and M_k emigrants. The skill-specific emigration rates are defined as $m_k = M_k/N_k$. Ex-ante (or before-migration) human capital is measured by the proportion of highly skilled among natives, i.e. $H \equiv \frac{N_s}{N_s+N_u}$. With these notations, it is clear we can write the average emigration used in (1) rate as:

$$m \equiv \frac{M_s + M_u}{N_s + N_u} = m_s H + m_u (1 - H),$$

and the ex-post (or after-migration) proportion of highly skilled in (1) is given by:

$$h \equiv \frac{H(1 - m_s)}{H(1 - m_s) + (1 - H)(1 - m_u)}$$

If we assume that human capital is independent of migration (i.e., we neglect any incentives to acquire additional human capital in a context of migration or, in other words, we take H as given), then we can write the partial derivatives of m and h with respect to m_s and m_u as:

$$\frac{\partial m}{\partial m_s} = H > 0; \quad \frac{\partial m}{\partial m_u} = 1 - H > 0$$

$$\frac{\partial h}{\partial m_s} = \frac{-H(1-H)(1-m_u)}{(1-m)^2} = \frac{-H(1-h)}{(1-m)} < 0$$

$$\frac{\partial h}{\partial m_u} = \frac{H(1-H)(1-m_s)}{(1-m)^2} = \frac{(1-H)h}{(1-m)} > 0$$

Recalling the empirical model and the estimated equation (1), we may consider that democracy, or more generally let's denote it institutional quality I , depends on the total emigration rate, m , and on the share of highly skilled human capital in the resident labor force, h .²⁴ We write:

$$I = \beta_1 h + \beta_2 m + \beta_3 X \tag{2}$$

$$\beta_1, \beta_2 > 0$$

The effect of emigration rates on institutions is then given by:

$$\frac{dI}{dm_s} = \beta_2 H - \beta_1 \frac{H(1-h)}{(1-m)} \leq 0$$

$$\frac{dI}{dm_u} = \beta_2(1-H) + \beta_1 \frac{(1-H)h}{(1-m)} > 0$$

This shows that the effect of low skilled emigration is always positive while the effect of high-skilled emigration is ambiguous. Focusing on skilled emigration, we can see that (i) high-skilled emigration improves institutions iff

$$\frac{\beta_2}{\beta_1} > \frac{1-h}{1-m} \equiv z_s,$$

²⁴Comparing (1) and (2), we have $I_{i,t} = Democracy_{i,t+5} - \beta_0 \cdot Democracy_{i,t}$. In the long-run (indexed by ss), we have $I_{i,ss} = (1-\beta_0) \cdot Democracy_{i,ss}$. Our variable I proxies long-run institutional quality or transitional institutional improvement.

(ii) the condition for a positive optimal brain drain ($\frac{dI}{dm_s} > 0$ for $m_s = 0$) is

$$\frac{\beta_2}{\beta_1} > \frac{1 - H}{1 - m_u(1 - H)} \equiv z_0,$$

(iii) and the condition for an interior optimal brain drain ($\frac{dI}{dm_s} < 0$ for $m_s = 1$) is

$$\frac{\beta_2}{\beta_1} < \frac{1}{(1 - m_u)(1 - H)} \equiv z_1.$$

In these conditions, variables z_s , z_0 and z_1 are country-specific and their distribution can be computed using the Gaussian kernel density estimator. The ratio $\frac{\beta_2}{\beta_1}$ results from our estimation. Using the point estimates from our baseline regressions both in an unbalanced and balanced sample (column (3) of tables 1, 2, 3, 4 for the unbalanced sample, and columns (1), (3), (5) of table 5 for the balanced sample), we can simulate these three conditions.²⁵ To save place and given that we are mostly interested in the first of these conditions, we do not present simulation results for the last two.²⁶

Figures 2a to 2d show the kernel distribution of $\frac{1-h}{1-m} \equiv z_s$ and vertical lines at values of $\frac{\beta_2}{\beta_1}$ for the 4 indicators (thick lines for the unbalanced sample; thin lines for the balanced sample). As can be seen from the figures, the marginal effect of the skilled emigration rate is positive for almost all of the countries (the kernel distribution is completely to the left of the thresholds, except for CL in a balanced sample).²⁷ However, when we calculate 90 percent confidence intervals (dashed lines) for the vertical lines at values of $\frac{\beta_2}{\beta_1}$ for each of the 4 indicators (both in the balanced and unbalanced samples²⁸),²⁹ this result turns out to be not significant since the bulk of the kernel distribution lies between the upper and lower bounds of the confidence interval.

²⁵For the simulations, we consider the unbalanced sample of 91 countries as in column (3) of tables 1 and 2. Data considered refer to year 2000 from Barro and Lee (2001) and Defoort (2008). When Barro and Lee (2001) data on human capital are missing for year 2000, the human capital data from Defoort (2008) are considered (this happens for 11 countries).

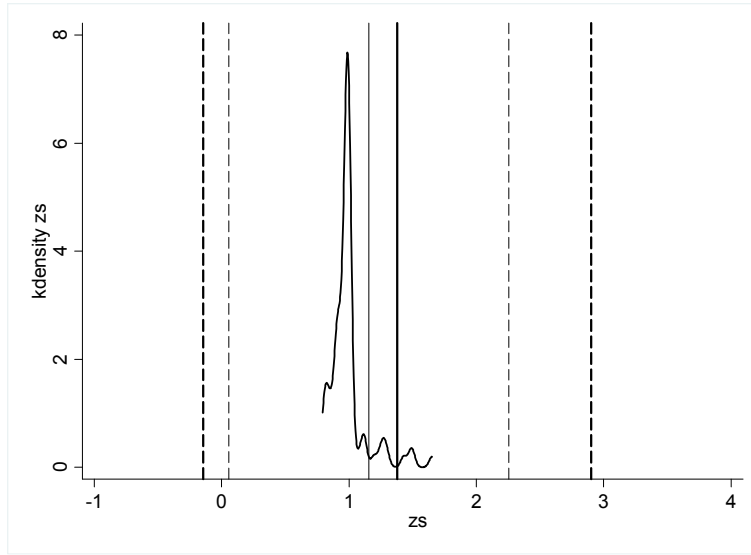
²⁶They are available from the authors upon request.

²⁷The few "losers" (negative marginal effect) are all with extremely high brain drain rates.

²⁸For EFW indicator, we consider only the unbalanced sample as in the balanced sample both β_1, β_2 are not significant and there are too few observations.

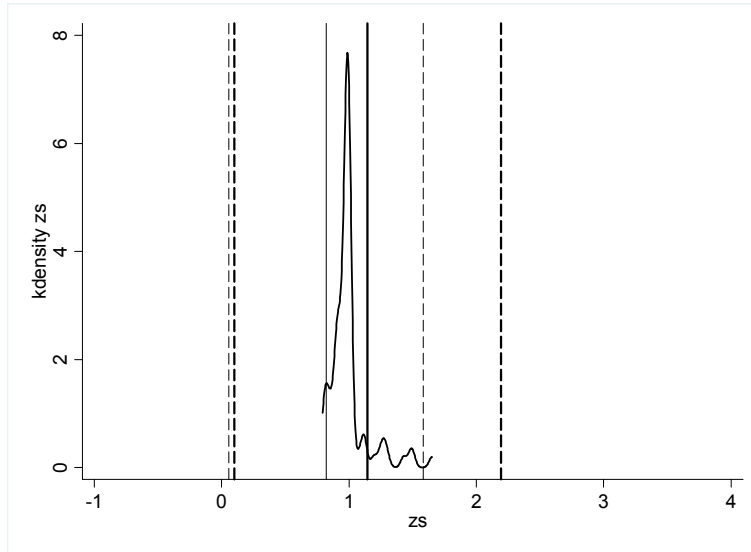
²⁹Confidence intervals of $\frac{\beta_2}{\beta_1}$ are computed using the Stata command *nlcom*. Calculations follows the *delta* method, a method based on Taylor series expansions used for deriving variance approximations and confidence intervals for transformed variables.

Figure 2.a. Simulation of a marginal increase in high-skilled emigration (PR)



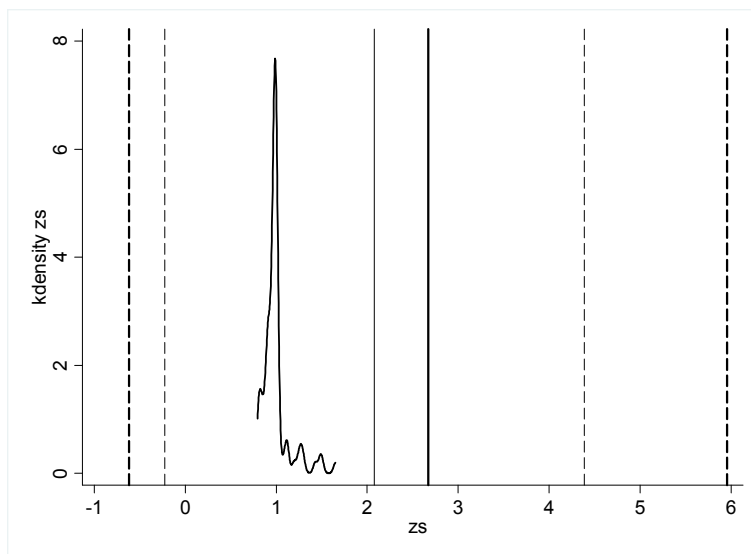
Legend: Kernel density of z_s . Vertical continuous lines = mean value of $\frac{\beta_2}{\beta_1}$. Vertical dashed lines = confidence interval at 90 percent (thick lines for unbalanced panel, thin lines for balanced panel)

Figure 2.b. Simulation of a marginal increase in skilled emigration (CL)



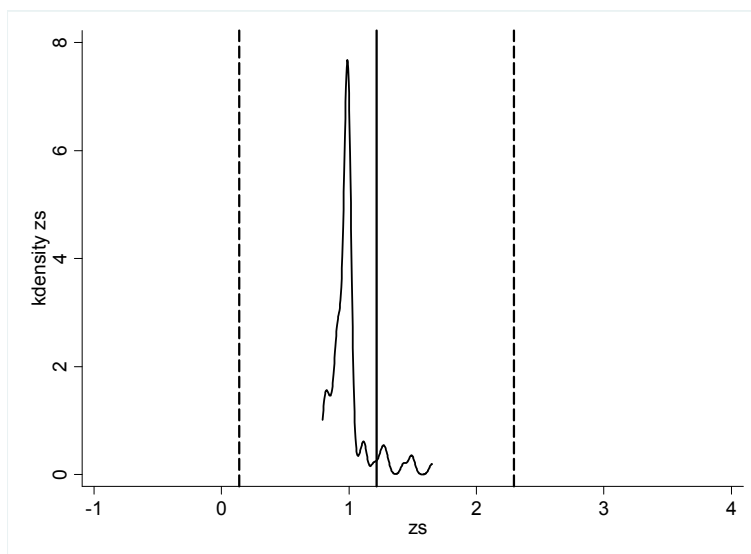
Legend: Kernel density of z_s . Vertical continuous lines = mean value of $\frac{\beta_2}{\beta_1}$. Vertical dashed lines = confidence interval at 90 percent (thick lines for unbalanced panel, thin lines for balanced panel)

Figure 2.c. Simulation of a marginal increase in skilled emigration (Polity2)



Legend: Kernel density of z_s . Vertical continuous lines = mean value of $\frac{\beta_2}{\beta_1}$. Vertical dashed lines = confidence interval at 90 percent (thick lines for unbalanced panel, thin lines for balanced panel)

Figure 2.d. Simulation of a marginal increase in skilled emigration (EFW)



Legend: Kernel density of z_s . Vertical continuous lines = mean value of $\frac{\beta_2}{\beta_1}$. Vertical dashed lines = confidence interval at 90 percent obtained with the unbalanced panel.

Turning to the second of our questions (i.e., which countries would lose/gain from having a non-selective emigration), we simulate the counterfactual quality of institutions obtained when the skilled emigration rate is set to be equal to the unskilled emigration rate. This assumption implies a decrease in the skilled emigration rate. Noting that:

$$I = \beta_1 h + \beta_2 m + \beta_3 X$$

and assuming $m_s = m_u = \tilde{m}$, we have:

$$\tilde{I} = \beta_1 \tilde{h} + \beta_2 \tilde{m} + \beta_3 X$$

where \tilde{h} is the resulting share of skilled in the resident labor force.

The change in institutional quality is given by:

$$\Delta I \equiv \tilde{I} - I = \beta_1(\tilde{h} - h) + \beta_2(m_u - m)$$

It is straightforward to see that $\Delta I < 0$ if:

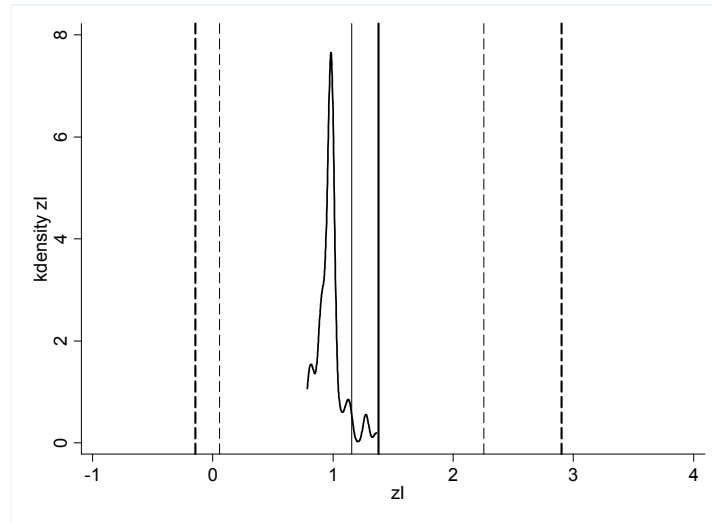
$$\beta_1(\tilde{h} - h) + \beta_2(m_u - m) < 0$$

or, similarly, if:

$$\frac{\beta_2}{\beta_1} > \frac{(\tilde{h} - h)}{(m - m_u)} \equiv z_I$$

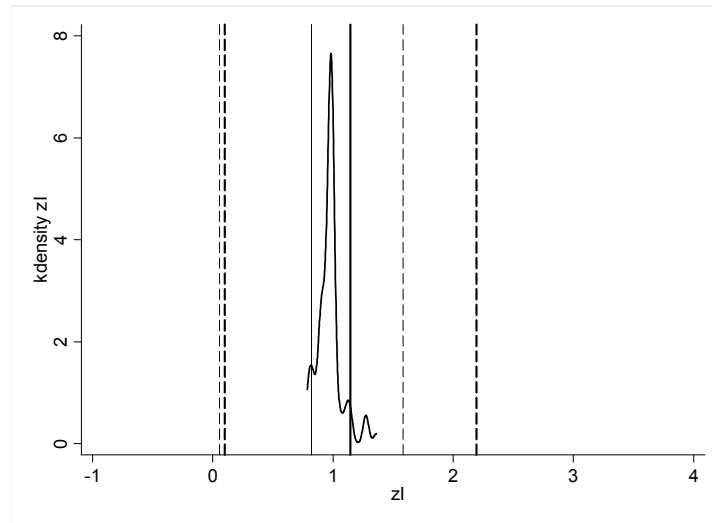
Figures 3a to 3d show the kernel distribution of z_I and vertical lines at values of $\frac{\beta_2}{\beta_1}$ for the 4 indicators (CL, PR, Polity2 and EFW) in the unbalanced (thick lines) and balanced (thin lines) samples. The results from this second simulation are similar to those of the first exercise. The fact that the kernel distribution is almost completely to the left of the thresholds suggests an institutional gain for nearly all the countries in our sample from having positively selected migrants (i.e., the counterfactual simulation from equating migration propensities across education groups yields an institutional loss), however this effect is not statistically significant as the distribution lies between the upper and the lower bounds of the confidence intervals.

Figure 3.a. Counterfactual simulation of the effect of skilled emigration on institutions ($m_s \rightarrow m_u$) - PR



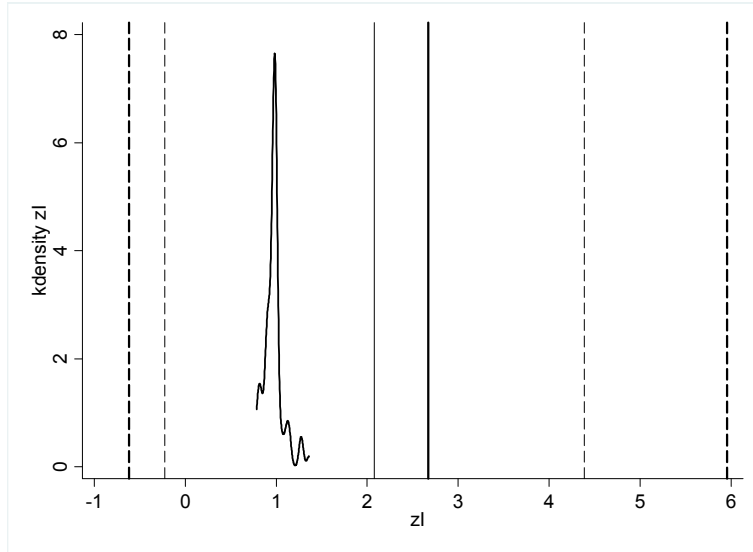
Legend: Kernel density of z_I . Vertical continuous lines = mean value of $\frac{\beta_2}{\beta_1}$. Vertical dashed lines = confidence interval at 90 percent (thick lines for unbalanced panel, thin lines for balanced panel)

Figure 3.b. Counterfactual simulation of the effect of skilled emigration on institutions ($m_s \rightarrow m_u$) - CL



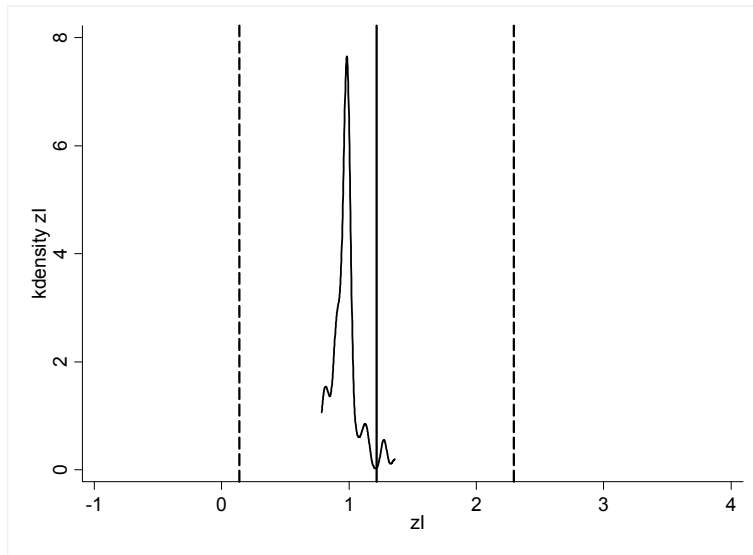
Legend: Kernel density of z_I . Vertical continuous lines = mean value of $\frac{\beta_2}{\beta_1}$. Vertical dashed lines = confidence interval at 90 percent (thick lines for unbalanced panel, thin lines for balanced panel)

Figure 3.c. Counterfactual simulation of the effect of skilled emigration on institutions ($m_s \rightarrow m_u$) - Polity2



Legend: Kernel density of z_I . Vertical continuous lines = mean value of $\frac{\beta_2}{\beta_1}$. Vertical dashed lines = confidence interval at 90 percent (thick lines for unbalanced panel, thin lines for balanced panel)

Figure 3.d. Counterfactual simulation of the effect of skilled emigration on institutions ($m_s \rightarrow m_u$) - EFW



Legend: Kernel density of z_I . Vertical continuous lines = mean value of $\frac{\beta_2}{\beta_1}$. Vertical dashed lines = confidence interval at 90 percent obtained with the unbalanced panel.

In both simulations exercises, however, we have neglected any possible incentive (or "brain gain") effects of migration on human capital formation. Following an important theoretical literature, such incentive effects have been identified in a series of recent micro (e.g., Gibson and McKenzie, 2011) and macro (Beine et al., 2008, 2010) empirical papers.³⁰ The last question we want to ask, therefore, is the following: which countries would lose/gain from a marginal increase in skilled emigration if incentive effects were accounted for? If we assume there are additional incentives to invest in human capital when there is a positive differential emigration probabilities for highly educated (which is an empirical regularity), then we should write:

$$H = H(m_s - m_u),$$

$$\text{with } H' = \frac{dH}{dm_s} > 0 \text{ and } \frac{dH}{dm_s} \cdot \frac{m_s - m_u}{H} = \delta$$

The derivative of institutional quality with respect to skilled emigration now becomes:

$$\frac{dI}{dm_s} = \beta_2 \frac{dm}{dm_s} + \beta_1 \frac{dh}{dm_s} = \beta_2 H + \beta_2 H' (m_s - m_u) - \frac{\beta_1 H (1 - h)}{1 - m} + \frac{\beta_1 H' (1 - m_s)(1 - m_u)}{(1 - m)^2}$$

After some manipulations, it is easy to see that the marginal impact of skilled emigration on institutional quality will be positive (i.e., $\frac{dI}{dm_s} > 0$), if:

$$\frac{\beta_2}{\beta_1} > \frac{1 - h}{(1 - m)(1 + \delta)} - \frac{\delta(1 - m_s)(1 - m_u)}{(1 + \delta)(m_s - m_u)(1 - m)^2} \equiv z_h$$

To be able to simulate the marginal impact of an increase in skilled migration in a context of endogenous human capital, we need to give a value to δ , the effect of emigration on human capital formation. We will use several values for δ , in particular $\delta = 0.05$ (closer to the point estimate in Beine, Docquier and Rapoport (2010): $\delta = 0.054$)³¹, $\delta = 0.01$, $\delta = 0.10$ (which correspond to the values of the lower and upper bounds of the 90 percent confidence interval of the estimated point), and $\delta = 0.23$ which corresponds to the long-run elasticity of human capital formation to high-skill migration prospects.³²

³⁰See Docquier and Rapoport (2009) for a survey of this recent brain drain literature.

³¹Note that since this value is for the short-run elasticity of human capital formation to the differential emigration probability, it can be seen as a conservative estimate.

³²In unreported simulations, we also use the lower and upper bound of the 90 percent confidence

As can be seen from Figures 4a to 4d, which simulate this condition for our four indicators, the kernel distribution shifts downward and to the left for higher values of δ . We also include a 90 percent confidence interval for the threshold $\frac{\beta_2}{\beta_1}$ as we did in the previous simulations (dashed and thick lines for the unbalanced sample, dashed and thin lines for the balanced one). In the presence of incentive effects on human capital, in the short-run the marginal effect of skilled emigration on institutional quality now appears positive and significant for a limited set of countries. For example, if we consider the Polity2 indicator (which has the lowest lower bound of confidence intervals among all the 4 indicators) in the unbalanced sample and the kernel distribution with $\delta = 0.05$, the countries with a marginal positive effect will be: Argentina, Brazil, Bulgaria, Indonesia, Namibia, Russia, Swaziland, Thailand, Turkey, Venezuela. If we consider the PR indicator instead we will have in addition to the previous ones: Bangladesh, Bolivia, Botswana, Burma, Chile, China, Costa Rica, Egypt, India, Lesotho, Lybia, Mexico, Paraguay, Peru, Syria. On the contrary, if we believe in major human capital enhancement effects in the long-run, and we consider the kernel distribution with $\delta = 0.23$, the marginal effect of skilled emigration on institutional quality becomes positive for many countries.³³

4 Conclusion

Emigration affects institutions in developing countries in many ways. By providing people with exit options and a safety-net through remittance income, emigration can lower incentives to voice internally and, eventually, delay democratic reform and political change; on the other hand, emigrants can voice from abroad and support diverse political groups and views at home; they can also contribute to the diffusion of democratic values and norms, be it directly, through return migration and contacts with relatives, or indirectly, through their belonging to social networks connecting diasporas and home-country populations. Finally, since migration is a non-random

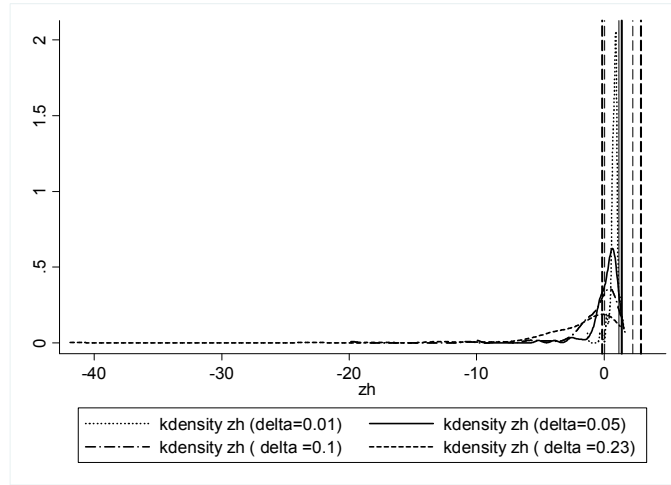
interval of the long-run estimated elasticity, i.e. $\delta = 0.06$, and $\delta = 0.4$. See Table 13 of the Appendix for the list of countries in our sample and the simulation results country by country.

³³Human capital incentive long-run effects take about 50 years to operate (convergence = 20% per decade). We have to observe that the long-run elasticity is calculated considering estimations in Beine et al. (2010), where a beta-convergence model is estimated using OLS. We know that in this case the estimated beta convergence coefficient may be biased. This may cause a larger magnitude of our long-run elasticity. At the same time, we have also to note that larger long-run elasticities were obtained for poor countries in Beine, Docquier and Oden-Defoort (2011), who use panel regressions.

process, emigration alters the composition of the home-country population on several dimensions (notably education and ethnicity) that can in turn affect democracy at home.

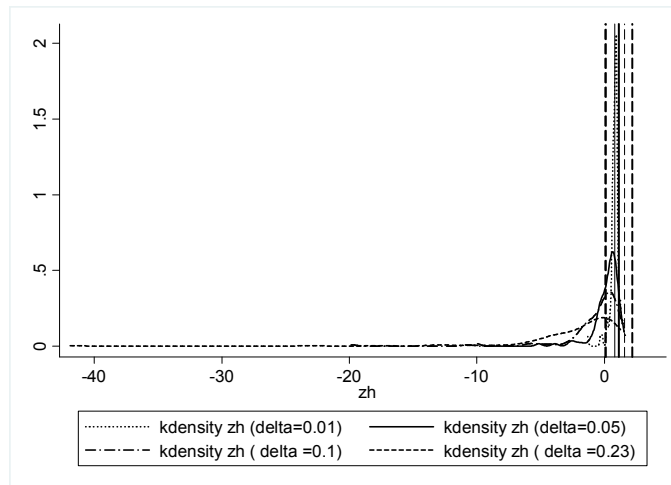
In this paper we first document these channels and then consider dynamic-panel regressions to investigate the overall impact of emigration on institutions in a large sample of developing countries. We find that openness to migration (measured by the total emigration rate) contributes to improve institutional quality (as measured by standard indicators of democracy and economic freedom) in the migrants' origin countries. This result is robust to the use of balanced/unbalanced panels, to the exclusion of certain groups of countries (e.g., former socialist countries) or to accounting for the low-coverage of certain countries in our migration data. We also find that human capital (measured by the share of tertiary educated in the resident labor force) has a positive and significant effect on institutional quality. Since skilled emigration both increases total emigration and reduces average human capital, this raises the question of whether some countries can achieve an institutional gain through a brain drain. To answer this question we rely on numerical simulations which show a positive but generally statistically insignificant effect of skilled emigration. However, once we account for the fact that emigration prospects provide additional incentives to invest in human capital, then the effect of skilled emigration on institutional quality becomes positive for a limited set of countries in the short run and for many countries in the longer run.

Figure 4.a. Simulation of a marginal increase in skilled emigration with incentive effects (PR)



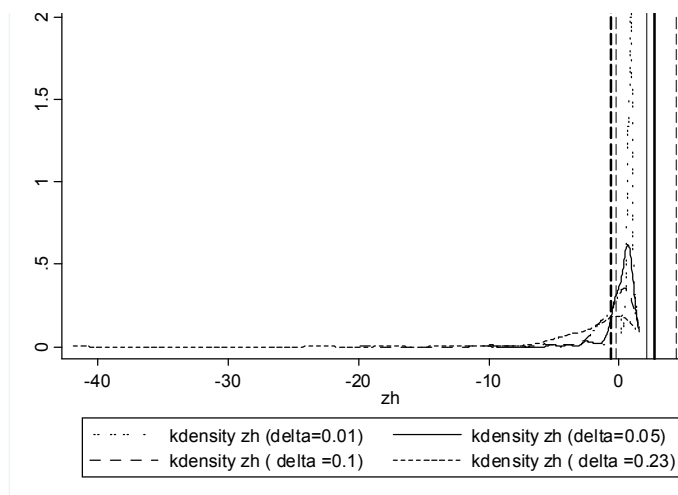
Legend: Kernel densities of z_h computed with $\delta=(0.01;0.05;0.10;0.23)$ Vertical continuous lines = mean value of $\frac{\beta_2}{\beta_1}$. Vertical dashed lines = confidence interval at 90 percent (thick lines for unbalanced panel, thin lines for balanced panel)

Figure 4.b. Simulation of a marginal increase in skilled emigration with incentive effects (CL)



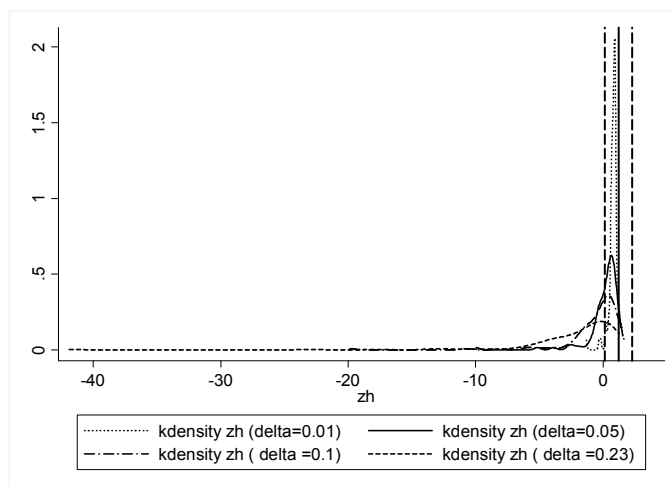
Legend: Kernel densities of z_h computed with $\delta=(0.01;0.05;0.10;0.23)$ Vertical continuous lines = mean value of $\frac{\beta_2}{\beta_1}$. Vertical dashed lines = confidence interval at 90 percent (thick lines for unbalanced panel, thin lines for balanced panel)

Figure 4.c. Simulation of a marginal increase in skilled emigration with incentive effects (Polity2)



Legend: Kernel densities of z_h computed with $\delta=(0.01;0.05;0.10; 0.23)$ Vertical continuous lines = mean value of $\frac{\beta_2}{\beta_1}$. Vertical dashed lines = confidence interval at 90 percent (thick lines for unbalanced panel, thin lines for balanced panel)

Figure 4.d. Simulation of a marginal increase in skilled emigration with incentive effects (EFW)



Legend: Kernel densities of z_h computed with $\delta=(0.01;0.05;0.10; 0.23)$ Vertical continuous lines = mean value of $\frac{\beta_2}{\beta_1}$. Vertical dashed lines = confidence interval at 90 percent obtained with the unbalanced panel.

5 References

Acemoglu, D., S. Johnson and J.A. Robinson (2005): Institutions as a Fundamental Cause of Long-Run Growth and Growth, in Aghion, P. and Durlauff, S., eds.: Handbook of Economic Growth, North Holland. Chapter 6, pp. 385-472.

Acemoglu, D., S. Johnson, J.A. Robinson and P. Yared (2005): From Education to Democracy, American Economic Review, AEA Papers and Proceedings, 95, 2: 44-49.

Agrawal, A.K, Kapur, D., J. McHale and A. Oettl (2011): Brain drain or brain bank? The impact of skilled emigration on poor country innovation, Journal of Urban Economics, 69, 1: 43-55.

Alesina, A., A. Devleeschauver, W. Easterly, S. Kurlat and R. Wacziarg (2003): Fractionalization, Journal of Economic Growth, 8: 155-94.

Alesina, A., W. Easterly and J. Matuszeski (2008): Artificial States, Journal of the European Economic Association, forthcoming.

Arellano, M., and S. R. Bond (1991): Some Specification Tests for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, Review of Economic Studies, 58: 277-298.

Arellano, M. and O. Bover (1995): Another Look at the Instrumental Variable Estimation of error-Components Models, Journal of Econometrics, 68: 29-51.

Barro, R. and J.W. Lee (2001): International data on educational attainment: updates and implications, Oxford Economic Papers, 53: 541-563.

Batista, C. and P.C. Vicente (2011): Do migrants improve governance at home? Evidence from a voting experiment, World Bank Economic Review, forthcoming.

Beine, M., F. Docquier and C. Oden-Defoort (2011): A Panel data analysis of the brain gain, World Development, forthcoming.

Beine, M., F. Docquier and H. Rapoport (2001): Brain Drain and Economic Growth: Theory and Evidence, Journal of Development Economics, 64, 1: 275-89.

Beine, M., F. Docquier and H. Rapoport (2007): Measuring international skilled migration: a new database controlling for age of entry, World Bank Economic Review, 21, 2: 249-54.

Beine, M., F. Docquier and H. Rapoport (2008): Brain drain and human capital formation in developing countries: winners and losers, The Economic Journal, 118, 4: 631-52.

Beine, M., F. Docquier and H. Rapoport (2010): On the robustness of brain gain estimates, Annales d'Economie et de Statistique, forthcoming.

Beine, M., F. Docquier and M. Schiff (2008): International migration, transfer of norms and home-country fertility, IRES Discussion Paper No 2008-43.

Blundell, R. and S. Bond (1998): Initial Conditions and Moment Restrictions in Dynamic Panel Data Models, *Journal of Econometrics*, 87: 115-43.

Bobba, M. and D. Coviello (2007): Weak instruments and weak identification in estimating the effects of education on democracy, *Economics Letters*, 96: 301-306.

Bond S., Hoeffler A. and J. Temple (2001): GMM Estimation of Empirical Growth Models, CEPR Discussion Paper No. 3048, London: CEPR.

Castello-Climent A. (2008): On the distribution of education and democracy, *Journal of Development Economics*, 87: 179-190.

Defoort, C. (2008): Long-term trends in international migration: an analysis of the six main receiving countries, *Population*, 63, 2: 317-352.

Djuric, I. (2003): The Croatian Diaspora in North America: Identity, Ethnic Solidarity, and the Formation of a Transnational National Community, *International Journal of Politics, Culture and Society*, 17, 1: 113-130.

Docquier, F. and A. Marfouk (2006): International migration by education attainment, 1990-2000; in C. Ozden & M. Schiff (eds): *International Migration, Brain Drain and Remittances*, New York: McMillan and Palgrave. Chapter 5, pp. 151-99.

Docquier, F. and H. Rapoport (2003): Ethnic discrimination and the migration of skilled labor, *Journal of Development Economics*, 70, 159-72.

Docquier, F. and H. Rapoport (2009): Skilled migration: the perspective of developing countries, Chapter 9 in J. Bhagwati and G. Hanson, eds.: *Skilled immigration: problems, prospects and policies*, Oxford University Press, pp. 247-84.

Easterly, W. (2001): The middle class consensus and economic development, *Journal of Economic Growth*, 6, 4: 317-36.

Epstein, G.S., A. Hillman and H.W. Ursprung (1999): The King never emigrates, *Review of Development Economics*, 3, 2: 107-21.

Fargues, P. (2007): The demographic benefit of international migration: a hypothesis and its application to Middle Eastern and North African Countries; in C. Özden and M. Schiff (eds), *International migration, economic development and policy*, Washington DC: World Bank and Palgrave Macmillan.

Ferguson, J. (2003): *Migration in the Caribbean: Haiti, the Dominican Republic and Beyond*, Minority Rights Group International.

Gibson, J. and D. McKenzie (2011): *The Microeconomic Determinants of Emigration*

and Return Migration of the Best and Brightest: Evidence from the Pacific, *Journal of Development Economics*, forthcoming.

Gould, D.M. (1994): Immigrant links to the home country: empirical implications for U.S. bilateral trade flows, *Review of Economics and Statistics*, 76, 2: 302-16.

Haney, P. and W. Vanderbush (1999): The Role of Ethnic Interest Groups in US Foreign Policy: The Case of the Cuban American National Foundation, *International Studies Quarterly*, 43, 2: 341–361.

Haney, P. and W. Vanderbush (2005): *Cuban Embargo: Domestic Politics of American Foreign Policy*. Pittsburgh, PA: University of Pittsburgh Press.

Hansen, L.O. (1988): The political and socio-economic context of legal and illegal Mexican Migration to the U.-S. (1942-1984), *International Migration*, 26, 1: 95-107.

Holland, J. (1999): *The American Connection: U.S. Guns, Money and Influence in Northern Ireland*, New York: National Book Network.

Iranzo, S. and G. Peri (2009): Migration and trade: theory with an application to the Eastern-Western European integration, *Journal of International Economics*, 79, 1: 1-19.

Javorcik, B.S., C. Ozden, M. Spatareanu and I.C. Neagu (2011): Migrant Networks and foreign Direct Investment, *Journal of Development Economics*, 94, 2: 151-90.

Katz, E. and H. Rapoport (2005): On human capital formation with exit options, *Journal of Population Economics*, 18, 2: 267-74.

Kauffman, D., Kraay, A. & Mastruzzi, M. (2005): *Governance matters IV: governance indicators for 1996-2004*, Mimeo., World Bank

Kerr, W. R. (2008): Ethnic Scientific Communities and International Technology Diffusion, *Review of Economics and Statistics*, 90: 518-37.

Kugler, M., and H. Rapoport (2007): International labor and capital flows: complements or substitutes?, *Economics Letters*, 94, 2: 155-62.

Lahneman, W.J. (2005): *Impact of diaspora communities on global and national politics*. Report on survey of the literature, CIA Strategic Assessment Group and University of Maryland, July.

La Porta, R., Lopez-de-Silanes, F., Shleifer A. and R. Vishny (1999): The quality of government, *Journal of Law, Economics and Organization*, 15, 1: 222-279.

Li, X. and J. McHale (2009): Does brain drain lead to institutional gain? A cross-country empirical investigation, Mimeo, Queen's University.

Mariani, F. (2007): Migration as an antidote to rent-seeking, *Journal of Development Economics*, 84, 2: 609-30.

Meyer, J.B. (2001): Network approach versus brain drain: lessons from the diaspora, *International Migration*, 39, 5: 91-110.

Mountford A. (1997): Can a brain drain be good for growth in the source economy?, *Journal of Development Economics*, 53, 2: 287- 303.

Nickell, S. (1981): Biases in dynamic models with fixed effects, *Econometrica*, 49, 1417-1426.

Omar Mahmoud, T., H. Rapoport, A. Steinmayr and C. Trebesch (2010): Do migrants remit political change? Evidence from Moldova, Work in Progress, The Kiel Institute for the World Economy.

Ragazzi, F. (2009): The invention of the Croatian Diaspora: Unpacking the politics of “diaspora” during the war in Yugoslavia, Center for Global Studies, George Mason University, Working Paper No 10.

Rauch, J.E. and A. Casella (2003): Overcoming informational barriers to international resource allocation: prices and ties, *The Economic Journal*, 113: 21 42.

Rauch, J.E. and V. Trindade (2002): Ethnic Chinese networks in international trade, *Review of Economics and Statistics*, 84, 1: 116-30.

Rodrik, D. (2007): One economics, many recipes: globalization, institutions and economic growth, Princeton: Princeton University Press.

Rodrik, D., A. Subramanian and F. Trebbi (2004): Institutions rule: the primacy of institutions over geography and integration in economic development, *Journal of Economic Growth*, 9, 2: 131-65.

Roodman, D. (2009): A Note on the Theme of Too Many Instruments, *Oxford Bulletin of Economics and Statistics*, 71, 1: 135-158.

Ross, M.L. (2001): Does oil hinder democracy?, *World Politics*, 53, 3: 325–61.

Spilimbergo, A. (2009): Foreign students and democracy, *American Economic Review*, 99, 1: 528-43.

Tsui K.K. (2010): More oil, less democracy: evidence from worldwide crude oil discoveries, *The Economic Journal*, forthcoming.

Vanderbush, W. (2009): Exiles and the Marketing of U.S. Policy toward Cuba and Iraq, *Foreign Policy Analysis*, 5, 287–306.

Wilson, A.J. (1995): *Irish America and the Ulster Conflict, 1968-1995*, Washington, D.C.: Blackstaff Press.

Wilson, J.D. (2011): Brain drain taxes for non-benevolent governments, *Journal of Development Economics*, forthcoming.

6 Appendix: robustness analysis

Table 7: Dep. Var. - Political Rights Index - PR(no socialist countries)

	(1)	(2)	(3)	(4)
PR _{t-5}	0.606*** (0.0667)	0.585*** (0.0684)	0.561*** (0.0620)	0.619*** (0.0678)
Human capital _{t-5}	0.657** (0.318)	0.805** (0.342)	0.231 (0.389)	0.717** (0.337)
Total emigration rate _{t-5}	0.840*** (0.289)	0.851*** (0.282)	0.596* (0.332)	0.973*** (0.314)
Log population _{t-5}	0.0280 (0.0194)	0.0277 (0.0185)	0.00832 (0.0175)	0.0417* (0.0238)
Log GDP per capita _{t-5}			0.0568** (0.0287)	
Time dummies	yes	yes	yes	yes
AR(1) test	0.000	0.000	0.000	0.000
AR(2) test	0.687	0.697	0.543	0.704
Hansen J test	0.330	0.449	0.468	0.526
Observations	425	425	386	414
N. countries	79	79	74	69
N. instr.	74	62	76	62

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors clustered by country in parentheses. One step system GMM estimator. Unbalanced and balanced sample include data at five year interval between 1980 and 2005. AR(1) and AR(2) are the p-values of Arellano-Bond test for serial correlations. The values reported for the Hansen J test are the p-values for the null hypothesis of instrument validity. All the variables are treated as pre-determined. They are instrumented for using their own first to further lags, according to the specification. In particular, column (1) and (2) show the results for our baseline specification when control variables of interest are instrumented for using their own to their third and second lags respectively. Column (3) introduces gdp per capita (in logs) as a control variable, while column (4) considers a balanced sample. In column (3) and (4) control variables are instrumented for using their own to their second lags. In addition to these instruments, the system GMM also uses as instruments for the level equations the explanatory variables in the first differences lagged one period.

Table 8: Dep. Var. - Civil Liberties Index - CL(no socialist countries)

	(1)	(2)	(3)	(4)
CL _{t-5}	0.603*** (0.0657)	0.585*** (0.0709)	0.535*** (0.0577)	0.606*** (0.0650)
Human capital _{t-5}	0.579** (0.236)	0.618** (0.242)	0.193 (0.258)	0.572** (0.231)
Total emigration rate _{t-5}	0.655*** (0.240)	0.712*** (0.252)	0.568** (0.252)	0.627** (0.262)
Log population _{t-5}	0.0134 (0.0143)	0.0142 (0.0150)	0.000336 (0.0145)	0.00692 (0.0163)
Log GDP per capita _{t-5}			0.0433** (0.0206)	
Time dummies	yes	yes	yes	yes
AR(1) test	0.000	0.000	0.000	0.000
AR(2) test	0.300	0.314	0.593	0.306
Hansen J test	0.239	0.108	0.326	0.188
Observations	425	425	386	414
N. countries	79	79	74	69
N. instr.	74	62	76	62

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors clustered by country in parentheses. One step system GMM estimator. Unbalanced and balanced sample include data at five year interval between 1980 and 2005. *AR(1)* and *AR(2)* are the p-values of Arellano-Bond test for serial correlations. The values reported for the Hansen J test are the p-values for the null hypothesis of instrument validity. All the variables are treated as pre-determined. They are instrumented for their own first to further lags, according to the specification. In particular, column (1) and (2) show the results for our baseline specification when control variables of interest are instrumented using for their own to their third and second lags respectively. Column (3) introduces gdp per capita (in logs) as a control variable, while column (4) considers a balanced sample. In column (3) and (4) control variables are instrumented using for their own to their second lags. In addition to these instruments, the system GMM also uses as instruments for the level equations the explanatory variables in the first differences lagged one period.

Table 9: Dep. Var. - Polity2 Index (no socialist countries)

	(1)	(2)	(3)	(4)
Polity2 _{t-5}	0.608*** (0.0700)	0.596*** (0.0724)	0.568*** (0.0673)	0.592*** (0.0708)
Human capital _{t-5}	0.381 (0.289)	0.391 (0.306)	0.0384 (0.354)	0.373 (0.307)
Total emigration rate _{t-5}	1.232*** (0.345)	1.372*** (0.382)	0.911** (0.367)	1.307*** (0.378)
Log population _{t-5}	0.0642*** (0.0233)	0.0758*** (0.0247)	0.0329 (0.0238)	0.0737*** (0.0268)
Log GDP per capita _{t-5}			0.0580** (0.0293)	
Time dummies	yes	yes	yes	yes
<i>AR</i> (1) test	0.000	0.000	0.000	0.000
<i>AR</i> (2) test	0.495	0.492	0.527	0.648
Hansen J test	0.497	0.307	0.533	0.407
Observations	408	408	375	390
N. countries	73	73	68	65
N. instr.	74	62	76	62

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors clustered by country in parentheses. One step system GMM estimator. Unbalanced and balanced sample include data at five year interval between 1980 and 2005. *AR*(1) and *AR*(2) are the p-values of Arellano-Bond test for serial correlations. The values reported for the Hansen J test are the p-values for the null hypothesis of instrument validity. All the variables are treated as pre-determined. They are instrumented for their own first to further lags, according to the specification. In particular, column (1) and (2) show the results for our baseline specification when control variables of interest are instrumented using for their own to their third and second lags respectively. Column (3) introduces gdp per capita (in logs) as a control variable, while column (4) considers a balanced sample. In column (3) and (4) control variables are instrumented using for their own to their second lags. In addition to these instruments, the system GMM also uses as instruments for the level equations the explanatory variables in the first differences lagged one period.

Table 10: Dep. Var. - Econ. Freedom Index - EFW(no socialist countries)

	(1)	(2)
EFW _{t-5}	0.760*** (0.0562)	0.833*** (0.0488)
Human capital _{t-5}	0.155* (0.0887)	0.0472 (0.120)
Total emigration rate _{t-5}	0.172** (0.0752)	0.158* (0.0850)
Log population _{t-5}	0.00160 (0.00538)	-0.00115 (0.00474)
Log GDP per capita _{t-5}		0.00520 (0.0107)
Time dummies	yes	yes
<i>AR</i> (1) test	0.000	0.000
<i>AR</i> (2) test	0.0537	0.0508
Hansen J test	0.515	0.860
Observations	337	329
N. countries	64	64
N. instr.	62	76

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors clustered by country in parentheses. One step system GMM estimator. Unbalanced and balanced sample include data at five year interval between 1980 and 2005. *AR*(1) and *AR*(2) are the p-values of Arellano-Bond test for serial correlations. The values reported for the Hansen J test are the p-values for the null hypothesis of instrument validity. All the variables are treated as pre-determined. Column (1) shows the results for our baseline specification when control variables of interest are instrumented for using their own to their second lags. Column (2) introduces gdp per capita (in logs) as a control variable. In column (2) control variables are also instrumented for using their own to their second lags. In addition to these instruments, the system GMM also uses as instruments for the level equations the explanatory variables in the first differences lagged one period. Results in a balanced sample are not reported, because of too few number of groups (and no socialist countries).

Table 11: Robustness for Oil exporting countries

	PR		CL		Polity2		EFW	
	Unbalanced	Balanced	Unbalanced	Balanced	Unbalanced	Balanced	Unbalanced	Balanced
PR_{t-5}	0.585*** (0.0642)	0.593*** (0.0633)	0.537*** (0.0649)	0.526*** (0.0621)	0.581*** (0.0643)	0.591*** (0.0617)	0.688*** (0.0493)	0.650*** (0.0514)
CL_{t-5}							0.253*** (0.0712)	0.163* (0.0928)
$Polity2_{t-5}$							-0.483*** (0.110)	-0.511*** (0.0906)
EFW_{t-5}							0.221*** (0.0841)	0.153 (0.129)
Human capital $_{t-5}$	1.052*** (0.341)	1.102*** (0.333)	0.955*** (0.272)	0.983*** (0.261)	0.765** (0.313)	0.675** (0.304)	0.0474 (0.0984)	0.136*** (0.0653)
Human cap. $_{t-5}$ *Oil exp. dum.	-1.864** (0.849)	-1.863** (0.816)	-1.918** (0.793)	-1.899** (0.804)	-1.744* (1.048)	-1.310 (0.953)	0.00457 (0.00473)	0.00358 (0.00635)
Total emigration rate $_{t-5}$	0.707** (0.281)	0.700** (0.290)	0.543** (0.236)	0.432* (0.248)	1.066*** (0.330)	0.919*** (0.310)		
Tot. em. rate $_{t-5}$ *Oil exp. dum.	0.555* (0.332)	0.473 (0.385)	0.767*** (0.279)	0.619* (0.347)	0.398 (0.588)	0.222 (0.620)		
Log population $_{t-5}$	0.0365* (0.0219)	0.0378* (0.0207)	0.0162 (0.0150)	0.00604 (0.0147)	0.0674*** (0.0223)	0.0546** (0.0224)		
Time dummies	yes	yes	yes	yes	yes	yes	yes	yes
$AR(1)$ test	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.0006
$AR(2)$ test	0.526	0.534	0.672	0.689	0.681	0.862	0.0735	0.197
Hansen J test	0.942	0.932	0.658	0.803	0.810	0.832	0.958	1.000
Observations	476	456	476	456	459	432	372	216
N. of countries	91	76	91	76	85	72	74	36
N. of instruments	90	90	90	90	90	90	90	88

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors clustered by country in parentheses. One step system GMM estimator. Unbalanced and balanced samples include data at five year interval between 1980 and 2005. $AR(1)$ and $AR(2)$ are the p-values of Arellano-Bond test for serial correlations. The values reported for the Hansen J test are the p-values for the null hypothesis of instrument validity. All the variables are treated as pre-determined. They are instrumented for using their own first to second lags. In addition to these instruments, the system GMM also uses as instruments the explanatory variables in the first differences lagged one period.

Table 12: Robustness for Sub-Saharan African countries

	PR		CL		Polity2		EFW	
	Unbalanced	Balanced	Unbalanced	Balanced	Unbalanced	Balanced	Unbalanced	Balanced
PR_{t-5}	0.606*** (0.0641)	0.583*** (0.0647)	0.575*** (0.0640)	0.548*** (0.0637)	0.600*** (0.0616)	0.603*** (0.0583)	0.749*** (0.0609)	0.755*** (0.0599)
CL_{t-5}								
$Polity2_{t-5}$								
EFW_{t-5}								
Human capital $_{t-5}$	0.939*** (0.355)	1.168*** (0.376)	0.766*** (0.268)	0.913*** (0.288)	0.765** (0.342)	0.673** (0.321)	0.228*** (0.0880)	0.0718 (0.0939)
Human cap. $_{t-5}$ *SSA dum.	1.807* (0.992)	1.348 (1.047)	1.189 (0.852)	0.882 (0.884)	1.433 (1.247)	0.631 (0.971)	-0.227 (0.525)	0.302** (0.153)
Total emigration rate $_{t-5}$	0.917*** (0.275)	0.875*** (0.283)	0.727*** (0.238)	0.674*** (0.259)	1.339*** (0.365)	1.039*** (0.310)	0.210*** (0.0734)	0.0699 (0.0845)
Tot. em. rate $_{t-5}$ *SSA dum.	1.236 (1.200)	3.163*** (1.203)	0.226 (1.019)	1.910** (0.902)	2.988*** (1.151)	2.402** (1.081)	0.692** (0.279)	0.294* (0.163)
Log population $_{t-5}$	0.0405** (0.0193)	0.0389** (0.0171)	0.0196 (0.0130)	0.0169 (0.0135)	0.0734*** (0.0223)	0.0526** (0.0206)	0.00323 (0.00536)	-0.00162 (0.00487)
Time dummies	yes	yes	yes	yes	yes	yes	yes	yes
$AR(1)$ test	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.0006
$AR(2)$ test	0.578	0.594	0.586	0.619	0.705	0.874	0.0713	0.207
Hansen J test	0.461	0.786	0.378	0.764	0.462	0.800	0.894	1.000
Observations	476	456	476	456	459	432	372	216
N. of countries	91	76	91	76	85	72	74	36
N. of instr.	90	90	90	90	90	90	90	90

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors clustered by country in parentheses. One step system GMM estimator. Unbalanced and balanced sample include data at five year interval between 1980 and 2005. $AR(1)$ and $AR(2)$ are the p-values of Arellano-Bond test for serial correlations. The values reported for the Hansen J test are the p-values for the null hypothesis of instrument validity. All the variables are treated as pre-determined. They are instrumented for using their own first to second lags. In addition to these instruments, the system GMM also uses as instruments for the level equations the explanatory variables in the first differences lagged one period.

Table 13: Country specific impact of skilled migration on institutional quality

Countries	z_s	z_I	z_h ($\delta = 0.05$)	z_h ($\delta = 0.10$)	z_h ($\delta = 0.01$)	z_h ($\delta = 0.23$)
Afghanistan	0.990	0.984	0.754	0.540	0.941	0.064
Algeria	0.976	0.973	-0.013	-0.912	0.771	-2.908
Argentina	0.809	0.806	-2.122	-4.787	0.199	-10.701
Bangladesh	0.971	0.969	-0.304	-1.463	0.705	-4.036
Barbados	1.300	1.150	1.165	1.043	1.272	0.771
Belize	1.267	1.133	1.151	1.045	1.243	0.810
Benin	0.986	0.984	0.540	0.135	0.893	-0.765
Bolivia	0.867	0.861	-0.233	-1.232	0.638	-3.451
Botswana	0.964	0.963	-0.591	-2.005	0.641	-5.143
Brazil	0.918	0.917	-2.535	-5.674	0.200	-12.640
Bulgaria	0.819	0.815	-1.039	-2.728	0.433	-6.476
Burma (Myanmar)	0.969	0.968	-0.313	-1.478	0.703	-4.065
Burundi	0.983	0.981	0.278	-0.362	0.836	-1.783
Cameroon	0.987	0.984	0.677	0.396	0.923	-0.229
Central African Republic	0.985	0.984	0.262	-0.396	0.835	-1.855
Chile	0.854	0.848	-0.273	-1.297	0.620	-3.571
China	0.976	0.975	-0.454	-1.753	0.678	-4.636
Colombia	0.920	0.912	0.324	-0.219	0.796	-1.423
Congo, Dem. Rep. of the	0.989	0.988	0.347	-0.238	0.856	-1.534
Congo, Rep. of the	0.982	0.972	0.726	0.494	0.929	-0.022
Costa Rica	0.838	0.830	-0.166	-1.078	0.629	-3.103
Croatia	1.013	0.998	0.601	0.226	0.927	-0.606
Cuba	0.986	0.956	0.745	0.526	0.936	0.041
Czech Republic	0.907	0.900	0.144	-0.549	0.748	-2.087
Dominica	1.429	1.269	1.258	1.103	1.393	0.759
Dominican Republic	0.973	0.954	0.514	0.097	0.878	-0.828
Ecuador	0.849	0.841	-0.137	-1.033	0.644	-3.022
Egypt	0.900	0.896	-0.324	-1.436	0.646	-3.906
El Salvador	1.111	1.091	0.715	0.355	1.029	-0.444
Ethiopia	0.983	0.981	0.454	-0.027	0.873	-1.096
Fiji	1.123	1.043	1.012	0.912	1.100	0.689
Gambia, The	1.011	1.005	0.929	0.854	0.994	0.687
Ghana	1.006	0.997	0.898	0.799	0.984	0.581
Guatemala	1.024	1.011	0.740	0.481	0.965	-0.092
Guyana	1.652	1.361	1.555	1.467	1.632	1.271
Haiti	1.118	1.069	1.052	0.992	1.104	0.859
Honduras	1.014	1.000	0.737	0.485	0.957	-0.074
Hungary	0.908	0.897	0.384	-0.092	0.799	-1.150
India	0.956	0.954	-0.205	-1.259	0.715	-3.600
Indonesia	0.951	0.950	-2.682	-5.985	0.195	-13.315
Iran	0.947	0.938	0.559	0.206	0.867	-0.577
Iraq	0.931	0.925	0.288	-0.297	0.797	-1.595
Jamaica	1.481	1.270	1.389	1.306	1.462	1.121
Jordan	0.811	0.801	-0.081	-0.891	0.626	-2.691
Kenya	1.004	0.996	0.871	0.750	0.976	0.483
Lesotho	0.990	0.990	-0.392	-1.648	0.703	-4.437
Liberia	1.007	0.987	0.896	0.795	0.984	0.570
Libya	0.899	0.895	-0.359	-1.503	0.637	-4.041

Countries	z_s	z_I	$z_h (\delta = 0.05)$	$z_h (\delta = 0.10)$	$z_h (\delta = 0.01)$	$z_h (\delta = 0.23)$
Malawi	0.998	0.996	0.758	0.540	0.948	0.056
Malaysia	0.938	0.930	0.442	-0.008	0.835	-1.007
Mali	1.003	1.002	0.638	0.306	0.927	-0.430
Mauritania	0.988	0.986	0.497	0.051	0.886	-0.939
Mauritius	1.076	1.046	0.973	0.880	1.055	0.672
Mexico	1.014	1.009	-0.470	-1.819	0.705	-4.812
Mozambique	0.999	0.999	0.751	0.526	0.948	0.026
Namibia	0.962	0.936	-0.887	-2.567	0.577	-6.296
Nepal	0.978	0.977	-0.057	-0.998	0.763	-3.086
Nicaragua	0.997	0.969	0.781	0.585	0.952	0.150
Niger	0.996	0.995	0.150	-0.619	0.820	-2.326
Pakistan	0.982	0.978	0.571	0.198	0.896	-0.631
Panama	0.852	0.829	0.453	0.089	0.769	-0.717
Papua New Guinea	0.995	0.990	0.813	0.647	0.957	0.279
Paraguay	0.923	0.920	-0.586	-1.958	0.609	-5.003
Peru	0.793	0.786	-0.521	-1.715	0.519	-4.366
Philippines	0.814	0.791	0.388	0.001	0.725	-0.859
Poland	0.922	0.911	0.456	0.031	0.825	-0.910
Romania	0.927	0.920	0.345	-0.184	0.806	-1.359
Russia	0.802	0.791	-3.881	-8.139	-0.171	-17.589
Rwanda	0.998	0.996	0.783	0.587	0.953	0.154
S. Lucia	1.239	1.142	1.144	1.058	1.219	0.867
S. Vincent and the Gren.	1.502	1.283	1.407	1.321	1.482	1.130
Senegal	0.995	0.991	0.676	0.386	0.929	-0.257
Seychelles	1.193	1.112	1.113	1.040	1.176	0.878
Sierra Leone	1.010	1.000	0.917	0.832	0.991	0.643
Slovakia	0.903	0.896	0.145	-0.544	0.745	-2.073
South Africa	0.907	0.901	0.154	-0.531	0.751	-2.050
Sri Lanka	0.994	0.984	0.811	0.645	0.956	0.276
Sudan	0.983	0.982	0.248	-0.420	0.830	-1.903
Swaziland	0.958	0.958	-9.952	-19.870	-1.310	-41.884
Syria	0.882	0.877	-0.180	-1.146	0.661	-3.289
Thailand	0.892	0.890	-1.683	-4.024	0.357	-9.219
Togo	0.991	0.986	0.723	0.479	0.935	-0.061
Trinidad and Tobago	1.275	1.118	1.191	1.114	1.257	0.944
Tunisia	0.983	0.978	0.322	-0.280	0.846	-1.614
Turkey	0.958	0.958	-5.140	-10.685	-0.310	-22.990
Uganda	1.001	0.997	0.862	0.736	0.972	0.455
Uruguay	0.892	0.885	0.023	-0.767	0.711	-2.520
Venezuela	0.826	0.822	-0.986	-2.633	0.450	-6.289
Vietnam	0.994	0.982	0.796	0.616	0.953	0.216
Zambia	0.987	0.984	0.673	0.387	0.922	-0.248
Zimbabwe	0.960	0.953	0.532	0.144	0.871	-0.718

Indicators	Unbalanced sample			Balanced sample		
	$\frac{\beta_2}{\beta_1}$	CI(LB) 90%	CI(UB)90%	$\frac{\beta_2}{\beta_1}$	CI(LB) 90%	CI(UB)90%
PR	1.378	-0.14563	2.90214	1.155	0.055266	2.255557
CL	1.146	0.099587	2.192358	0.821	0.057627	1.583564
Polity2	2.670051	-0.61846	5.958563	2.079336	-0.22539	4.386075
EFW	1.215187	0.13698	2.293393			