Journal of International Economics 92 (2014) 231–251

Contents lists available at ScienceDirect

Journal of International Economics

journal homepage: www.elsevier.com/locate/jie

Openness and income: The roles of trade and migration☆

Francesc Ortega [a], Giovanni Peri [b,c,*]

[a] Economics Department, Queens College, City University of New York, Powdermaker Hall, 65-30 Kissena Blvd, Flushing, New York 11367, United States
[b] University of California, Davis, One Shields Avenue, Davis, CA 95616, United States
[c] NBER, United States

ARTICLE INFO

Article history:
Received 16 September 2012
Received in revised form 14 November 2013
Accepted 23 November 2013
Available online 6 December 2013

JEL classification:
F22
J61
O4

Keywords:
International migration
Trade
Income per person
Geography
Institutions
Diversity

ABSTRACT

This paper explores the relationship between openness to trade, immigration, and income per person across countries. To address endogeneity concerns we extend the instrumental-variables strategy introduced by Frankel and Romer (1999). We build predictors of openness to immigration and to trade for each country by using information on bilateral geographical and cultural distance (while controlling for country size). Since geography may affect income through other channels, we also control for climate, disease environment, natural resources, and colonial origins. Most importantly, we also account for the roles of institutions and early development. Our instrumental-variables estimates provide evidence of a robust, positive effect of openness to immigration on long-run income per capita. In contrast, we are unable to establish an effect of trade openness on income. We also show that the effect of migration operates through an increase in total factor productivity, which appears to reflect increased diversity in productive skills and, to some extent, a higher rate of innovation.

1. Introduction

Interactions with other countries can be a powerful engine of economic development and technological change, especially for small countries (Alesina et al., 2000, 2005; Frankel and Romer, 1999). For several decades economists have focused on a country's openness to trade, measured by policies (as in Sachs and Warner, 1995; Lucas, 2009), or by trade flows as a share of GDP (as in Frankel and Romer, 1999; Rodrik, 2000; Alcalá and Ciccone, 2004) to quantify the importance of cross-country interactions on income. They realized early on, however, that openness to trade could be a consequence, as much as a cause, of high income per person across countries. To address this endogeneity, Frankel and Romer (1999) (FR from now on) proposed using cross-country variation in trade flows arising from bilateral geography in order to identify the causal effects of trade openness on income per capita. Subsequent works by Rodriguez and Rodrik (2001) and others have pointed out that the exclusion restriction behind this identification approach is likely to be violated unless one controls for other channels through which geography is likely to affect income per capita, such as natural endowments, climate, disease environment, colonization history, and so on. Rodrik et al. (2004) further argued that once one controls for institutional quality, neither geography nor trade matter much in determining a country's income per person.

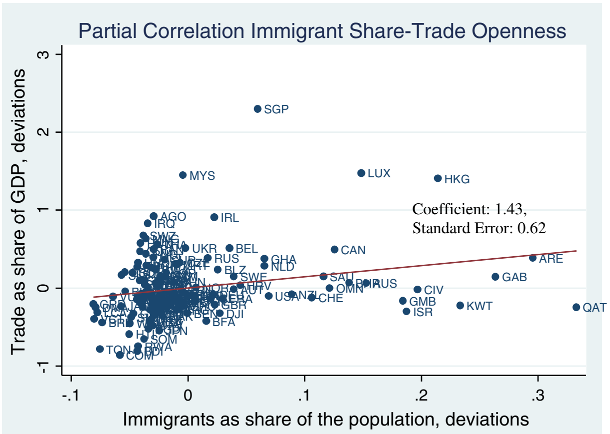
There is yet another potential problem with the approach proposed by FR. Trade openness is correlated with openness to migration.[1] Furthermore bilateral migration flows are well explained by a gravity relationship, just like trade flows (Mayda, 2007; Clark et al., 2007; Grogger and Hanson, 2011). Hence, the original specification used by FR may also suffer from a potential omitted-variables problem. Geographical proximity and accessibility also affect other forms of bilateral interactions between countries such as flows of ideas, technology and investments. However, unless these interactions are perfectly disembodied (and hence hard to measure), such flows would be

☆ The authors thank two anonymous referees for very helpful comments. Antonio Ciccone, John Devereux, Jesus Fernandez-Huertas, Andrei Levchenko, Joan Llull, Petra Moeser, Enrico Moretti, Jonathan Portes, Kevin Shih, Ryuichi Tanaka and Nico Voigtlaender provided helpful discussions. We also benefitted from comments from seminar participants at GRIPS (Tokyo), Collegio Carlo Alberto, UC Berkeley, University of Colorado, UC Santa Cruz, Harvard University, Queens College CUNY, and All UC History Conference.

* Corresponding author at: University of California, Davis, One Shields Avenue, Davis, CA 95616, United States.
E-mail addresses: fortega@qc.cuny.edu (F. Ortega), gperi@ucdavis.edu (G. Peri).

[1] Fig. 1 reports the partial correlation between trade as a share of GDP and the foreign-born share across the 146 countries included in the Frankel and Romer (1999) sample. Each variable is a residual, after we control for country size (measured by the logarithms of population and area) to purge its effect on openness to trade and migration. The Figure illustrates a clear positive and significant (but far from perfect) correlation between openness to trade and to migration.

0022-1996/$ – see front matter
http://dx.doi.org/10.1016/j.jinteco.2013.11.008



Note: The data are relative to 147 countries in year 2000. We plot the residuals after adjusting by log population and log area. The sources and construction of the trade as share of GDP and of the foreign-born share are described in the text.

Fig. 1. Migration share and trade share. Note: The data are relative to 147 countries in year 2000. We plot the residuals after adjusting by log population and log area. The sources and construction of the trade as share of GDP and of the foreign-born share are described in the text.

reflected in the mobility of goods (including capital goods) and of people. Thus we focus our analysis on these two vehicles of globalization.

This paper extends the approach proposed by FR using a new global immigration dataset and estimates the effects of economic openness, jointly considering migration and trade, on income per person. The first step in the analysis is to produce gravity-based predictors for both trade and migration. Our predictors are based on bilateral regressions that separately fit migration and trade flows on the basis of proxies for bilateral geographical and cultural distance. By examining jointly the roles played by these two dimensions of globalization, our work extends the recent analysis of the effect of trade and it connects with the research by economic historians on the First Globalization era.[2]

We also recognize that a country's geographic location may have a direct effect on income per capita (besides its effect through the channels of trade and migration), which threatens our instrumental-variables strategy. While it is infeasible to perfectly control for all possible channels in a cross-sectional setting, we consider the most plausible suspects and directly control for them in our econometric specifications. Namely, we explicitly account for the roles of climate, natural resources, disease environment, colonial origin, early development, and, perhaps most importantly, the quality of institutions. In a series of influential papers, Hall and Jones (1999), Acemoglu et al. (2001), Rodrik et al. (2004) and many others, have argued that institutions are the main factor accounting for cross-country disparities in income per capita.

Our analysis produces the following main findings. First, our gravity-based predictors appear to be highly relevant when appropriate controls for the direct effect of geography are included in the specification. Even though the predictor for the share of immigrants performs better than the predictor for the trade share, we are able to identify fairly well the roles of both trade and migration on income per capita. Second, our two-stage least-squares estimates imply that the share of immigrants in the population has a significant and robust estimated effect on long-run income per capita, although there is substantial uncertainty around its exact magnitude. On the basis of our point estimates, we find a qualitatively large effect: a 10 percentage-point difference in the share of foreign born in the population, which is close to the standard deviation in our sample, is associated with differences in income per person by a factor close to 2. If we attach a causal interpretation to this coefficient it would imply that if Japan, with a foreign-born share below 1% in year 2000, adopted a degree of openness to immigration equal to that of the US (about 11% of foreign born in 2000) its long-run income per capita would double. To the contrary, we do not find a robust effect of trade openness once we control for other effects of geography. We also show that our finding of the positive effect of migration is clearly distinct from the effects of early development and institutional quality, which we also document.

Then we empirically investigate the mechanism behind our main finding. First, we show that the estimated effect of migration on income operates mainly by increasing total factor productivity (TFP). Next, we show that underlying this finding there is a positive diversity effect. Namely, we show that the degree of diversity by country of origin within the immigrant population has an additional positive effect on income per person. Our interpretation is that diverse immigration expands the set of differentiated skills in the labor force. Finally, we also provide some suggestive evidence indicating that immigration appears to increase innovation activity, as measured by patents. This may also account for a part of the TFP effect that we uncovered. It may also imply that immigrants bring new ideas to a country, along with a wider set of skills.

While our results are consistent with immigration playing an important role in increasing productivity, two important caveats are in order. First, our cross-sectional approach is unable to control for persistent country-specific unobserved characteristics that may affect income. Short of longitudinal data, we cannot fully rule out the possibility of omitted-variable bias.[3] Second, disembodied flows of knowledge that affect productivity and are also influenced by geography may bias our estimates of the effect of migration (and trade). While we interpret our instrumental-variables estimates throughout the paper as uncovering causal effects, these two caveats should always be kept in mind.

There is a vast theoretical literature linking several aspects of openness (or globalization) to income levels and growth.[4] Some authors emphasize the role of openness to trade in promoting innovation, technological diffusion and catch-up (Grossman and Helpman, 1991; Rivera-Batiz and Romer, 1994; Eaton and Kortum, 1996; Lucas, 2009, to name a few). Others have focused on the effect of market size via trade on innovation and growth. Acemoglu (2003) has argued that the size of the market can affect the speed (as well as the direction) of technological adoption. Matsuyama (1992) and Galor and Mountford (2008) have argued that market size may encourage specialization and learning by doing. Finally, Weil (2005) has focused on the efficiency gains experienced by firms subject to international competition.

More closely related to this paper are empirical studies that estimate the effects of openness to trade on income per capita. We have already discussed the important contribution by FR, extended by Alcalá and Ciccone (2004), Noguer and Siscart (2005), and others, and the critiques by Rodrik (2000), Rodriguez and Rodrik (2001), and Rodrik et al. (2004).[5] As summarized earlier, the literature is inconclusive. Several authors have reported positive and significant effects of trade openness on income while others have raised concerns about the robustness of those findings. Two important recent contributions to this debate provide evidence based on longitudinal data. Feyrer (2009a) provides within-country estimates of the effect of trade on income that exploit the rising importance of international trade carried by air, particularly for country pairs that are connected by relatively short air routes

[2] Economic historians have argued that migration was an important vehicle for economic convergence in terms of factor prices and income levels between the 1870s and World War I, the so-called First Globalization era (Taylor and Williamson, 1997; Taylor, 1997a, 1997b). The sustained increase in international migration flows since the early 1990s has rekindled the interest in the role of migration in accounting for cross-country differences on income per capita. Recently, Putterman and Weil (2010) have argued that migration played an important role in the early economic development of many countries and that its effects have been extremely persistent.

[3] Feyrer (2009a, 2009b) shows that longitudinal data is very important to identify the effects of trade on income. These papers are reviewed below.

[4] For excellent textbook treatments of openness and economic growth, see Acemoglu (2009) chapters 18 and 19, on the roles of knowledge diffusion and trade; Barro and Sala i Martin (2004) chapter 8, discuss technology diffusion and endogenous growth. Weil (2005), chapter 11, describes the relationship between economic growth and openness.

[5] An influential early contribution was Sachs and Warner (1995) who analyzed the effect of trade policies (over the period 1965–1990) on economic growth.

relative to the corresponding sea routes. Feyrer (2009b) exploits the closing of the Suez canal as a natural experiment to try to identify the causal effects of distance on trade, and trade on income. Both papers find evidence of a positive causal effect, with some disagreement regarding the exact magnitude. On the basis of these findings Feyrer argues that longitudinal variation is crucial for identifying the effect of trade. We largely agree with this view. Our results suggest a limited ability of cross-sectional data to identify the effect of trade on income while controlling for the direct effect of geography. Nevertheless our findings also suggest that the cross-sectional approach is much more informative regarding the effect of migration on income.

This paper is also related to several studies that analyze the determinants of bilateral migration flows using a gravity equation (such as Adsera and Pytlikova, 2012; Beine et al., 2011; Bertoli and Fernandez-Huertas, 2013; Clark et al., 2007; Grogger and Hanson, 2011; Llull, 2011; Mayda, 2007, 2010; Pedersen et al., 2006, to name a few). Much more scant is the literature that employs cross-country variation to attempt to identify the causal effects of migration on income per person.[6] The closest paper to ours is Andersen and Dalgaard (2011). The main goals of this paper are similar to ours. However, these authors measure openness to migration on the basis of data on short-run cross-border movements of people (travel). As most travel is driven by tourism and business, it is strongly correlated with trade flows.[7] Still, they are able to find a positive effect of travel on income per person while controlling for trade openness. Our estimates for openness to migration and the role of institutions are robust to more demanding empirical specifications than those used in Andersen and Dalgaard (2011). Our interpretation is that the foreign-born share in a country's population may better capture the channels through which immigration affects long-run income.

Our paper is also related to the recent work of diversity on economic development. Ashraf and Galor (forthcoming) argue that there is a hump-shaped effect of genetic diversity on country-level productivity. High diversity leads to a wider spectrum of genetic traits, which makes a society more adaptable to a changing technological environment. On the other hand, high genetic diversity may undermine trust. They provide empirical evidence for this non-monotonic relationship and argue that the current levels of diversity in the US are close to the optimum implied by their estimates. Recently, Alesina et al. (2013) have analyzed the impact of birthplace diversity on economic development. These authors build diversity indicators for a large set of countries for years 1990 and 2000, disaggregated by education and nativity. Using these data they estimate a positive effect of birthplace diversity on income per capita, which appears to be larger for college-educated migrants and high-income receiving countries.

Finally, our work is also related to the strand of literature studying the role of institutions and early development on economic growth. According to Hall and Jones (1999) and Acemoglu et al. (2001, 2002), the main reason why geography appears to be a crucial determinant of cross-country differences in income per capita is that geography decisively shaped a country's history of colonization, cementing the foundations for the existing institutional arrangements. In particular, good early institutions may have allowed for policies aimed at sustaining free markets, democracy, checks and balances and well-functioning legal and judicial systems. Current cross-country income differences are also closely related to differences in development several centuries earlier (Diamond, 1997; Comin et al., 2010). Putterman and Weil (2010) show that existing measures of a country's early development substantially increase their explanatory power over current income differences when we take into account the countries of origin of the *ancestors* of the current population. Thus they argue that a country's immigration history is a crucial determinant of its current level of development. We will discuss in Section 6 the role of a country's immigration history relative to the role of its current immigrant population.

The rest of the paper is organized as follows. In Section 2 we present our empirical strategy. Section 3 presents the data and descriptive statistics. In Section 4 we reproduce the analysis of the effect of trade openness on income per person. Section 5 focuses on the effect of openness to migration on income. Section 6 analyzes the roles of institutions and ancestors. Section 7 explores the role of diversity as a channel that can account for our empirical results. Section 8 concludes. The Appendix A contains some additional material.

2. Empirical approach

2.1. Specification

Our empirical specification can be seen as a natural extension of the specification proposed by FR. We postulate that the log of income per capita in country c (y_c) is given by:

$$lny_c = \beta_0 + \beta_T TSH_c + \beta_M MSH_c + \beta_S lnS_c + \beta_C \mathbf{Controls_c} + u_c, \tag{1}$$

where TSH_c represents total trade (import plus export) as a share of GDP, MSH_c is the migration share in the population, S_c controls for country size, X_c collects all other regressors, and u_c accounts for unobserved determinants of log income per capita. To better explain the rationale behind this empirical model we present (in the Appendix A) a simple multi-country model that features trade and migration flows both across country borders and across regions within the same country. The presence of within-country flows necessitates controlling for country size. The model is based on Alesina et al. (2000) and has two main features. In the style of Armington (1969), each region is endowed with a differentiated good and a differentiated type of labor. Secondly, international trade and migration costs are higher than the analogous costs across regions within the same country (normalized to zero). Moreover, these costs are not perfectly observable. The model can be used to derive the following equilibrium relationship between (the log of) income per worker and the theoretical measures of international trade and migration openness, θ_c^T and θ_c^M, which are, respectively, inverse measures of trade and migration costs:

$$lny_c = \beta_0 + \beta_1\theta_c^T + \beta_2\theta_c^M + \beta_3 lnS_c + \beta_4\mathbf{X}_c + \varepsilon_c. \tag{2}$$

Coefficients β_1 and β_2 represent the long-run semi-elasticity of income per person to trade and to migration openness, respectively. S_c is a measure of country size. X_c is a vector that includes other determinants of long-run output per person, such as the quality of institutions, natural resources, climate, and so on. The zero-mean term ε_c allows for idiosyncratic deviations of $\ln y_c$ from its steady state and is uncorrelated with the other explanatory variables X_c. Eq. (2) cannot be directly estimated because we do not observe the latent openness of trade and migration (θ_c^T and θ_c^M), which depend on physical, cultural and policy factors. We do observe, however, the volume of trade and migration flows. Specifically, we have data on the migration shares, defined as the share of immigrants (foreign-born) in the total population, MSH_c, and the international trade flows (export plus imports) as a share of the country's GDP, TSH_c. Within our theoretical model (in the Appendix A), we derive the following relationships between the (unobserved) ideal measures of trade and migration openness and their empirical counterparts:

$$MSH_c = \Upsilon + a_1\theta_c^M - a_2 S_c + \mathbf{a}\Xi_c^M \tag{3}$$

$$TSH_c = \Psi + b_1\theta_c^T - b_2 S_c + \mathbf{b}\Xi_c^T. \tag{4}$$

---

6 Peri (2012) looks at the long-run effect of immigration on productivity and income per person across US states.

7 Their main measure is based on arrivals and departures of people traveling to, and staying in, places outside their usual place of residence, normalized by the size of the workforce. These are short-term stays (no more than one consecutive year) and include business as well as leisure travel.

As one would expect, international trade and migration openness (an inverse function of the respective international trade and migration costs) affect the equilibrium trade and migration shares. In addition, country size enters these equations. The reason is that larger countries enjoy greater domestic variety in terms of goods and factors. Since domestic trade and migration flows are less costly than international ones, larger countries will display lower trade and lower migration (in terms of TSH and MSH) than comparable countries of smaller size. Terms Ξ_c^M and Ξ_c^T collect other determinants of these shares, such as labor demand shocks or exchange rate volatility. We assume that some of those factors are not observable to the econometrician. Combining Eqs. (3) and (4) with Eq. (2) we obtain Eq. (1).[8] It is important to note that the unobserved terms in Ξ_c^M and Ξ_c^T, are now housed in the error term of Eq. (1). Some of those may affect output per worker directly and are certainly correlated with MSH_c and TSH_c. Hence, OLS estimates of Eq. (1) will suffer from some degree of omitted-variable bias. Other unobserved terms in Ξ_c^M and Ξ_c^T, uncorrelated with output per worker will act as classical measurement error.

2.2. Gravity-based instruments

Recognizing the econometric concerns discussed above, FR proposed an instrumental-variables strategy based on exploiting cross-country differences in trade and migration openness arising from the geography-based trade and migration costs. These costs are proxied by bilateral geographic and cultural characteristics. The implicit assumption is that these costs only determine output per worker by affecting access to international trade and migration.

We begin by building a predictor for bilateral trade and migration shares of country c:

$$\begin{aligned} lnx_{cj} &= \gamma_1 ln(Dist)_{cj} + \gamma_2 ln(Pop)_c + \gamma_3 ln(Pop)_j + \gamma_4 ln(Area)_c + \gamma_5 ln(Area)_j \\ &+ \gamma_6 (Landlocked)_c + \gamma_7 (Border)_{cj} + \gamma_8 (ComLang)_{cj} + \gamma_9 (Colony)_{cj} \\ &+ \gamma_{10} ln(Dist)_{cj} (Border)_{cj} + \gamma_{11} ln(Pop)_c (Border)_{cj} + \gamma_{12} ln(Pop)_j (Border)_{cj} \\ &+ \gamma_{13} ln(Area)_c (Border)_{cj} + \gamma_{14} ln(Area)_j (Border)_{cj} \\ &+ \gamma_{15} ln(Landlocked)_j (Border)_{cj} + u_{cj}. \end{aligned} \quad (5)$$

The dependent variable x_{cj} is either MSH_{cj}, the stock of immigrants from country j to country c relative to the population of country c, or TSH_{cj}, the value of trade (export + imports) between country c and j divided by the GDP of country c. The explanatory variables are the distance between the two countries, the population and area of each country, the number of countries in the pair that is landlocked, a dummy for whether country c and j share a border, a dummy for speaking a common language and a dummy for shared colonial past.[9] The interactions of the border dummies with the distance, population area, and landlocked dummies are also included to increase the predictive power of the regression. In one specification we include origin and destination dummy variables, which absorb the origin-specific and the destination-specific regressors. In that case we omit area, population and the landlocked dummies that only vary by origin or by destination.

Once we have estimated the gravity regressions (5) we aggregate them across destinations j to obtain the predicted trade and migration shares for each country c. More specifically, define Z_{cj} to be the vector of explanatory variables included in Eq. (5) and γ_M to be the vector of coefficients in the regression for migration flows, while γ_T is the vector of coefficients in the bilateral trade regression. Then we define the trade share predicted by bilateral costs for country c as:

$$\widehat{TSH}_c = \sum_{j \neq c} exp\left(\hat{\gamma}_T Z_{cj}\right). \quad (6)$$

Similarly we define the migration share predicted by bilateral costs in country c as:

$$\widehat{MSH}_c = \sum_{j \neq c} exp\left(\hat{\gamma}_M Z_{cj}\right). \quad (7)$$

These predictors reflect the variation in bilateral trade and migration flows driven by bilateral costs and partners' size. Hence, once we control for country size, variations in the predicted values of $\widehat{TSH}_c$ and $\widehat{MSH}_c$ will be driven solely by the relative position of a country in terms of its geographic and cultural coordinates. We note that the right-hand side of the gravity regressions is identical for migration and trade flows. How can then one hope to obtain two distinct predictors for openness to trade and migration from these regressions? What is crucial here is that we allow the data to assign potentially different coefficients to these explanatory variables for trade and migration flows and this will generate different predictions when interacted with the partner country characteristics. The degree of correlation between the two resulting predictors is an empirical issue, however, that needs to be examined below. The trade and migration literature have estimated gravity equations like Eq. (5) repeatedly. Our goal is not to have a structural interpretation of the coefficients $\hat{\gamma}_T$ and $\hat{\gamma}_M$ but rather to use the predictors (6) and (7) as instruments for the trade and migration shares.[10] We also note that our strategy here is in the same spirit as Do and Levchenko (2007) and di Giovanni and Levchenko (2009) who estimate a set of similar bilateral trade models at the sector level. Variation in their sector-level predictors is also based on the different sensitivity across sectors to the same determinants of cultural and geographic distance.

2.3. Identification strategy

As discussed earlier our main estimating equation is given by

$$ln\, y_c = \beta_0 + \beta_T\, TSH_c + \beta_M\, MSH_c + \beta_S\, lnS_c + \beta_C\ \mathbf{Controls_c} + u_c.$$

Compared to the original FR specification, we account for migration and trade jointly. More importantly, we take seriously the criticism by Rodriguez and Rodrik (2001) and address the threats to the validity of the instrumental variables by explicitly accounting for the main channels through which geographical and cultural features may directly affect income per capita. On the basis of the empirical economic growth literature these channels are the effects of geography on early political-economic development (Putterman and Weil, 2010), on colonization

[8] In Eq. (1) β_T is equal to β_1/a_1, β_M is equal to β_2/a_2, and $\beta_S = \beta_1/b_1 + \beta_2/b_2 + \beta_3$. Term $\beta_5\Xi_c$ is a linear combination of the residual determinants of trade, $\mathbf{b}\Xi_{T,c}$, and immigration, $\mathbf{a}\Xi_{M,c}$.

[9] The role of language in shaping international migration flows has been firmly established by Adsera and Pytlikova (2012). Their findings also show that sharing a common language matters more for non-English-speaking destinations. One may be tempted to include as regressors measures of immigration policy, which have been shown to be important determinants of migration flows, Bertoli et al. (2011, 2013), and so on. However, immigration policies may not be exogenous with respect to economic conditions in the country, as emphasized in political-economy models of immigration, such as Benhabib (1996) or Ortega (2005, 2010).

[10] Nevertheless, we note that the more recent model-based implementations of the gravity equation to predict trade (e.g. Anderson and van Wincoop, 2003) and migration (e.g. Ortega and Peri, 2009, 2012) include a full set of country of origin and of country of destination fixed effects. These are needed to capture the effect of "multilateral resistance" and not including them may introduce omitted-variable bias. Hence, in one empirical implementation we estimate Eq. (5) augmented by a set of country of origin and country of destination fixed effects, which naturally greatly increases the goodness of fit of the regression. Obviously, this is because the country dummies absorb all the country-specific factors that account for the bilateral flows. This includes the roles of country size (population and area) but also expected income levels at destination. The latter is the source of the endogeneity bias that we are trying to purge. Hence, when we build the predictors for migration (and trade) we do not include the estimated coefficients associated to these country dummies. The resulting predictors are more credibly exogenous but, naturally, their ability to predict the migration flows in the data is greatly diminished. One promising intermediate step is to build the fixed-effects gravity predictor using the estimated source-country fixed-effects but leaving out the destination fixed-effects. Since in our particular application this did not make much of a difference we opted for the simpler and more clearly exogenous predictor that does not use any of the estimated country fixed effects.

and institutional quality (Hall and Jones, 1999; Acemoglu et al., 2001), on climate and the disease environment (Weil, 2007), and on agricultural productivity and availability of natural resources (Comin et al., 2010).

In order to deal with these concerns we use two approaches. Our first approach is to include an extensive vector of control variables aiming at accounting for all the main potential channels through which geography can affect income. In this way the exogeneity assumptions required for the validity of the instruments are weakened substantially. Specifically, we include distance from the equator and regional dummy variables (sub-Saharan Africa, Latin America, and East Asia) to deal with differences in culture, and type of colonization history, we include the percent of land in the tropics, a measure of soil quality, a landlocked dummy, average distance to the coast, average temperature and average humidity to control for agricultural productivity, measures of general accessibility to the country, and characteristics of its climate and measure of oil resources. We also include morbidity variables (incidence of malaria and yellow fever) that may affect health and human capital and colonial-history controls (former French colony, former English colony) that may affect the legal origin of a country (La Porta et al., 1999).

Our second approach is more ambitious, since we also attempt to provide causal estimates for the role of institutions, in addition to the role of trade and migration shares. The reason to do this is twofold. First, it is another route to relax the exclusion restrictions behind our instrumental-variables approach. Good institutions, such as protection of property rights, granting balance of powers and ensuring economic freedom, are certainly a key determinant of a country's current productivity. Moreover, institutional quality is extremely persistent over time and can be traced back to a country's colonization history, which was shaped by geographic factors (Acemoglu et al., 2001). So failing to include the quality of institutions as a regressor in Eq. (1) requires the rather heroic assumption of no correlation between our gravity-based predictors for trade and migration and the (omitted) quality of institutions. A second reason to include institutional quality as a regressor is that we will be able to compare our estimated effects of trade and migration on income to the effect of institutional quality, which has often been considered as the most important factor accounting for cross-country differences in income per capita. Clearly, this approach requires estimating a regression model with more than one endogenous regressor. Following Hall and Jones (1999) and Alcalá and Ciccone (2004), we exploit distance from the equator, that proxies for European settlement, as a source of exogenous variation for a country's current institutional quality.

In our analysis we also pay attention to the recent work by Putterman and Weil (2010). These authors have argued that the origin countries of our ancestors played an important role in shaping early political institutions. Due to the extreme persistence of institutional quality over the centuries a country's migration history is an important determinant of present-day cross-country differences in income. Controlling for it is important to isolate the effect of more recent mobility on income. Finally, we also note that the trade and migration shares we employ are imperfect proxies for the underlying theoretical openness of movements of goods and people. Our instrumental-variables estimates will also help address the resulting measurement error.

3. Data and summary statistics

Our bilateral trade data is from the NBER-UN dataset (Feenstra et al., 2005). This database uses National Accounts in order to obtain bilateral trade data and checks the importing as well as the exporting country statistics in order to improve on accuracy. We also cross-examined these data with the International Trade database (BACI) available at CEPII.[11] The UN-NBER database has slightly larger coverage, filling some missing values, especially for smaller bilateral trade values. This dataset has information on imports for over thirty thousand bilateral pairs for the year 2000. We then replace missing values with zeros.[12] The bilateral migration data are from Docquier et al. (2010) and are described there in greater detail. They measure the number of people (older than 25) born in each of 194 world countries and residing in any of these countries in 2000. The original sources of these data are national censuses conducted around the year 2000. Specifically, for 194 countries we have their working-age population broken down by country of birth and level of education (with or without college education). There are 38,031 bilateral cells, none of which have missing values, however a large fraction contain zeros, corresponding to the fact that there are no migrants between many country pairs. We complement the bilateral dataset with data on geography (bilateral distance, a dummy for sharing a border, and the number of landlocked countries in the pair), country size (in terms of population and area), language (common languages), and colonial ties. These data are from the BACI dataset, provided by CEPII and described in Head et al. (2010). The resulting dataset has over 33,000 bilateral observations for trade and migration flows, around 24,000 of which have nonzero observations for trade flows, and about 8000 have nonzero observations for migration flows (see the number of observations in Table 2). In comparison FR had only 3220 bilateral trade flows and Noguer and Siscart (2005) had 8906. Hence the coverage of our trade data is significantly larger than in the previous studies and the migration data are completely new.

We now turn to our country-level dataset, which spans 188 countries, 146 of which were present in the FR dataset. To maintain comparability we estimate our main models on this sub-sample. The remaining 42 countries tend to be low-income and small in size, which raises some issues about the quality of their data. However, we made a significant effort to extend the coverage for most variables, and thus we also present results for the full sample.[13] Our main variables of interest are real GDP per person (PPP-adjusted), a measure of income inequality (Gini coefficient), the trade share in GDP (defined as imports plus exports over PPP-adjusted GDP), real trade openness (as in Alcalá and Ciccone, 2004), the foreign-born share (both in terms of population and of human capital), an index of institutional quality and a measure of patents per person. The GDP and trade shares are from the Penn World Tables (version 7.0), the foreign-born share is calculated using the Docquier et al. (2010) data. Along the lines of Hall and Jones (1999) and Alcalá and Ciccone (2004) we build a measure of institutional quality. Our index of institutional quality is based on data in Acemoglu et al. (2001) and is built as a simple average of an index of average protection against expropriation risk and an index of constraints on the executive (around year 1990).[14] Acemoglu et al. (2001) is also our source for several additional variables that measure absolute geography, disease environment, natural resources, climate, institutional characteristics and cultural traits. We use the database from Alesina and La Ferrara (2005) for ethnic, linguistic and religious fractionalization.

Table 1 reports some basic descriptive statistics and the source for the main variables of the paper. The mean real GDP per person is $10,682, with a standard deviation that is 20% larger than the mean. The mean Gini coefficient (from the UNU-WIDER dataset) is 41.53 (standard deviation 11.04). The mean trade share is 90%, with a standard deviation of 50 percentage points.[15] The average degree of real

[11] The correlation coefficient with the CEPII bilateral trade data for year 2000 is 0.99 when restricting to the same country pairs. These data can be downloaded at http://www.cepii.fr/anglaisgraph/bdd/baci.htm.

[12] We note that this will have no effect on our linear-in-logs predictors since the zero values will be dropped anyway. However, it will allow us to increase the number of observations in the non-linear estimation (Poisson pseudo-maximum likelihood). We build the trade flow for each country pair by adding imports and exports.

[13] We have also performed most of the regressions on the full dataset, with very similar findings (available upon request).

[14] For more details see page 1397 in Acemoglu et al. (2001).

[15] As small countries have very large degree of trade openness, if one weights each country by its GDP the average trade share is 54%.

Table 1
Descriptive statistics and data sources for the main variables.

Variable	Obs	Mean	Std. Dev.	Min	Max	Source
Dummy Frankel and Romer sample	188	0.78				Frankel and Romer (1999)
Real GDP per person in 2000 (PPP, chain-weighted 2005 USD)	184	10,682	12,881	117	74,162	PWT, 7.2
TSH = trade flows/GDP	184	0.9	0.5	0.02	3.78	PWT, 7.2
Real TSH	184	0.5	0.42	0.01	2.72	Alcalá and Ciccone (2004), PWT 7.2
MSH = foreign-born/resident pop.	188	0.04	0.08	0	0.52	Docquier et al. (2010)
Emigrated/resident population	188	0.06	0.09	0	0.49	Docquier et al. (2010)
MSH in terms of human capital	175	0.09	0.15	0	0.8	Docquier et al. (2010)
Institutional quality index	157	5.45	2.01	1	8.5	Acemoglu et al. (2001)
Diversity index immigration	168	0.7	0.22	0.02	0.96	Own calculations
Diversity index trade flows	168	0.87	0.1	0.39	0.96	Own calculations
Logarithm of population	183	1.71	2.01	−3.12	7.14	PWT, 7.2
Logarithm of area	186	11.34	2.68	3.22	16.65	BACI dataset
Distance to equator	187	25.07	17	0	67.47	BACI dataset
Share of tropical land	153	0.49	0.48	0	1	BACI dataset
Pct. Euro. descent in 1900	153	28.38	40.97	0	100	Acemoglu et al. (2001)
PW share of foreign ancestors	188	0.24	0.32	0	1	Putterman and Weil (2010)
PW early political dev. (Statehist)	160	0.48	0.23	0	0.96	Putterman and Weil (2010)
Pct. population speaking a European Language in 1975	149	31.01	43.01	0	100	Acemoglu et al. (2001)
Gini coefficient	130	41.53	11.04	21.8	76.6	UNU-WIDER
90–10 income ratio	71	11.57	11.21	3.16	67.58	UNU-WIDER
Predicted TSH (FR specification)	188	0.16	0.11	0	0.69	Own calculations
Predicted TSH (linear specification)	188	0.27	0.3	0	2.43	Own calculations
Predicted TSH (non-linear spec.)	188	0.85	0.42	0	2.14	Own calculations
Predicted TSH (linear FE)	188	0	0	0	0.01	Own calculations
Predicted MSH (FR specification)	188	0.01	0.01	0	0.04	Own calculations
Predicted MSH (linear specification)	188	0.01	0.01	0	0.06	Own calculations
Predicted MSH (non-linear spec.)	188	0.04	0.03	0	0.16	Own calculations
Predicted MSH (linear FE)	188	0.01	0.01	0	0.03	Own calculations

trade openness is 0.50 (with a standard deviation of 0.42).[16] The correlation coefficient between the two variables is 0.76. The foreign-born share, defined as the foreign-born population over the total population in the country has a mean of 0.04 (standard deviation 0.08), and ranges from virtually zero to 0.52. When we build the migration share in terms of human capital (as opposed to population), we rely on estimates of Mincerian returns and the share of college-educated. The resulting migration share (in terms of human capital) is 0.09 on average (standard deviation 0.15), and ranges from zero to 0.80. These figures reflect the fact that immigrants are more educated than natives in many countries. As one would expect, the correlation coefficient between the two definitions of the migration share is very high (0.96).

Among the remaining variables let us comment on two important control variables from Putterman and Weil (2010). The first is an index of early political development (the so-called *Statehist* variable). This index characterizes the level of sophistication of the sociopolitical institutions in the countries of origin of the ancestors around year 1500 of the current population for each country. This index is available for 160 of the countries in our sample. We also use their data, specifically their bilateral matrix of ancestry, to compute the share of the current population (year 2000) in each country whose ancestors in year 1500 lived in a different country. This is a measure of openness to international migration over the very long run. The average value is 0.24, with a large standard deviation (0.32), and ranges from zero to 1. In addition the Table reports descriptive statistics on some of our main control variables (population, area, percent of the population speaking European languages), measures of income inequality (used as dependent variables later in the analysis), and a series of variations on our gravity-based predictors for the trade share (*TSH*) and migration share (*MSH*), which are the core of our instrumental-variables strategy. We discuss their construction in detail below.

[16] Following footnote 4 in Alcalá and Ciccone (2004), real trade openness is defined as (nominal) openness times the price level, which undoes the dependence on relative nontradeable goods prices.

4. Preface: Trade and income

We preface our empirical analysis by briefly presenting the estimates of the gravity models for bilateral trade flows, and reproducing the results of the previous literature that focused only on the effect of trade openness on income.

4.1. Gravity estimates for trade flows

Table 2 (specifications 1 to 3) reports the estimates of the gravity model for bilateral trade flows, based on Eq. (5) where the dependent variable is the log of the bilateral trade share. Column 1 reports the estimates of a linear-in-logs model. Column 2 reports the estimates of a similar model that includes country of origin and country of destination dummy variables. This specification will be helpful in assessing if the coefficients estimated with the standard predictor (column 1) suffer from omitted-variable bias. Moreover the fixed-effects specification is better motivated theoretically (see Anderson and van Wincoop, 2003 regarding trade flows, and Ortega and Peri, 2009; Bertoli and Fernandez-Huertas, 2013 in the context of international migration).[17] In column 3, we follow Silva and Tenreyro (2006) and adopt a non-linear estimation method (Poisson pseudo-maximum likelihood). As argued by these authors, the latter estimation method addresses important heteroskedasticity issues and also boosts the sample size because it can naturally accommodate observations with zero bilateral values.[18]

Qualitatively, the point estimates are similar across the three columns and have the expected signs: geographical distance is associated with lower bilateral trade shares, while sharing a common language and having colonial ties are all associated to larger bilateral trade shares. In particular, we note that the coefficient on log distance is very similar

[17] It is important to keep in mind that our goal here is not to identify the structural parameters of the underlying model for trade and migration flows. Our aim is to build predictors of these flows that can be considered plausibly exogenous. For an instance of convincing identification of the effects of distance on trade flows see Feyrer (2009b).

[18] To reduce the computational burden we do not include country fixed effects in the non-linear model.

Table 2
Gravity models for bilateral trade share (TSH) and migration share (MSH).

Estimation	OLS	FE	Poisson	OLS	FE	Poisson
Dep. Var.	Ln bil. TSH	Ln bil. TSH	Ln bil. TSH	Ln bil. MSH	Ln bil. MSH	Ln bil. MSH
	(1)	(2)	(3)	(4)	(5)	(6)
Ln distance	−1.82***	−1.71***	−0.87***	−1.38***	−1.37***	−1.46***
	(0.04)	(0.03)	(0.08)	(0.04)	(0.04)	(0.08)
Ln pop. dest	0.02		−0.21***	−0.40***		−0.30***
	(0.01)		(0.03)	(0.02)		(0.04)
Ln pop. origin	1.08***		0.83***	0.63***		0.74***
	(0.01)		(0.04)	(0.02)		(0.07)
Ln area origin	−0.07***		0.04	0.20***		0.15***
	(0.01)		(0.03)	(0.02)		(0.04)
Ln area dest.	−0.25***		−0.21***	−0.08***		−0.08
	(0.01)		(0.05)	(0.02)		(0.05)
Sum landlocked	−0.82***	0.05	−0.64***	−0.25***	−2.50***	−0.67***
	(0.03)	(0.45)	(0.07)	(0.05)	(0.95)	(0.14)
Border	−4.71***	−7.64***	−1.95	−1.01	−1.45	−2.49**
	(1.00)	(0.95)	(1.25)	(0.94)	(1.09)	(1.19)
Border * (ln dist.)	0.69***	−0.04	0.23	−0.07	0.11	0.97***
	(0.21)	(0.20)	(0.39)	(0.23)	(0.24)	(0.36)
Border * (ln pop origin)	−0.32***	−0.49***	0.01	−0.21**	−0.06	−0.08
	(0.08)	(0.07)	(0.09)	(0.09)	(0.10)	(0.11)
Border * (ln pop dest.)	−0.34***	−0.54***	−0.28***	−0.25***	−0.35***	−0.58***
	(0.08)	(0.07)	(0.10)	(0.09)	(0.09)	(0.12)
Border * (ln area origin)	0.05	0.41***	−0.11	−0.06	−0.04	−0.34***
	(0.09)	(0.08)	(0.13)	(0.10)	(0.11)	(0.12)
Border * (ln area dest.)	0.11	0.45***	0.21	0.31***	0.25**	0.20
	(0.09)	(0.08)	(0.22)	(0.10)	(0.10)	(0.15)
Border * landlocked	0.81***	0.80***	0.83***	0.32**	0.06	0.49**
	(0.11)	(0.11)	(0.14)	(0.13)	(0.14)	(0.20)
Common language	0.60***	0.21***	1.00***	0.88***	0.50***	0.85***
	(0.08)	(0.07)	(0.26)	(0.10)	(0.10)	(0.19)
Common official lang.	0.01	0.69***	−0.38	0.47***	0.64***	0.13
	(0.08)	(0.07)	(0.27)	(0.10)	(0.09)	(0.20)
Time zone diff.	0.13***	0.01	0.02	0.09***	0.02*	0.02
	(0.01)	(0.01)	(0.03)	(0.01)	(0.01)	(0.03)
Colonial ties	3.09***	0.94***	1.43***	1.27***	1.49***	1.02***
	(0.13)	(0.09)	(0.13)	(0.17)	(0.11)	(0.22)
Origin hegemon	−2.23***		−1.78***	1.02***		0.53*
	(0.18)		(0.23)	(0.22)		(0.30)
Observations R-squared	24,627	24,627	33,108	8022	8022	34,782
	0.4	0.71	0.22	0.42	0.70	0.23
Country fixed effects	No	Yes	No	No	Yes	No

Note: All models contain an intercept (not shown here). The trade share (TSH) is defined as the sum of bilateral imports and exports over GDP of the receiving country, the migration share (MSH) is the number of foreign-born in the country over the total population. The fixed-effects estimator includes a full set of origin and destination dummy variables (not reported). The estimated fixed effects are not used in building the predictors for TSH and MSH. In parenthesis we report the heteroskedasticity-robust standard errors. *, **, and *** significant at 10%, 5% and 1% confidence level.

in the first two columns. This suggests that the vector of explanatory variables included in the first column is large enough to help identify the crucial role of bilateral distance in determining trade flows.[19] We also note that the point estimates of destination population are much smaller (even negative) than the corresponding origin coefficients. This reflects the construction of trade shares where the denominator is the destination GDP. The goodness of fit is obviously substantially higher in the specification including fixed effects (column 2). Compared to the original exercise performed by FR, our gravity model includes information on past colonial ties, along the lines of Head et al. (2010), which increases the explanatory power of the model and the resulting strength of the predictor for the trade share.

As explained earlier, we use our estimates of the vector of coefficients γ_T, obtained from specifications (Acemoglu, 2003, 2009; Acemoglu et al., 2001) in Table 2, to build predicted values for all bilateral country pairs (not just those pairs used in the estimation). We then aggregate these predicted values following Eq. (6) to obtain the predicted trade share for each country. The right panel of the Table reports the estimates for the migration gravity regressions. For now it suffices to note that the overall pattern of coefficient signs is similar to that obtained for bilateral trade flows. We will return to the migration gravity regressions in Section 5 below.

4.2. Replication of the literature

In order to assess our contribution we show briefly that we can replicate the finding by FR. The benchmark of our replication is the initial work of FR, and a more updated version of the same exercise by Noguer and Siscart (2005). Following these authors, we estimate the following model:

$$lny_c = \beta_0 + \beta_T TSH_c + \beta_P lnPop_c + \beta_A lnArea_c + \beta_C \textbf{Controls} + u_c. \qquad (8)$$

In Eq. (8) the dependent variable is the log of income per person in country c measured in 2000 US Dollars, corrected for PPP as in the Penn World Tables. We include as explanatory variables the logarithm of area ($\ln Area_c$) and population ($\ln Pop_c$) to capture the effect of country size. As an instrument for the trade share we use the gravity-based predictor proposed by FR and constructed using the estimates of Table 2

[19] The same is true regarding bilateral migration flows (the right panel). We note though that the coefficient on log distance in column 6 is very similar to those in columns 4 and 5, while this is not the case for trade flows (column 3). This suggests that our estimates for migration flows may be more robust than the estimates for trade flows.

Table 3
The effects of openness to trade, 2SLS estimates.

Specification	FR sample	Full sample	Dist. to equator	Controls	OLS
	(1)	(2)	(3)	(4)	(5)
TSH	3.03***	3.49***	−0.33	−0.19	0.33
	(1.14)	(1.35)	(0.68)	(0.72)	(0.22)
ln population	0.10	0.11	0.03	−0.18*	−0.13*
	(0.13)	(0.14)	(0.08)	(0.09)	(0.07)
ln area	0.17	0.14	−0.23**	0.05	0.07
	(0.19)	(0.17)	(0.11)	(0.08)	(0.07)
Dist. to equator			0.05***	0.03***	0.02***
			(0.01)	(0.01)	(0.01)
Pct. Land tropics				0.39	
				(0.46)	
Observations	146	181	146	122	122
Controls					
Region	No	No	No	Yes	Yes
Geo/climate/disease/oil	No	No	No	Yes	Yes
Colonial origin	No	No	No	Yes	Yes
First-stage regression					
Instruments	Pred. TSH	Pred. TSH	Pred. TSH	Pred. TSH	
KP weak identif. F test	6.71	6.25	5.04	7.85	

Note: The dependent variable is the log of income per capita. All regressions include an intercept. Regional dummies for sub-Saharan Africa, East Asia, and Latin America. Geography, climate and disease controls include the percentage of land in the tropics, a landlocked dummy, average distance to the coast, average yearly temperature, average yearly humidity, an index of soil quality, an index of the incidence of malaria, and an index of the incidence of yellow fever. Colonial origin controls include dummy variables for former French colony, former English colony, and a dummy for the 4 rich "young" countries (US, Canada, Australia and New Zealand), and the share of 1900 population of European origin. For columns 1 through 4 (one endogenous variable and one excluded instrument) the Stock and Yogo (2005) critical values (maximal IV size) range from 5.53 to 16.38, respectively, from the less stringent to the most stringent test (the 25% to 15% maximal IV size). Predictors based on fixed-effects gravity regression, but not using the estimated country fixed effects. In parenthesis we report the heteroskedasticity-robust standard errors. *, **, and *** significant at 10%, 5% and 1% confidence level. The KP weak identification test is the Kleibergen–Paap rk Wald F statistic. In the case of one endogenous regressor, as here, it coincides with the Angrist and Pischke (2009) test.

(column 1) described above. Table 3 reports the two-stage least-squares estimates for Eq. (8) and heteroskedasticity-robust standard errors. Columns 1 and 2 of Table 3 report the estimates of the basic model, which includes only controls for country size (logs of area and population). Our main sample is the one used by FR and contains 146 countries. We also report results with the largest sample that we could assemble (181 countries, in column 2). Column 1 reproduces the finding in FR, where the trade share appears to have a positive and significant effect on income per person. Specifically, the point estimate is around 3, implying that a one percentage point increase in the trade is associated with a 2.5% increase in long-run income per person. These estimates are close to those found by FR, who report estimates between 1.97 and 2.96, and also hold in a larger sample of countries (column 2). Columns 3 and 4 include further controls, and represent the essence of the Rodriguez and Rodrik (2001) critique: the direct effect of geography on income overshadows the effect of the trade share. Column 3 includes distance from the equator as an additional control. This variable is highly significant, confirming the results in Hall and Jones (1999). Importantly, the coefficient on the trade share falls dramatically (by an order of magnitude) and becomes statistically insignificant. Column 4 includes three continental dummies (sub-Saharan Africa, East Asia and Latin America) and additional variables to control for geography, climate, soil quality, disease environment, and the colonial past. The point estimate of the trade share coefficient remains very small and insignificant.[20] The lack of a significance could be due to a problem of weak instruments, revealing little or no correlation between the trade share and its predictor. At the bottom of column 4 we report the Kleibergen–Paap F-statistic for weak identification. We obtain a value of 7.8. This value is higher than the least demanding critical values (5.53) reported by Stock and Yogo (2005) although lies below the highest ones (16.38). We take this as indicating that while there may be a concern about weak instruments, the problem does not seem to be extremely severe. We shall conduct a more systematic analysis of non- and weak identification in the next section. Perhaps more informative are the results reported in column 5, which report the OLS estimates of the model that includes the geography and colonial controls. Interestingly, the partial correlation between the trade share and income appears to be very small and, in fact, we cannot reject a value of zero.[21]

5. Openness to migration

The empirical growth literature has almost exclusively focused on trade data to measure overall economic openness.[22] This viewpoint neglects the well established fact that migration has played a very important role historically in disseminating ideas across the globe.[23] Research on the economic effects of immigration, instead, has taken a narrower focus, stressing the identification of labor market effects. As argued by Hanson (2009), a more general approach is needed to carry out a comprehensive analysis of the aggregate economic effects of migration.

It is certainly plausible that openness to migration may play an important role in accounting for cross-country differences in income per capita. Fig. 2 shows that there exists a robust positive partial correlation between the migration share and the logarithm of income per person across countries, after controlling for country size (population and area).[24] Naturally, these correlations may be driven by the confounding effect of trade, by reverse causality or by other dimensions of openness. To address this point we examine the joint effects of openness to trade and migration on income in a more formal regression setting. Building on the basic FR specification, Table 4 includes openness to migration (measured by the share of immigrants in the population) as an additional explanatory variable. We estimate Eq. (1), treating both the trade and migration shares as *endogenous* regressors and use the respective gravity-based predictors as instrumental variables.

Table 2 (columns 4 through 6) reports the estimates of the coefficients in the gravity migration model. The signs of the coefficients are largely as expected. As was the case with trade flows, bilateral distance reduces bilateral migration, while sharing a common language and colonial ties appear to significantly increase migration. While not dramatic, there are some noticeable and intuitive differences between the marginal roles played by some variables in accounting for trade and migration flows.[25] Consider, for instance, the simplest model (linear in logs

[20] We also run specifications (not reported) using the non-linear and the fixed-effect gravity predictors for trade as instruments. The estimates are less precise but the results are similar: the coefficient on the trade share is significant only if we do not include any control for geography.

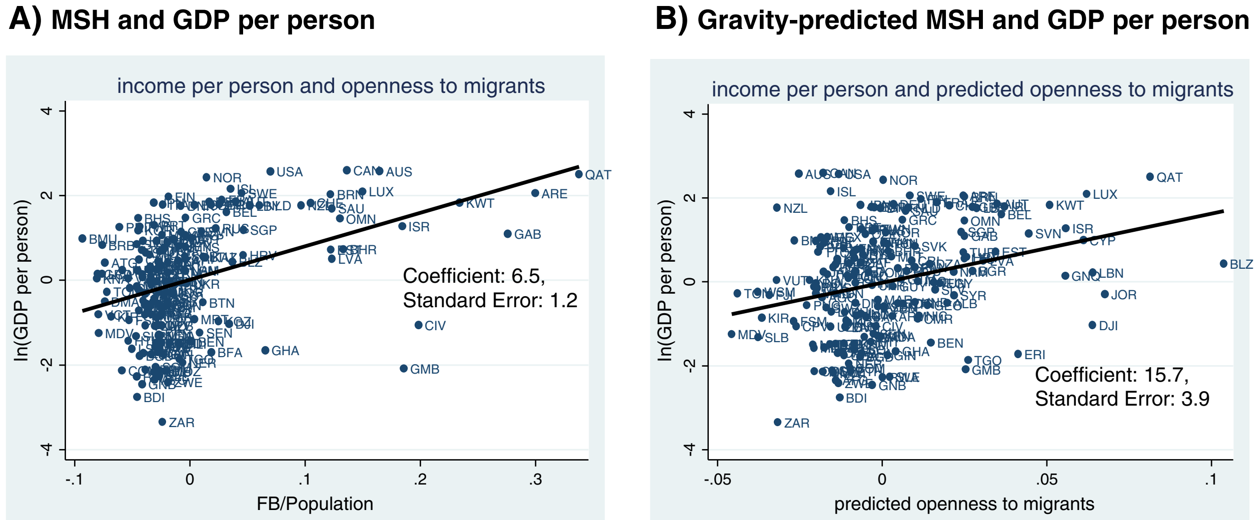
[21] Our results differ from those of Noguer and Siscart (2005), who find that the positive effect of trade openness on income is robust to the inclusion of the geographic controls. We use different (more complete and updated) data, which accounts for the disparity in results. At minimum our results suggest that the effect of trade openness uncovered by these authors using the Frankel and Romer methodology is sensitive to the data used in the estimation. It is also possible that over time the trade to GDP ratio has become an increasingly worse proxy of openness to trade.

[22] See, for instance, the review in the textbook by Weil (2007).

[23] See, for instance, Acemoglu et al. (2001), Comin et al. (2010), Diamond (1997), and more recently, Putterman and Weil (2010).

[24] Fig. 2A plots log income per person against the foreign-born share in the country. The associated regression coefficient is 6.5 with a standard error of 1.18. Fig. 2B plots the gravity-predicted migration share (after partialling out population and area) and income per capita. The regression coefficient is 15.7 with a standard error of 3.95. In both cases the correlation is robust to dropping outliers. It is also not driven by the US, Canada, or Australia—countries that are both highly economically developed and have a high foreign-born share. This is particularly clear for the predicted migration share (Fig. 2B) since the large size and relative remoteness of these countries lead to relatively low predicted immigration shares.

[25] The point estimates display some variation across estimation methods. For comparison see the estimates provided by Ortega and Peri (2013) who emphasize the role of immigration policies.



Note: The scatterplot shows each variable after adjusting for logarithm of population and area. The predictor for immigration share used is the linear gravity predictor.

Fig. 2. Openness to immigration (MSH) and GDP per person, adjusted for country size. A: MSH and GDP per person. B: Gravity-predicted MSH and GDP per person. Note: The scatterplot shows each variable after adjusting for logarithm of population and area. The predictor for immigration share used is the linear gravity predictor.

and estimated by OLS) displayed in columns 1 and 4. Bilateral distance seems to affect trade flows much more than migration flows. On the contrary, sharing a common language appears to be much more important in accounting for migration than for trade flows. Colonial ties instead seem to affect trade more than migration. It is useful to examine the relationship between actual and predicted trade and migration shares. Fig. 3 displays the corresponding scatterplots. Clearly, the predicted migration share is strongly correlated with the actual data (as seen in Fig. 3A). This correlation is large, statistically significant, and not driven by outliers (as shown in Fig. 3B). This is in clear contrast with the ability of the predicted trade share to account for the actual data (Fig. 3C and D). In this case the positive correlation between predicted and actual values depends strongly on a few influential observations. When the observations for Ireland, Luxembourg and Singapore are omitted, the correlation is weakened substantially and loses its statistical significance.[26] While the gravity-based predictors are successful in explaining both bilateral trade and migration, when aggregating to obtain the total openness of a country, the gravity predictor works noticeably better for migration.

It is also interesting to note that while the raw data exhibits a strong positive relationship between immigration shares and income per person across countries, the correlation between *predicted* immigration shares and income per person is still positive but diminished.[27] The reason for the weakening of the relationship between income per person and immigration shares is clear. Our gravity-based predictors do not incorporate in any direct way the fact that rich countries tend to attract more immigrants, as this is precisely the endogeneity bias we are trying to eliminate. The moderate remaining positive relationship arises from the fact that small, easily accessible countries (of any income level) near large populous countries are more likely to receive large immigration flows relative to their population.

5.1. Trade and migration jointly

Table 4 reports the 2SLS joint estimates of the coefficients of trade and migration openness on income per person.[28] In column 1 we only control for country size (area and population). Here only the trade share appears to be statistically significant. Specifications 2 through 5 control for distance to the equator, the key geographic control identified by Rodriguez and Rodrik (2001) on the basis of its role in determining the history of a country's institutions. The coefficient on the trade share falls dramatically and becomes statistically insignificant whereas the migration share has a large and significant effect. Column 3 reports estimates based on the full sample, column 4 uses a predictor for the migration share based on the fixed-effects gravity regression (but not using the estimated fixed effects to form the predictions, as mentioned above). Column 5 includes a comprehensive set of control variables to account for the effects of geography, disease environment, natural resources, climate, and colonial past on income.[29] Throughout, the point estimate on the MSH remains very robust.

Let us provide at this point a thorough analysis of the power of our instruments or, more formally, let us test the hypothesis of weak identification of the parameters of interest. The bottom of the table reports information regarding the performance of the instrumental variables we are using. For each regression model we report the Kleibergen–Paap F test (KP), which allows us to test the null of jointly weak instruments.[30] While in several specifications we are not able to reject the null of weak instruments, in our most preferred specifications, columns 5 and 8, we can reject the null. Specifically, for column 5 (specification with all the controls) the statistic is 5.92, which lies above all the Stock and Yogo critical values except for the most demanding one (7.03).

[26] The role of influential observations in the prediction power of the gravity-based trade shares had already been noticed in the previous literature (see Fig. 1 and following discussion in Frankel and Romer, 1999).

[27] When we compute the country-level predicted immigration shares we find that the ratio for the average rich country (income per person above 10,000 dollars) to the average poor country is 1.11, much smaller than the 3.67 ratio in the raw data. So, in practice, our gravity-based predictor predicts moderately larger immigration shares for high income countries.

[28] On the basis of these estimates we experiment to find the version of the aggregated predictors for country-level openness to trade and migration that perform better. We find that the strongest predictions for the actual shares are obtained by the linear-in-logs OLS estimation for the trade share and the non-linear estimation for the migration share. Most of our two-stage least-squares estimates are based on this vector of instruments.

[29] The controls included are the same as those in specification 4 of Table 3 and they are listed in the footnote to Table 4.

[30] This is the KP test for weak identification (not underidentification). The Stock and Yogo (2005) critical values we report are only strictly appropriate under homoskedasticity. We report heteroskedasticity-robust standard errors, which in our application tend to be higher than those obtained under the assumption of homoskedasticity. See the discussion in Baum et al. (2007).

Table 4
The effects of openness to migration. 2SLS estimates.

Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	FR sample	Dist. equator	Full sample	FE pred. MSH	Controls	Reduced form	OLS	Only MSH	VATSH
MSH	5.06	6.09***	8.51***	13.24***	7.32***		6.34***	7.30***	10.14**
	[3.43]	[1.82]	[2.18]	[4.54]	[2.05]		[1.22]	[2.15]	[5.13]
TSH	2.80**	−0.29	−0.11	0.05	−0.01		0.29*		−1.22
	[1.15]	[0.57]	[0.96]	[1.22]	[0.56]		[0.17]		[1.62]
Dist. equator		0.05***	0.04***	0.03***	0.03***	0.03***	0.03***	0.03***	0.02
		[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Pred. MSH						10.92**			
						[5.19]			
Pred. TSH						−0.24			
						[0.27]			
Observations	146	146	181	181	119	119	119	119	69
Controls									
Region dummies	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Geo/climate/disease/oil	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Colonial origin	No	No	No	No	Yes	Yes	Yes	Yes	Yes
First stage reg.									
K–P weak identif. F-test	3.7	2.47	1.69	1.41	5.92	.	.	12.27	2.26
Angrist–Pischke. F-test for MSH	11.46	12.14	12.82	5.28	11.43	.	.	12.27	8.65
Angrist–Pischke F-test for TSH	6.6	4.9	3.2	3.25	10.2	.	.	.	6.8
Instruments:	Pred. TSH	Pred. TSH	Pred. TSH	Pred. TSH	Pred. TSH	.	.		Pred. TSH
	Pred. MSH	Pred. MSH	Pred. MSH	Pred. MSH fe	Pred. MSH	.	.	Pred. MSH	Pred. MSH
SY 10% max IV size	7.03	7.03	7.03	7.03	7.03	.	.	16.38	7.03
SY 25% max IV size	3.63	3.63	3.63	3.63	3.63	.	.	5.53	3.63

Note: The dependent variable is the log of income per capita. Unless noted otherwise the predicted TSH is based on the OLS gravity estimates and the predicted MSH is based on the non-linear estimation. All regressions include an intercept, log population, log area, and the percent of land in the tropics. Regional dummies for sub-Saharan Africa, East Asia, and Latin America. Geography, climate and disease controls include the percentage of land in the tropics, a landlocked dummy, average distance to the coast, average yearly temperature, average yearly humidity, an index of soil quality, an index of the incidence of malaria, an index of the incidence of yellow fever, and a measure for oil reserves; colonial controls includes dummy variables for former French colony, former English colony, and a dummy for the 4 rich "young" countries (US, Canada, Australia and New Zealand). Predictors based on fixed-effects gravity regression do not use the estimated fixed effects. In parenthesis we report the heteroskedasticity-robust standard errors. *, **, and *** significant at 10%, 5% and 1% confidence level. The KP weak identification test is the Kleibergen–Paap rk Wald F statistic. The AP weak identification test is the Angrist and Pischke (2009) test for weak identification of individual regressors. The reported critical values are those provided by Stock and Yogo (2005) under the assumption of IID errors.

We also report the Angrist and Pischke (2009) F-statistic (AP) for the migration share and for the trade share, separately. This test is very useful in the context of our application since it evaluates whether each individual endogenous regressor, in our case the migration share or the trade share, is well identified separately, after partialling out the other endogenous regressor. A strong AP statistics for one of the endogenous regressors is an indication that we can perform inference on that coefficient, even if the other endogenous regressor is weakly identified. Across the majority of specifications we can reject the null that MSH is weakly identified for most of the Stock and Yogo critical values. In particular, for our preferred specification (column 5) we can reject the null for all but the most demanding critical value. As for the TSH, the AP F-statistic is uniformly lower. In columns 3 and 4 we cannot reject the null even at the lowest critical value. However, when we add the controls and focus on our preferred specification (columns 5), the coefficient becomes better identified. The AP F-statistic takes the value 10.2, which again is above all but the most demanding critical value (and above the "rule of thumb" of 10). On the basis of these tests we conclude that neither of our two coefficients suffer from severe weak identification in the preferred specification, provided the controls for geography, climate, and colonial past are included in the specification.[31]

As further check that weak instrument bias is not a severe concern, we report the reduced-form regression in column 6, where we include the geography-predicted trade and migration as explanatory variables. Here the power of the first stage is not an issue, as the instruments appear directly in the regression. Even in this specification the migration share predictor is positively and significantly correlated with income per person, while the trade predictor is not. This indicates that when including our extensive set of controls there does not appear to be a cross-sectional correlation between the trade share and income per capita. Column 7 reports the OLS estimates of the model with all the controls. Of course, these estimates should not be interpreted causally. Nevertheless, it is worth noting that there is a strong association between income and the migration share in the cross-section, even after controlling for many other variables.[32] In contrast, this is not the case for the trade share, which has a small point estimate (compared to the FR estimate) and is only marginally significant. Finally, column 8 presents a model that omits the TSH from the regression but includes all the controls. The point estimate of the migration share is hardly affected, which suggests the two are not strongly correlated after controlling for all other regressors.[33] Column 9 reports estimates based on a model where we have defined openness to trade on the basis of value added, following Johnson and Noguera (2012).[34] Such definition makes the trade share measure more internally consistent as only the value added of trade is divided by GDP, which is itself measured as value added. We note that the sample size falls almost by half in this case. This reduces the

[31] There is some disagreement on what critical values to use for the tests of weak instruments reported here. The Stock and Yogo critical values are only strictly appropriate when errors are IID. In the presence of heteroskedasticity, it is common practice to use these critical values as well but only as a rough guideline. It is also common to use the rule of thumb of a value of 10. In this context we choose to be cautious and we also perform a series of tests that are robust to weak instruments (and to heteroskedasticity). Specifically, we have conducted three such tests for the null hypothesis that the coefficients on the MSH and the TSH are jointly equal to zero, namely, the Anderson–Rubin chi-square and F tests, and the Stock and Wright S test. In two cases we reject the null of no joint effect at the 10% significance level and, in the third case the p-value is 0.11. This provides additional evidence of a significant effect of our regressors on income per person.

[32] We also note that the OLS estimate for the effect of migration is similar to the corresponding instrumental-variables estimate (column 5). This suggests that endogeneity and measurement error OLS biases may be of similar magnitude.

[33] We have also performed the weak-instrument robust inference on the coefficient of the MSH in specification 8. The resulting confidence set is [1.77, 11.47], which does not contain the zero value. The interval is rather wide, thus there is a fair amount of uncertainty regarding the exact magnitude of the effect.

[34] Recent studies such as Bems et al. (2010) and Johnson and Noguera (2012) have emphasized that with the increasing fragmentation of production across countries, (gross) trade flows may be much larger than the value added content of trade, especially for those countries doing a lot of processing trade.

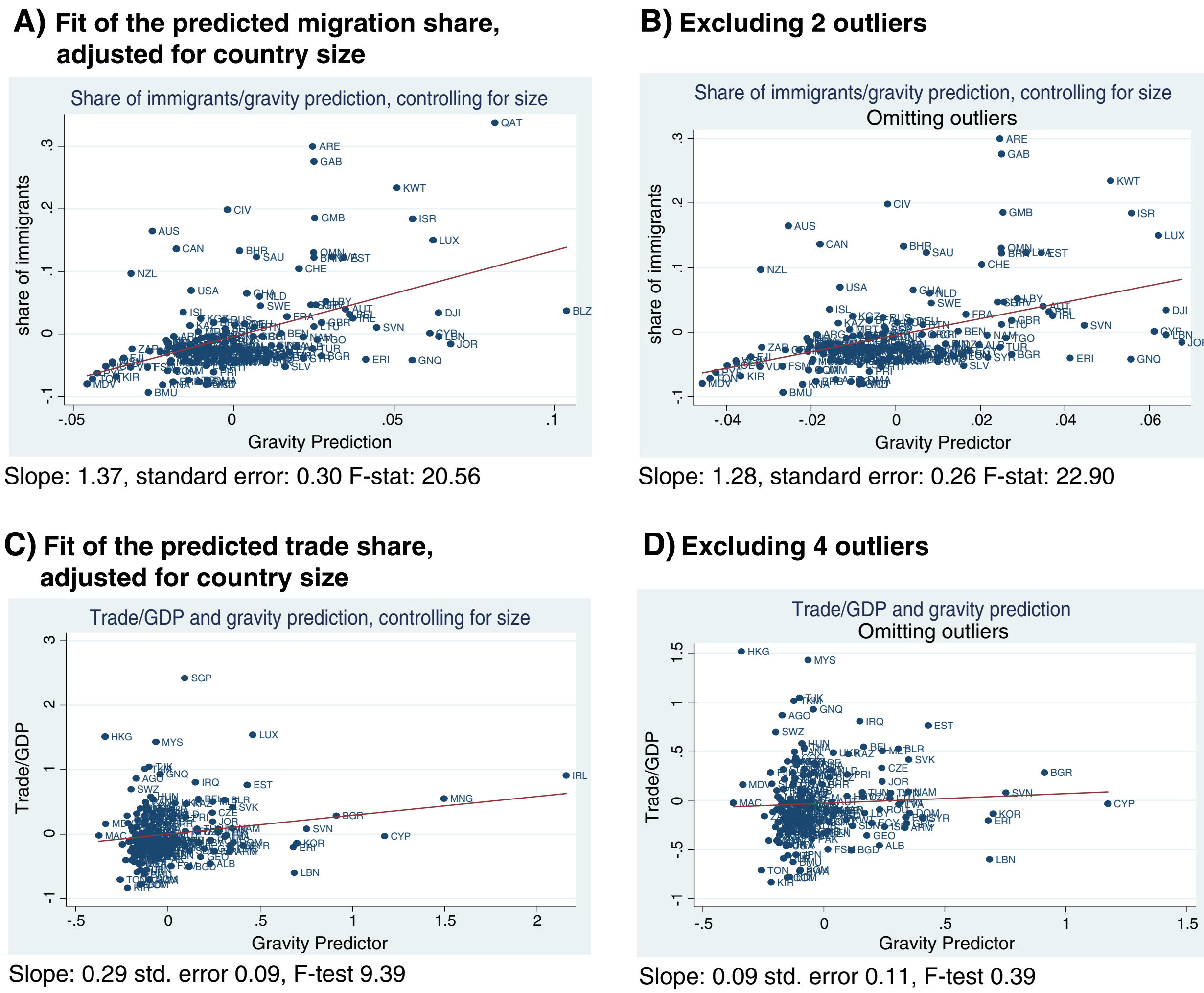


Fig. 3. A: Fit of the predicted migration share, adjusted for country size. B: Excluding 2 outliers. C: Fit of the predicted trade share, adjusted for country size. D: Excluding 4 outliers.

precision of the estimates. Nevertheless, we still find a positive and significant effect of the migration share and no evidence of an effect of openness to trade. Finally, we observe a stronger first stage in column 8, when only one endogenous regressor is present in the regression model. The F-statistic of the first stage is now above 12.

Let us illustrate here the magnitude of the effect of openness to immigration on income per person using a point estimate of 7.32 (the median estimate in Table 4). Assuming a causal interpretation of our estimates, if country A has a migration (foreign-born) share that is 10 percentage points larger than country B, we would expect it to have a long-run level of income per capita that is about twice the level of country B.[35] Ten percentage points in the migration share is the difference between the migration shares of the tenth and ninetieth percentiles in the country distribution by income per capita. By way of comparison, Hall and Jones (1999) reported that cross-country differences in schooling levels would account for income per capita differences between these two groups of countries by a factor or 3. Before concluding this section we conduct three important robustness exercises on our measure of openness to migration.

5.2. Human capital

Thus far our analysis has ignored educational differences among natives and immigrants by measuring openness to migration as the migrant (foreign-born) share in the population. Here we take into account that the human capital content of immigration flows, relative to natives, varies across countries. It is possible that migrants with high education generate a larger contribution to income than those with lower education levels. On the other hand, several authors have argued that migrants' formal education is only a rough measure of the productive skills of immigrants (Dustmann et al. (2013)). To investigate this question we distinguish between individuals with a college degree and those without, and we compute the share in the human capital of a country that is accounted for by its foreign-born population.

Specifically, we assume that the average college-educated worker has higher efficiency units of labor than the average non-college educated worker. These efficiency units are assessed following a simple Mincerian approach as in Hall and Jones (1999). We assume that the return to each additional year of education is 6.8% in terms of wages across all countries in our sample. Assuming that the average gap in years of schooling between college educated and non-college educated workers is 6 years, we obtain that the efficiency units of skilled workers are 1.503 times the units for unskilled workers.[36]

Column 1 in Table 5 reports estimates for the main specification (that is, column 8 in Table 4, which includes all the controls) but using now the migration share in terms of human capital, rather than in

[35] Since $e^{0.1 \times \beta} = e^{0.1 \times 7.3} = 2.08$.

[36] We define country c's stock of human capital as $H_c = U_c + 1.503 * S_c$, where S_c and U_c denote the number of college graduates and non-college graduates in the population, respectively.

Table 5
Human capital, net migration and heterogeneous effects, 2SLS estimates.

Specification	Human capital	Emigration	Net immig.	TSH	Human capital
	(1)	(2)	(3)	(4)	(5)
MSH		7.75***			
		(2.11)			
MSH HK	3.60***				
	(1.17)				
Emig./Pop.		1.38**			
		(0.63)			
Imm. − Emig./Pop			5.50**		
			(2.15)		
MSH * High				7.50***	8.08***
				(1.85)	(1.86)
MSH * Low				6.31**	3.52
				(3.05)	(4.71)
Observations	119	119	119	118	118
R-squared	0.82	0.84	0.77	0.85	0.86
Controls					
Region dummies	Yes	Yes	Yes	Yes	Yes
Geo/climate/disease/oil	Yes	Yes	Yes	Yes	Yes
Colonial origin	Yes	Yes	Yes	Yes	Yes
First stage regression					
KP weak ident. F test	19.06	12.02	9.75	4.51	2.61
Instruments:	Pred. MSH	Pred. MSH	Pred. MSH	Pred. MSH * High	Pred. MSH * High
				Pred. MSH * Low	Pred. MSH * Low
SY 10% max IV size	16.38	16.38	16.38	7.03	7.03
SY 25% max IV size	5.53	5.53	5.53	3.63	3.63

Note: The dependent variable in all these specifications is the log of income per capita. Predicted values for the Migration share are based on the non-linear gravity estimates. All regressions include an intercept, log population, log area, and distance to the equator. For columns 4–5 countries are classified as having a high (low) value if they are above (below) the median along the respective dimension. The instruments are the gravity-predictor of share of foreign-born interacted with the dummy "high" and "low" for the considered dimension. Regional dummies for sub-Saharan Africa, East Asia, and Latin America. Other geography, climate and disease controls include the percentage of land in the tropics, a landlocked dummy, average distance to the coast, average yearly temperature, average yearly humidity, an index of soil quality, an index of the incidence of malaria, and an index of the incidence of yellow fever, and a measure for oil reserves. Colonial Controls includes dummy variables for former French colony, former English colony, and a dummy for the group of 4 young, rich countries (US, Canada, Australia and New Zealand), and the share of 1900 population of European origin. In parenthesis we report the heteroskedasticity-robust standard errors. *, **, and *** = significant at 10%, 5% and 1% confidence level. The KP weak identification test is the Kleibergen–Paap rk Wald F statistic. The reported critical values are those provided by Stock and Yogo (2005) under the assumption of IID errors.

terms of population. The estimates confirm our previous finding: the migration share has a positive and very significant effect on income per capita. Compared to the estimates in column 8 in Table 4, both the point estimate and the associated standard errors have been cut in half. The scale of the coefficient is smaller because the range of variation of the human capital migration share is larger (almost twice as large) than the analogous measure in terms of population. However, the implications in terms of accounting for cross-country disparities in income per capita remain essentially unchanged.

5.3. Net migration

Our measure of openness to migration is based on *gross* immigration. Namely, it is the foreign-born share in the country, which does not take into account the magnitude of emigration flows of natives to the rest of the world. Next we examine whether *emigration* has an additional effect on income, and whether the implications for net immigration, defined as immigration minus emigration, differ from those of gross immigration. Let us first focus on the emigration share. It is possible that emigration has negative effects on income per capita by depriving a country of valuable skills (brain drain). However, the converse may also be true since remittances, human capital gains from return migration, and the transfer of knowledge through the diaspora may compensate for the loss of workers (Stark et al., 1997; Beine et al., 2008).[37]

Column 2 in Table 5 estimates a regression model that features both the immigration share (based on gross immigration) and the emigration share, defined as the number of emigrants over the country's total population, besides the extensive set of controls used in the previous table. We note that the point estimate for the immigration share remains largely unaffected and the coefficient on the emigration share is positive and significant. We also note that here we are treating the emigration share as an exogenous regressor.[38] Since it is very likely that this regressor is correlated with unobserved determinants of income per person, we do not wish to draw any strong conclusions from this estimate. Nevertheless, one may argue the estimated coefficient should be seen as a lower bound for the true effect (as higher income should be correlated with lower emigration).[39] Column 3 in this table presents a possibly more interesting set of estimates. The key regressor here is the net immigration share, defined as stock of immigrants minus the stock of emigrants divided by the country's total population. We treat this as an endogenous regressor and we still instrument it with the gravity-based predictor for the (gross) immigration share. We find that the net immigrant share has a positive effect on income. The coefficient is only slightly lower than the one we found using MSH as the main regressor (5.50 versus 7.75), and we cannot reject equality. Hence, this provides an important robustness check on our main finding.

5.4. Heterogeneous effects

Next we explore whether the effects of immigration on income per capita differ across host countries in any systematic way. For example, it is possible that countries that are more open to international trade, or that have a more educated labor force, benefit to a larger extent from openness to immigration.

To address this question we classify countries between those that have high and low levels of openness to trade and of average human

[37] di Giovanni et al. (2012) argue that the gains from remittances more than compensate for the loss of labor associated to emigration.

[38] We tried to instrument the emigration share of a country with the gravity-predicted emigration, but this turns out to be a very weak instrument when including both emigration and immigration as endogenous variables.

[39] It is also worth noting that emigration data are typically of poorer quality than immigration data so there is potentially more measurement error.

capital per person. Then we estimate an augmented model where we allow for the effect of immigration on income to differ for these two sets of countries. Specifically, we interact MSH with dummy variables for "high" (above the median) and "low" (below the median) values along each one of these dimensions. Analogously, we expand the vector of instruments by including interactions of the predicted migration share with the same dummy variables. The resulting estimates are reported in columns 4 and 5 in Table 5. On the basis of column 4, we find that immigration appears to have similar effects on all countries, regardless of their openness to trade. The point estimate is slightly lower for low-trade-openness countries but we cannot reject equality of the two coefficients. Turning now to column 5, the estimates here reveal a positive and significant effect of migration on countries with above-median levels of human capital in their labor force. With regard to low human capital host countries, the point estimate is still positive, but much smaller (3.5 versus 8.1) and not statistically significant. This suggests that countries with a highly educated domestic labor force benefit more from immigration. This result is reminiscent of research finding that countries endowed with higher human capital are better at absorbing knowledge created abroad. This will be the case, for instance, if immigrants are vehicles of knowledge and ideas.

6. Institutions and ancestors

In our previous empirical model we included a fairly complete set of controls to address the Rodriguez and Rodrik critique. According to Hall and Jones (1999), Acemoglu et al. (2001) and Rodrik et al. (2004), the main reason why geographical variables are relevant (latitude, in particular) is that geography was decisive in determining a country's history of colonization. They argue that those initially less-developed countries that were colonized by a Western European power through long-term settlements were endowed with good institutions. Since good institutions beget good institutions, those countries are likely to enjoy high institutional quality today in the form of well-functioning markets, protection of property rights, and constraints on the power of government which contributed substantially to their economic success. On the contrary, countries that were colonized but not settled by Europeans experienced "exploitative" early institutions, which became a persistent burden on their economic development, lacking checks and balances and furthering concentration of power.

It is also plausible that good early institutions may have led to sustained openness to international trade and migration. Since our predictors for the trade and migration shares are based on geography, which influenced a country's history of colonization and its resulting institutions, it is important to attempt to separately identify the roles of economic openness and good institutions on income. Our measure of institutional quality follows Acemoglu et al. (2001) and is the average between their indices for "protection against expropriation risk" and "constraints on the executive". Both are measured over the period 1975–85. These indices capture some fundamental aspects of protection of private property rights and the limitation of the power of government, which have been found to be crucial for an institutional setting conducive to economic growth.[40] Of course, institutional quality is likely to be endogenous to economic development. Following Hall and Jones (1999) and Alcalá and Ciccone (2004), we complement the gravity-based predictors for openness (to trade and migration) with plausibly exogenous determinants of early institutions. Namely, distance from the equator and the share of the population of European descent measured in 1975. The former has been shown to affect the odds of having been settled by a European power. The latter provides a measure of the degree of social, economic and cultural influence from Europe, and is likely a good proxy for the size of historical European settlements in the country.

[40] The value of this index ranges between 0 and 8.

Table 6 reports the 2SLS estimates. Column 1 considers the roles played by the share of immigrants and our index of institutional quality on income per person, considering both as endogenous regressors. We include regional dummies and controls for the whole set of geography, climate, disease environment, and natural resource variables. Note that, unlike in the previous section, here we do not control for distance to the equator (as it is used as an excluded instrument) or dummies for the colonial past of a country (as its influence is mainly through institutions). Both the migration share and institutional quality are highly significant, with coefficients of 8.4 and 0.4, respectively and t-statistics of 4 and 5. In our sample the difference in the institutional quality index between the 90th and 10th income percentiles is around 6. Based on our point estimate, the resulting income difference explained by institutions is equal to a factor of 14.[41] In comparison, the migration share accounts for a factor of about 2 in the income gap between countries in the 90–10 percentiles. Hence, while institutions still appear to be the main determinant of income per capita disparities, openness to immigrants also has a sizable and distinct contribution. The total difference in income per capita between the 10th and the 90th percentile is a factor of 38, so institutions and migration together explain a substantial part of that gap ($14 * 2 = 28$).

While the vector of instruments is not strikingly strong, we do reject the null of weak instruments at the less stringent critical values using the Stock and Yogo statistic. The second column reports the estimates when we include the exogenous predictors of institutions (distance from the equator and share of population of European descent) directly as regressors, rather than using them as instruments. The positive and very significant effect of the migration share on income per capita hardly changes. In column 3 we perform a joint estimation of the effects of migration, trade and institutions, treating all of them as endogenous regressors. The trade share is not significant and including it hardly affects the point estimates of the migration share and institutional quality.

Next we examine another channel that may mediate the relationship between migration and development. Putterman and Weil (2010) argue that the birthplace of a country's ancestors has highly persistent economic effects. In particular, countries whose ancestors originated from countries with high early socioeconomic development, as measured by political and administrative institutions around year 1500, have higher income per capita in year 2000.[42] This finding suggests that historical migration played a crucial role as a vehicle for the dissemination of institutions. While related, this argument differs from ours. Their channel is fundamentally based on the countries of origin of the ancestors of today's *native* population and operates through the quality of institutions. In contrast, our MSH variable is the share of *immigrants* (foreign-born) in the current population and it is significant in a regression model where we control for institutional quality. In a way they focus on the effects of historical migration while our emphasis is on the effect of current migration flows.

We distinguish the two mechanisms in the following way. First, we use the data by Putterman and Weil (2010) to directly control for the long-run effect of migration through institutions. Namely, we include the ancestor-adjusted quality of political institutions before year 1500 (the so-called *Statehist* variable) and the share of the current population whose ancestors lived in a foreign country circa 1500.[43]

In columns 4 through 6 of Table 6 we introduce these two controls. As we see in column 4, the 'Statehist' variable is positive and significant,

[41] This is calculated as $\exp(0.4 * 6)$.

[42] Putterman and Weil (2010) also offer suggestive evidence indicating that greater variety in the composition by origin of ancestors may have had an additional positive effect. We return to this point later on.

[43] The raw *Statehist* variable is an index, ranging between 0 and 1, capturing the (discounted) length of time prior to year 1500 since the country had developed a supratribal government. The ancestor-adjusted variable, say for the US, is a weighted average of *Statehist* across all countries in the world, where the weights correspond to the shares by country of origin of the ancestors of the current US population around year 1500. The exact definition can be found in pages 1640 and 1641 of Putterman and Weil (2010).

Table 6
Institutions and early development, 2SLS estimates.

Specification	Main	Exog. Instit.	TSH, MSH, IQ	Putt.–Weil	PW	PW	PW	AC	AC
Dep. Var.	Ln GDP/pop	Ln GDP/pop	Ln GDP/pop	Ln GDP/pop	Ln GDP/pop	Ln GDP/pop	Inst. Qual.	Ln GDP/emp	Ln GDP/emp
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
MSH	8.39***	7.07***	7.54***	12.18***	9.77***	6.17**	−3.11		5.18***
	(2.09)	(2.08)	(1.84)	(3.28)	(2.74)	(2.51)	(3.64)		(1.55)
Institution quality	0.45***		0.44***		0.48***			0.23	0.26***
	(0.09)		(0.09)		(0.11)			(0.15)	(0.09)
TSH			0.27						
			(0.47)						
Ln real TSH								2.09***	0.33
								(0.71)	(0.34)
PW statehist.				2.54***	0.05	0.02	0.08		
				(0.50)	(0.44)	(0.37)	(0.82)		
PW sh. foreign ancestors				−0.05	−0.56	0.81**	2.13***		
				(0.33)	(0.50)	(0.35)	(0.57)		
Dist. equator		0.01*				0.03***	0.08***		
		(0.01)				(0.01)	(0.02)		
Sh. Euro. descent 1975		0.01***							
		(0.00)							
Observations	117	119	117		114	116		128	117
R-squared	0.78	0.84	0.80		0.75	0.84		0.21	0.83
Controls									
Region dummies	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes
Geo/climate/disease/oil	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes
Colonial origin	No	No	No	No	No	No	Yes	No	No
First stage regression									
KP weak ident. F test	8.99	9.32	4.65	8.68	6.24	7.32	16.67	2.12	3.7
AP weak ident. F-test for MSH	7.09	9.32	8.74	8.68	6.29	7.32	16.67		7.95
Instruments:	Pred. MSH	Pred. MSH	Pred. MSH	Pred. MSH	Pred. MSH	Pred. MSH			Pred. MSH
	EUdes75		EUdes75		EUdes75			EUdes75	EUdes75
	Dist. Equa.		Dist. Equa.		Dist. Equa.			Dist. Equa.	Dist. Equa.
			Pred. TSH					Ln pred. TSH	Ln pred. TSH
SY 10% max IV size	13.43	16.38		16.38	13.43	16.38	16.38	13.43	
SY 25% max IV size	5.45	5.53		5.53	5.45	5.53	5.53	5.45	

Note: The dependent variable is the log of income per capita or the log of income over employment. Unless noted otherwise the predicted TSH is based on the linear-in-logs gravity estimates and the predicted MSH is based on the non-linear estimation. All regressions include an intercept, log population, log size, and the percent of land in the tropics. PW refers to Putterman and Weil (2010). We use two variables from their study: "Statehist" and the share of foreign ancestors around year 1500. AC refers to Alcalá and Ciccone (2010). EDES75 refers to the share of European descendents in 1975. Regional dummies for sub-Saharan Africa, East Asia, and Latin America. Geography, climate, and disease controls include the percentage of land in the tropics, a landlocked dummy, average distance to the coast, average yearly temperature, average yearly humidity, an index of soil quality, an index of the incidence of malaria, an index of the incidence of yellow fever, and a measure for oil reserves. Predictors based on fixed-effects gravity regression do not use the estimated fixed effects. In parenthesis we report the heteroskedasticity-robust standard errors. *, **, and *** significant at 10%, 5% and 1% confidence level. The KP weak identification test is the Kleibergen–Paap rk Wald F statistic. The AP weak identification test is the Angrist and Pischke (2009) test for weak identification of individual regressors. The reported critical values are those provided by Stock and Yogo (2005) under the assumption of IID errors.

confirming the finding by Putterman and Weil (2010).[44] At the same time the coefficient on the immigration share, capturing current mobility, is still extremely significant and large. In column 5 we include institutional quality plus an extensive vector of controls. In this case neither of the two early history variables appears to be significant. Still the positive effects of the migration share and institutional quality remain fairly unaffected. This is not surprising since the mechanism emphasized by Putterman and Weil (2010) operates through the role of institutions and, hence, controlling for that diminishes its effect. When we do not control for the current quality of institutions (column 6), one of the Putterman and Weil variables does feature a positive and significant point estimate. To strengthen this point column 7 reports the estimates of a regression model with *institutional quality* as the dependent variable. Both of the Putterman and Weil variables are now highly significant even though we are including a very demanding set of controls. The current migration share, instead, is not a determinant of the quality of institutions. Hence, the historical mobility variables of Putterman and Weil (2010) affect income through their effect on institutional quality. The current migration share, proxying for recent cross-border mobility, seems to affect income per capita above and beyond the effect of institutions.

Finally, we address an interesting measurement issue regarding trade openness. Alcalá and Ciccone (2004) argue that because the prices of tradable goods are similar across countries while the prices of non-tradable goods are not, it is more appropriate to build the trade share by dividing total trade by PPP income, rather than by income in US dollars. They call this measure *real openness*. Column 8 reports the estimates of a specification analogous to the one they use in their study. The estimate for the (log of) real openness is positive, highly significant and similar to theirs. However, when we introduce our controls we are unable to reject the null of a zero effect (not shown here) on real openness. Column 9 includes also the migration share, institutional quality and controls for region, geography, climate, disease environment, and oil resources. The estimates show that the non-significant effect of trade openness that we find does not depend on measuring the trade share in nominal or real terms.

Our findings confirm that a country's history of migration had a historically important role in economic development by shaping its institutions, as first noted by Putterman and Weil (2010). However, in addition to this, contemporary levels of migration appear to increase income through channels other than institutions. In the remainder of the paper we discuss a potential explanation for that positive effect.

7. The channel: gains from diversity

The previous sections have provided evidence consistent with a large, positive effect of the immigration share on income per capita. We have also shown that the estimated effect is distinct from those of the current quality of institutions and a country's early migration

[44] Note that we are only controlling for country size and the current migration share. When we control for distance to the equator and regional dummies the Putterman and Weil regressors lose significance.

Table 7
Channels. The Hall and Jones, 1999 Decomposition, 2SLS Estimates.

Dep. Var.	ln Y/L	$(\alpha/1-\alpha) * \ln K/Y$	ln H/L	ln TFP	ln Y/L	ln TFP	Gini coeff.	P90/P10
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
MSH	7.52***	1.01	1.29	5.22**	9.01***	6.77***	−0.38	−82.56
	(2.49)	(1.00)	(0.80)	(2.58)	(2.48)	(2.34)	(0.42)	(54.99)
Instit. quality					0.48***	0.53***		
					(0.12)	(0.13)		
Observations	99	99	99	99	99	99	103	59
R-squared	0.84	0.31	0.77	0.77	0.79	0.65	0.59	0.53
Controls								
Region dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geo/climate/disease/oil	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Colonial origin	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First-stage regression								
AP weak ident. F-test for MSH	8.47	8.47	8.47	8.47	4.94	4.94	10.74	7.15
KP weak ident. F-test	8.47	8.47	8.47	8.47	6.29	6.29	10.74	7.15
Instruments:	Pred. MSH	Pred. MSH	Pred. MSH	Pred. MSH	Pred. MSH	Pred. MSH	Pred. MSH	Pred. MSH
SY 10% max IV size	16.38	16.38	16.38	16.38	13.43	13.43	16.38	16.38
SY 25% max IV size	5.53	5.53	5.53	5.53	5.45	5.45	5.53	5.53

Note: Dependent variables for columns 1–6 are normalized by the US value. Coefficient alpha is the capital share in the Cobb–Douglas production function underlying this decomposition (Hall and Jones, 1999). We have assumed alpha equal to 0.33. In columns 7 and 8 the dependent variables are the Gini coefficient and the 90–01 income percentile ratios. All regression models include an intercept and control for log population and log area, and distance to the equator (except for the last two columns where we treat institutional quality as an endogenous regressor and distance to the equator is part of the vector of instruments). Regional dummies for sub-Saharan Africa, East Asia, and Latin America. Geography, climate and disease controls include the percentage of land in the tropics, a landlocked dummy, average distance to the coast, average yearly temperature, average yearly humidity, an index of soil quality, an index of the incidence of malaria, an index of the incidence of yellow fever, and a measure of the country's oil reserves. Colonial controls include dummy variables for former French colony, former English colony, and a dummy for the 4 rich "young" countries (US, Canada, Australia and New Zealand), and the share of 1900 population of European origin. The predicted migration share is based on the non-linear Poisson pseudo-ML estimator. In parenthesis we report the heteroskedasticity-robust standard errors. *, **, and *** significant at 10%, 5% and 1% confidence level. The KP weak identification test is the Kleibergen–Paap rk Wald F statistic. The AP weak identification test is the Angrist and Pischke (2009) test for weak identification of individual regressors. The reported critical values are those provided by Stock and Yogo (2005) under the assumption of IID errors.

history, both of which have been emphasized as key factors in accounting for cross-country income disparities. The goal of this section is to provide a potential explanation for the channel that mediates the estimated effects of openness to migration on income. In particular we investigate a hypothesis based on gains from diversity. We postulate that countries may benefit from a more diverse immigrant population because this increases the available variety of skills, abilities and ideas, which may, in turn, increase average labor productivity in the long-run.

7.1. Decomposing the effect

Following Hall and Jones (1999) and Alcalá and Ciccone (2004), we postulate a simple Cobb–Douglas aggregate production function in which output is produced using human capital and physical capital. Income per worker (rather than per person) can be decomposed into: physical capital intensity, human capital intensity, and total factor productivity. Specifically,

$$lny_c = \frac{\alpha}{1-\alpha} ln\frac{K_c}{Y_c} + lnh_c + lnTFP_c, \tag{9}$$

where α is the labor share in income, which we set equal to 0.33, K_c/Y_c is the capital–output ratio, $h_c = \exp(\gamma S_c)$ is the average human capital per person, calculated as the exponential of average years of schooling times its Mincerian return. Finally, TFP_c is the total factor productivity, calculated as a Solow residual. The data on physical capital and output per worker are obtained from the Penn World Tables while the data on average schooling are from the Barro and Lee (2011) and the Cohen and Soto (2007) databases.[45]

Table 7 reports the 2SLS estimates for a series of models where the dependent variables are, in turn, the log of income per worker, and the terms on the right hand side of Eq. (9): the log of the capital–output ratio, the log of human capital per person, and the log of TFP (columns 1 through 4). Our main regressor of interest is the migration share and we include the same extensive set of controls as in the previous tables. A clear pattern emerges from these estimates: the migration share has a positive and high significant effect on income per worker, which operates through total factor productivity. We find no evidence of an effect of immigration on capital intensity, consistent with the prediction of the neoclassical growth model stating that changes in the size of the labor force will not affect capital per worker in steady state. Likewise we do not find an effect on the level of human capital per person in the receiving country. Columns 5 and 6 reproduce the specifications in columns 1 and 4, respectively, but we now include institutional quality as an endogenous regressor. This does not affect the finding that the main reason why immigration increases income per capita is its positive effect on total factor productivity.

The last two columns of Table 7 examine whether immigration affects the degree of income inequality in the country.[46] We consider two measures of inequality, the Gini coefficient and the 90–10 ratio of income percentiles. In neither case we find evidence of an effect. This is consistent with the previous findings that immigration alters neither the physical capital intensity nor the human capital intensity of the receiving economy. Those channels could alter the relative wage between more and less educated individuals or the return to capital relative to labor and lead to distributional effects on income. To the contrary an effect through TFP does not need to generate inequality. Recent studies have examined other mechanisms that may account for the mitigated effect of immigration on the wage structure (Manacorda et al., 2012; Ottaviano and Peri, 2012; Lewis, 2003; Gonzalez and Ortega, 2010, among many others).[47] The lack of an effect of immigration on the income distribution (beyond its effect on the mean) is consistent with the diversity channel that we focus on next.

[45] Where available the data on years of schooling have been obtained from the most recent version of the Barro and Lee (2011) database. For a dozen countries for which the information is not available in that database we rely on Cohen and Soto's (2007) data, available at their personal website. Following Hall and Jones (1999) all dependent variables have been normalized by the US value.

[46] There is an abundance of papers on the effects of international trade on income inequality. The debate has been reignited by the rise of trade flows with China and the public debate on the pros and cons of globalization. See Richardson (1995) for a survey and the recent studies by Autor et al. (2012) and Levchenko and Zhang (2012).

[47] For a good review of the literature see Raphael and Ronconi (2007). Few studies have examined the role of both international trade and migration. One influential contribution using US data is Borjas et al. (1992, 1997) and the response by DiNardo and Abowd to the 1997 article.

7.2. Diversity in skills

One way in which immigration may increase TFP and labor productivity is by increasing the diversity of skills and ideas in the labor force. The simplest way to conceptualize this is to consider that workers of different origin are differentiated factors of production, in the fashion of the Armington (1969) model of international trade.[48] For instance, this may reflect differences across countries in their social norms, language, cultural values, social prestige attached to science or the arts, and so on. In the context of trade flows, Broda and Weinstein (2006) find productivity gains arising from trading a wider variety of goods differentiated by country of origin. In the context of migration, Ottaviano and Peri (2006) find that U.S. cities with more diverse immigrant population (by country of origin) have higher productivity. Peri and Sparber (2009) find that immigration affects the supply of differentiated tasks and induces task-specialization that produces efficiency gains through deepening comparative advantage.[49] In related but independent research Alesina et al. (2013) emphasize the role of birthplace diversity as a source of productivity gains.[50]

We construct our *immigrant diversity index* by country of origin as follows.[51] The starting point is a breakdown of the foreign-born population in each (host) country according to birthplace. Consistent with the notation that we used in our bilateral migration regressions, we denote by MSH_{jc} is the share of individuals born in country j in the total foreign-born population residing in country c. Then we compute the following index:

$$DivM_c = 1-\sum_{j\neq c}\left(MSH_{jc}\right)^2, \tag{10}$$

where the summation range spans all countries in the world (with available data), except for host country c. A value of the index equal to zero indicates that one single origin country accounts for all foreign-born population, hence minimum diversity. This would be the case, for instance, if all immigrants in the US were born in Mexico. Higher values of the index correspond to a more balanced distribution of immigrants by country of origin. When all countries of origin have similar shares, the index reaches its highest value, 1-1/N representing maximum diversity. We also build an analogous diversity index for trade flows using the share of trade with a specific country relative to total trade. Before turning to the analysis, let us comment briefly on some features of the diversity index in our data. The value for the US is 0.91, which will be a useful benchmark. This value indicates that migration flows into the US are fairly diverse. Mexico plays a clearly dominant role in US immigration, however it is important to note that the US also hosts immigrants originating in all other countries in the world, and the shares of these countries in the total immigrant population in the US are fairly balanced. Several countries attain higher values: Israel (0.94), Spain (0.94), the UK (0.96), Denmark (0.96), and Canada (0.96), to name just a few. Many countries display much lower values: Bangladesh (0.06), Pakistan (0.09), India (0.60), and among OECD countries, Greece (0.70), or Japan (0.75).[52]

We now turn to the formal analysis. Table 8 reports two-stage least-squares estimates of models that include the diversity index. The first column simply reproduces the basic specification from Column 5 in Table 4: the migration share has a positive and significant effect on income, while the point estimate of the trade share is low and statistically insignificant. Column 2 adds the migration diversity index as a control. The coefficient for this variable is positive and highly significant and it reduces by almost 50% the point estimate for the effect of the immigration share. This suggests that the effect of immigration on income operates, at least in part, through the diversity channel. Column 3 adds the diversity index for trade. The point estimates for the migration share and the migration diversity index remain largely unaffected, although now the coefficient on MSH is marginally significant, and the point estimate for diversity in trade flows is in fact negative and marginally significant.

Before reading too much into these estimates we note that the diversity indices may be correlated with other determinants of income and certainly contain substantial measurement error. Ideally, one would like to treat them as endogenous regressors. In practice though the gravity predictors for the diversity indices perform poorly, thus hindering attempts at estimation by instrumental variables. To address this shortcoming we adopt a direct-regression approach and we use the gravity predictors for the diversity indices directly in the regression, rather than as instruments, while the trade and migration shares are still considered as endogenous and instrumented for. This is our preferred set of estimates and we report them in column 4. The migration share displays a positive and very significant effect, falling slightly from a coefficient of 7.3 down to 6.7. Furthermore, immigrant diversity has an additional positive and significant effect on income per capita. To the contrary diversity of trade flows continues to have a negative effect on income per person.[53] The income effect of immigrant diversity is large. An increase in the diversity of migrants from 0.05 (the value for Sri Lanka, whose immigrants are essentially all from India) to 0.96 (the value for the UK) implies a corresponding increase in output per person by a factor of 3.5. Beyond the positive effect of immigration, these results suggest that a diverse immigrant population has an additional positive effect on income per capita.

7.3. Ethnic fractionalization

The previous results suggest that large and diverse migration flows increase long-run income per capita. However, there may also be negative by-products associated with large and diverse migration flows. In particular, it may lead to ethnic or linguistic fractionalization, which has been related to conflict and under-provision of public goods.[54] Alesina et al. (1999) provide evidence indicating that ethnic or linguistic fractionalization increases conflict and reduces solidarity, leading to a reduction in the provision of public goods.[55] However, Alesina et al. (2003) examine the consequences of different types of fractionalization (ethnic, linguistic and religious) for economic growth and several other economic outcomes. While they find effects of ethnic and linguistic fractionalization on some economic outcomes (corruption, political rights), they report that these effects appear to be sensitive to the specification used, and they find much weaker and not consistent effects on economic growth.

[48] This idea is formalized in a simple model in the Appendix A.

[49] Amuedo-Dorantes and de la Rica (2011) largely confirm these findings using data for Spain, which experienced a very large immigration wave over the last decade.

[50] Alesina et al. (2013) argue that diversity among skilled immigrants has a larger productivity effect, particularly for rich host countries.

[51] Here we define the index for migration but we also build an analogous measure for trade flows. This index corresponds to one minus the Herfindahl–Hirschman concentration index.

[52] Here are some observations regarding the diversity of trade flows, denoted by *DivT*. We again use the US as the benchmark (value of 0.92). Several rich countries have more diverse trade flows, such as France (0.93), the UK (0.94), or Germany (0.95), reflecting the low trade costs within Europe. However, the countries with the highest values tend to be low income (Pakistan, India, Kenya or Tanzania are all in the top 10). At the other extreme, Mexico (0.39) and Canada (0.43) display very low values of the trade diversity index, reflecting the dominant position of the US as their main trading partner.

[53] We have also estimated a specification where we treat migration diversity as an endogenous regressor, instrumented with its gravity predictor and we omit the migration share from the regression. In that case (not reported in the table) we obtain a point estimate for migration diversity equal to 3.87, significant at the 10% level.

[54] See Alesina and La Ferrara (2005) for a review. Alesina et al. (2013) revisit this question.

[55] Garcia-Montalvo and Reynal-Querol (2005a, 2005b) argue that it is more appropriate to use polarization, as opposed to fractionalization measures. They find evidence of increased conflict and lower economic development.

Table 8
Diversity and fractionalization, 2SLS estimates.

Dep. var.	ln GDP/pop	ln GDP/pop	ln GDP/pop	ln GDP/pop	Ethnic Frac.	Ling. Frac.	ln GDP/pop
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
MSH	7.32***	4.01	4.34*	6.66***	2.04***	2.61***	9.81***
	(2.05)	(2.45)	(2.31)	(1.83)	(0.68)	(1.00)	(2.46)
TSH	−0.01	−0.07	−0.07	−0.02	−0.01	−0.04	
	(0.56)	(0.61)	(0.60)	(0.54)	(0.14)	(0.14)	
Diversity M.		1.30***	1.25***				
		(0.39)	(0.38)				
Diversity T.			−1.35*				
			(0.73)				
Pred. Div. M.				1.29**			
				(0.57)			
Pred. Div. T.				−1.76***			
				(0.44)			
Ethnic Frac.							0.61*
Ling. Frac.							−0.65**
Catholic 1980							0.00**
Muslim 1980							−0.01**
Protest. 1980							0.01***
Observations	119	117	117	117	118	115	113
R-squared	0.84	0.85	0.86	0.86	0.55	0.49	0.86
Controls							
Region dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geo/climate/disease/oil	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Colonial origin	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Instruments:	NL pred. MSH pred. TSH	NL pred. MSH pred. TSH	NL pred. MSH pred. TSH	NL pred. MSH pred. TSH	NL pred. MSH pred. TSH	NL pred. MSH pred. TSH	NL pred. MSH

Note: The predicted TSH is based on the linear in logs gravity estimates and the predicted MSH is based on the non-linear Poisson-ML. Predicted values for the TSH and the diversity (fractionalization) index for trade flows are based on the linear-in-logs gravity estimates and the analogous variables for Migration are based on the non-linear gravity estimates. All regressions include an intercept, log population, log area, and distance to the equator (not shown). Regional dummies for sub-Saharan Africa, East Asia, and Latin America. Other geography, climate and disease controls include the percentage of land in the tropics, a landlocked dummy, average distance to the coast, average yearly temperature, average yearly humidity, an index of soil quality, oil reserves, an index of the incidence of malaria, and an index of the incidence of yellow fever. Colonial controls include dummy variables for former French colony, former English colony, and a dummy for the group of 4 young, rich countries (US, Canada, Australia and New Zealand). Standard errors for some control variables have been omitted for lack of space. In parenthesis we report the heteroskedasticity-robust standard errors. *, **, and *** = significant at 10%, 5% and 1% confidence level.

We examine if immigration affects observed measures of fractionalization (taken from Alesina et al., 2003). Columns 5 and 6 in Table 8 estimate models using ethnic and linguistic fractionalization, respectively, as dependent variables. As one would expect, we find that immigration significantly increases both dimensions of fractionalization.[56] The increase in fractionalization and potential ethnic conflict may offset the aggregate productivity gains from greater variety in terms of skills and ideas. This begs an important question. Does increased ethnic and linguistic fractionalization lead to lower long-run income levels? We address this question as follows. Column 7 in Table 8 reports the estimates of a specification where we explicitly control for ethnic, linguistic, and religious fractionalization. The point estimate associated to linguistic fractionalization is negative and significant. In addition the point estimate for the coefficient on the migration share is 9.8, which is larger than the 7.3 point estimate in the baseline specification (column 1). While we cannot reject the null of equal coefficients across the two models, this suggests that, on average, the negative effects of fractionalization are more than offset by the gains arising from a more diverse labor force. The net effect of immigration on income per person estimated in column 1, in fact, includes the attenuation effect due to the indirect influence of ethnic and linguistic fractionalization, and is still significantly positive.

7.4. Diversity and idea creation

Another channel through which immigration might affect total factor productivity in the long-run is its effect on the rate of innovation of a country. Cross-country differences in their scientific and technological histories and in the structure of their research and academic institutions may shape the way talent is allocated across disciplines or the cognitive or non-cognitive abilities emphasized in the schooling system. As a result, individuals originating from a specific country may be more likely to innovate in some fields than in others. International migration exposes a country to the creative ideas of different people and may result in more innovation. We use data on patenting to measure innovation. While not all innovations are patented and while the patenting rate of innovations depends on the field and sector of discovery, statistics on patents have long been used as a measure of innovation.[57] Here we follow this approach by using patenting data, which is directly related to this question and widely available.

We are not the first in examining the relationship between immigration and their direct effect on innovation and entrepreneurship. Important contributions to this literature are Kerr and Lincoln (2010), Gauthier-Loiselle and Hunt (2010) and Hunt (2011). These studies provide evidence of high rates of patenting activity among the immigrant population in the U.S., compared to natives with similar educational attainment. Similarly, some recent studies link openness to trade to technology adoption and innovation (Bloom et al., 2011). Our data is from the World Intellectual Property Organization (WIPO), which collects data on patents granted by any patent office in the world to inventors residing in 108 countries between 1995 and 2010.[58] We construct the average yearly number of patents per million of inhabitants. In our data, the cross-country mean is 91 patents per year and per million of inhabitants, and ranges from 0.01 to 227.

Consistently with our treatment of other outcomes we consider the log of patents per capita as the relevant measure of innovation per

[56] We have also estimated an analogous specification where the dependent variable is an index of religious fractionalization but did not find a significant effect of migration.

[57] See for instance the book by Jaffe and Trajtenberg (2002).

[58] The data are available at the website http://www.wipo.int/ipstats/en/statistics/patents/.

Table 9
Patenting rates, 2SLS estimates.

Dep. Var.	Ln patents/pop	Ln patents/pop	Ln patents/pop
	(1)	(2)	(3)
MSH	13.37**	9.28*	
	(6.34)	(5.47)	
Instit. Qual.		1.31***	1.33***
		(0.31)	(0.34)
TSH			1.97
			(1.75)
Observations	105	103	103
R-squared	0.76	0.70	0.70
Controls			
Region dummies	Yes	Yes	Yes
Geo/climate/disease/oil	Yes	Yes	Yes
Colonial origin	Yes	No	No
Instruments:	NL pred. MSH	NL pred. MSH Dist. equator Euro. descent 1975	Pred. TSH Dist. equator Euro. descent 1975

Notes: The dependent variables in this table are the average annual patents per million people over the period 1995–2010 granted to applicants residing in the country by any patent office in the world, the log of the previous variable, and the log of the total number of patents. All regressions include an intercept, log population, log area, and distance to the equator. Regional dummies for sub-Saharan Africa, East Asia, and Latin America. Other geography, climate and disease controls include the percentage of land in the tropics, a landlocked dummy, average distance to the coast, average yearly temperature, average yearly humidity, an index of soil quality, an index of the incidence of malaria, an index of the incidence of yellow fever, and a measure of oil reserves. Colonial controls includes dummy variables for former French colony, former English colony, and a dummy for the group of 4 young, rich countries (US, Canada, Australia and New Zealand), and the share of 1900 population of European origin. In parenthesis we report the heteroskedasticity-robust standard errors. *, **, and *** = significant at 10%, 5% and 1% confidence level.

person in a country.[59] The key explanatory variable is the migration share. Table 9 reports our findings. The first column shows a positive estimated effect, significant at the 5% level, of the share of immigrants on log patenting per person, after controlling for the usual geographic and colonial variables. Column 2 includes in the regression the measure of institutional quality (treated as endogenous), and finds a strongly significant effect of institutions on innovative intensity. The point estimate on the immigrant share is reduced somewhat but remains significant at the 10% level. In contrast, the trade share, included in column 3, does not seem to play any role. Acknowledging the relatively high standard errors, these findings provide suggestive evidence that one of the channels through which immigration increases labor productivity in the long run may be by contributing to higher rates of innovation per capita. It is also worth noting that our findings do not necessarily imply that the immigrants themselves produce the whole increase in innovation. Combining different and complementary ideas can also make natives more innovative.

8. Discussion and conclusions

Our empirical findings support the idea that openness to migration, by increasing the range of skills and ideas in the host country, plays a role in accounting for cross-country differences in income per capita, beyond the important roles played by geography, history, and institutions. In our data the ratio between the income per person of the countries in the 90 and 10 percentiles is about 38. Approximately, this corresponds to Ireland and Uganda, respectively. On the basis of our estimates, if Uganda were to adopt immigration policies that equalized its immigration share to that of Ireland, its income per capita in the long run would increase by 70%. In comparison, if Uganda's institutional quality were brought to Irish levels, its long-run income per person would increase by a factor of 8.[60]

Our cross-sectional analysis is not able to uncover a significant role of trade openness on income. Even though we have been able to reproduce the positive effect found by Frankel and Romer (1999), the size and significance of this effect were sharply reduced when including geographical controls, confirming the critique by Rodriguez and Rodrik (2001). Ironically, the technique developed by Frankel and Romer appears to be more successful in identifying the effect of migration on income that turns out to be positive and significant.

In recent work Feyrer (2009a, 2009b) argues that longitudinal (or quasi-experimental) data substantially help identifying the effects of trade openness on income. His point estimates are not directly comparable to ours or to those reported by Frankel and Romer (1999). Feyrer's regression estimates the effect of the log of the total trade volume, not the trade to GDP ratio, on the log of income per capita. Let us denote this trade-income elasticity by b. Let also Y, y, and T denote, respectively, GDP, GDP per person, and total trade volume. Straightforward manipulation of Feyrer's regression model delivers the following relationship between income per capita ratios and trade to GDP ratios between two periods 0 and 1:

$$\frac{y_1}{y_0} = \left(\frac{T_1/y_1}{T_0/y_0}\right)^{\frac{b}{1-b}} = \left(\frac{T_1/Y_1}{T_0/Y_0}\right)^{\frac{b}{1-b}},$$

where we have assumed that the population is not affected by the increase in trade flows. Note that the expression above takes into account a feedback effect: an increase in the volume of trade increases income (per capita) which enter the expression of trade share. Now suppose that there is a one percentage point increase in the trade share so that $T_1/Y_1 = T_0 / Y_0 + 0.01$. Using Feyrer's preferred point estimate $b = 0.58$ (Table 5, Column 1) and evaluating the expression at the median trade share in year 2000 (0.80) the previous equation simplifies to

$$\frac{y_1}{y_0} = \left(\frac{0.81}{0.80}\right)^{\frac{b}{1-b}} = 1.017.$$

Hence, Feyrer's estimates imply that a one percentage point increase in the trade share (evaluated at the median trade share in the sample) leads to a 1.7% increase in income per capita.[61] This estimate is very similar to the original estimates reported by FR (about 2%).

In comparison, our estimates of the effect of migration (as share of the population) imply that a one percentage point increase in the immigration share in the population increases income per person by about 6%. A direct comparison of the semi-elasticities for these measures of trade and migration openness should be done cautiously.[62] However, this comparison emphasizes the quantitative relevance of immigration for productivity and hence our study calls for more research and more rigorous thinking about that relationship between openness to immigration and long-run economic growth. This is likely to be particularly relevant for policies, because international migration remains highly regulated when compared to trade flows, implying large unrealized efficiency gains, as recently emphasized by Clemens (2011).

[59] One may argue that if ideas are public goods the log of total patents in a country might be more appropriate. We are, however, more interested on the effect on the innovation rate, defined as innovations per person. At any rate, the results obtained using log of total patents as outcome (not reported) are consistent with those reported in the Table.

[60] Equalizing the levels of the residual factors, other than the share of immigrants in the population and institutional quality, would increase Uganda's income per capita by a factor close to 3.

[61] Another interesting statistic derived from Feyrer's estimate is the following elasticity: the percentage increase in income per capita associated to a 1% increase in the trade share equals $1.01^{\frac{0.58}{0.42}} - 1 = 0.014$, or 1.4%. Unlike the previous semi-elasticity, this elasticity does not depend on the level of the trade share.

[62] Migration shares are lower and less variable across countries than trade shares. In addition, as discussed earlier, trade to GDP ratios are inflated because the numerator is based on gross, rather than value-added, flows.

Appendix A. A theoretical framework

We present a simple model to justify our main empirical specifications for the effect of openness to international trade and migration on income per person, namely Eqs. (2), (4) and (3). The model is a minor extension of Alesina et al. (2000).[63] Consider N regions in the world, indexed by $i = 1,2,\ldots,N$. These regions are partitioned into C countries. The size of each country, S_c, is given by the number of regions it encompasses. Each region i is endowed with human capital (workers) H_i and physical capital K_i. Each region's capital stock is used to produce a differentiated intermediate good, one unit for one unit. Human capital is also differentiated by country of origin. All regions produce a common final good (used as numeraire) by means of the following aggregate production function:

$$Y_i = A_i\left(\sum_{j=1}^{N} H_{ij}^{\alpha}\right)\left(\sum_{j=1}^{N} X_{ij}^{1-\alpha}\right), \tag{11}$$

where $0 < \alpha < 1$. Expression (11) implies that producers in any region i have access to a full range of varieties for intermediate goods and human capital. H_{ij} denotes the units of human capital of variety j used in production of good i. Likewise, X_{ij} denotes the units of intermediate good j used in region i.

Intermediate goods and workers are geographically mobile but subject to iceberg-type costs. Intermediate goods are shipped costlessly across regions within the same country. However, when Z^X units of intermediate good X are shipped to a foreign region only $(1-\gamma^X)Z^X$ units reach the destination, where $0 \le \gamma^X \le 1$ denotes the cost of shipping internationally as share of the goods' value. We denote by p^i the price charged by the producer of intermediate good i to ship one unit. The shipping costs (zero for domestic shipments) are paid by the buyer. Likewise, there are costs associated to hiring a foreign-born worker. These costs can be thought of as the additional costs of recruiting abroad, sponsoring an immigrant or training costs paid by the employer to help adapt foreign skills to the host economy. When Z^H foreign workers are hired by a firm, only $(1-\gamma^H)Z^H$ units are available for production of the final good, where $0 \le \gamma^H \le 1$ is the immigration cost per unit of human capital. Factors are paid their marginal products. For tractability we impose the following symmetry assumptions: $A_i = A$, $K_i = K$, and $H_i = H$, for all regions $i = 1,2, \ldots, N$. We also assume that all countries have the same number of regions, $S_i = S$,which guarantees the existence of a symmetric equilibrium.

Let us now characterize the demand for domestic and foreign factors of production for a given region i. The marginal product of a unit of intermediate good purchased from a domestic producer from region j *in the same country* is

$$\frac{\partial Y_i}{\partial Z_{ij}^X} = A(1-\alpha)\left(Z_{ij}^X\right)^{-\alpha}\left(\sum_{k=1}^{N} H_{ik}^{\alpha}\right). \tag{12}$$

Let us now compute the marginal product of a unit of intermediate good purchased from *foreign* producer j', keeping in mind that only $X_{ij'} = (1-\gamma^X)Z_{ij'}^X$ units are available for production when $Z_{ij'}$ units are purchased. Then

$$\frac{\partial Y_i}{\partial Z_{ij'}^X} = A(1-\alpha)\left(1-\gamma^X\right)^{(1-\alpha)}\left(Z_{ij'}^X\right)^{-\alpha}\left(\sum_{k=1}^{N} H_{ik}^{\alpha}\right). \tag{13}$$

In a symmetric equilibrium all producers charge equal prices to all destinations (net of shipping costs), that is, $p_j = p_{j'}$. As a result each region purchases equal amounts of all domestically produced varieties (Z_D^X) and equal amounts of all foreign varieties (Z_F^X). Equal prices (net of shipping costs) and profit maximization imply that the marginal products of domestic and imported intermediate capital goods will be equalized:

$$Z_F^X = \theta^T Z_D^X \tag{14}$$

where $\theta^T = (1-\gamma^X)^{1-\alpha/\alpha}$.[64]

In similar fashion, wages (net of migration costs) will be equal across regions in a symmetric equilibrium. Thus profit maximization will lead to equalization of the marginal products of domestic (Z_D^H) and foreign workers (Z_F^H). Thus

$$Z_F^H = \theta^M Z_D^H \tag{15}$$

where $\theta^M = (1-\gamma^H)^{\alpha/(1-\alpha)}$. Let us now turn to the resource constraints for intermediate goods and workers. The stock of capital in a region is used to produce its own variety of intermediate good. Then Z_D^X units are shipped to each region within the same country and Z_F^X are shipped to each region in another country. Similarly, Z_D^H workers will migrate to each domestic region and Z_F^H will migrate to each foreign region. The resulting resource constraints for each variety of human capital and intermediate input satisfy

$$SZ_D^X + (N-S)Z_F^X = K \tag{16}$$

$$SZ_D^H + (N-S)Z_F^H = H. \tag{17}$$

We can use these equations to derive closed-form solutions:

$$Z_D^X = \frac{K}{S+(N-S)\theta^X} \tag{18}$$

$$Z_D^H = \frac{H}{S+(N-S)\theta^H} \tag{19}$$

plus $Z_F^X = \theta^T Z_D^X$, and $Z_F^H = \theta^M Z_D^H$.

Let us now use these expressions to derive the measures of openness to international trade and migration that we will employ in the empirical section. Let us define the trade to GDP ratio, for short, the *trade share* (TSH) as the sum of exports plus imports as a share of GDP in country i[65]:

$$TSH_i = 2(1-\alpha)\frac{Z_F^X(N-S)}{Z_F^X(N-S)+Z_D^X S} = 2(1-\alpha)\frac{\theta^T(N-S)}{\theta^T(N-S)+S}. \tag{20}$$

Clearly, given country size, an increase in trade openness, θ^T, would increase the trade share. And an increase in the size of the country, S, for a given degree of trade openness (as long as $\theta^T < 1$), will reduce the trade share.[66] Expression (20) shows that the trade share also depends on the elasticity of final output to intermediates $(1-\alpha)$ and on the overall size of the world economy N. Similarly, we define the

[63] As these authors show, this static model can be interpreted as the steady state of a growth model. Hence, we stress that our predictions relate openness to long-run income per capita levels across countries.

[64] Note that $Z_F^X < Z_D^X$ as long as $\gamma^X > 0$.

[65] Imports are equal to exports in this model. In symmetric allocations the price of intermediate goods is the same (net of shipping costs) for all regions. Hence, the value of imports plus exports relative to the total value of intermediate goods is equal to twice the ratio of exported quantities relative to total quantities. Coefficient $(1-\alpha)$ is the share of capital in total income in symmetric allocations.

[66] As the model is symmetric across countries, an increase in the country size S should be thought of as an increase in the size of each country (in number of regions), and consequently also as a reduction of the number of countries in the world.

migration share (MSH) as the foreign-born share in the population.[67] That is, for country i,

$$MSH_i = \frac{Z_F^H(N-S)}{Z_D^H S + Z_F^H(N-S)} = \frac{\theta^M(N-S)}{S+\theta^M(N-S)}. \tag{21}$$

It is easy to see that, for a given country size S, the migration share depends positively on openness to immigration. Conversely, for given openness θ^M, the migration share depends negatively on the size of the country. Log linearizing expressions (20) and (21) we obtain expressions (3) and (4) in the text. Finally, substituting Eqs. (18) and (19) into Eq. (11), we can express real GDP in country c, which is constant across regions within a country, as:

$$Y = A\left[S\left(1-\left(\theta^T\right)^{1-\alpha}\right) + N\left(\theta^T\right)^{1-\alpha}\right]\left[S\left(1-\left(\theta^M\right)^{\alpha}\right) + N\left(\theta^M\right)^{\alpha}\right]H^\alpha K^{1-\alpha}. \tag{22}$$

Dividing by the initial population in the region, H_c, we can now compute GDP per capita:

$$y_c = TFP\left(A_c, \theta_c^M, \theta_c^T, S_c\right)\left(\frac{K_c}{H_c}\right)^{1-\alpha}, \tag{23}$$

where we have reintroduced the country subindices for all variables. The first term collects all the determinants of total factor productivity in this model and the second is the factor intensity (capital-labor ratio). The previous expressions make clear that openness to migration θ^M and openness to trade θ^T affect positively TFP and income per person.[68] Similarly, for a given degree of openness, an increase in the size of the country, S, also increases productivity. We note also that this expression allows other factors to also affect TFP, such as government policies, institutions or social norms, which are absorbed in the term A_c. Taking a log-linear approximation if 23 we obtain 2 in the text.

References

Acemoglu, D., 2003. Patterns of skill premia. Rev. Econ. Stud. 70 (2), 199–230.
Acemoglu, D., 2009. Modern Economic Growth. Princeton University Press, Princeton, NJ.
Acemoglu, D., Johnson, S., Robinson, J., 2001. The colonial origins of comparative development: an empirical investigation. Am. Econ. Rev. 91, 1369–1401.
Acemoglu, D., Johnson, S., Robinson, J., 2002. Reversal of fortunes: geography and institutions in the making of the modern world income distribution. Q. J. Econ. 117 (4), 1231–1294.
Adsera, A., Pytlikova, M., 2012. The role of language in shaping international migration. IZA Working Paper 6333.
Alcalá, F., Ciccone, A., 2004. Trade and productivity. Q. J. Econ. 119 (2), 612–645.
Alesina, A., La Ferrara, E., 2005. Ethnic diversity and economic performance. J. Econ. Lit. 43 (3), 762–800.
Alesina, A., Baqir, R., Easterly, W., 1999. Public goods and ethnic divisions. Q. J. Econ. 114 (4), 1243–1284.
Alesina, A., Spolaore, E., Wacziarg, R., 2000. Economic integration and political disintegration. Am. Econ. Rev. 90 (5), 1276–1296.
Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., Wacziarg, R., 2003. Fractionalization. J. Econ. Growth 8 (2), 155–194.
Alesina, A., Spolaore, E., Wacziarg, R., 2005. Trade, growth and the size of countries, In: Aghion, P., Durlauf, S. (Eds.), 1st ed. Handbook of Economic Growth, vol. 1, pp. 1499–1542 (Ch. 23).
Alesina, A., Harnoss, J., Rapoport, H., 2013. Birthplace Diversity and Economic Prosperity. Working Paper 18699. National Bureau of Economic Research.
Amuedo-Dorantes, C., de la Rica, S., 2011. Complements or substitutes? Task specialization by gender and nativity in Spain. Labour Econ. 18 (5), 697–707.
Andersen, T.B., Dalgaard, C.J., 2011. Flows of people, flows of ideas, and the inequality of nations. J. Econ. Growth 16 (1), 1–32.
Anderson, J.E., van Wincoop, E., 2003. Gravity with gravitas: a solution to the border puzzle. Am. Econ. Rev. 93 (1), 170–192.
Angrist, J., Pischke, S., 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press.
Armington, P.S., 1969. A theory of demand for products distinguished by place of production. Staff Pap. Int. Monet. Fund 16 (1), 159–178.

[67] Note that the TSH and MSH are the same across regions within a country.

[68] Recall that N, the number of regions in the world, is obviously larger than S, the number of regions in a given country.

Ashraf, Q., Galor, O., 2013. The "Out of Africa" hypothesis, human genetic diversity, and comparative economic development. Am. Econ. Rev. (forthcoming).
Autor, D., Dorn, D., Hanson, G.H., 2012. The China Syndrome: Local Labor Market Effects of Import Competition in the United States. Working Paper 18054. National Bureau of Economic Research.
Barro, R., Lee, J.W., 2011. Educational attainment dataset. http://www.barrolee.com/.
Barro, R., Sala i Martin, X., 2004. Economic Growth, 2nd ed. MIT Press, Cambridge, MA.
Baum, C., Schaffer, M., Stillman, S., 2007. Enhanced routines for instrumental variables/GMM estimation and testing. Boston College Working Papers in Economics, 667. Boston College Department of Economics.
Beine, M., Docquier, F., Rapoport, H., 2008. Brain drain and human capital formation in developing countries: winners and losers. Econ. J. 118 (528), 631–652.
Beine, M., Docquier, F., Özden, Ç., 2011. Diasporas. J. Dev. Econ. 95 (1), 30–41.
Bems, R., Johnson, R., Yi, K.M., 2010. The role of vertical linkages in the propagation of the global downturn of 2008. IMF Econ. Rev. 58 (2), 295–326.
Benhabib, J., 1996. On the political economy of immigration. Eur. Econ. Rev. 40 (9), 1737–1743.
Bertoli, S., Fernández-Huertas Moraga, J., 2013. Multilateral resistance to migration. J. Dev. Econ. 102 (C), 79–100.
Bertoli, S., Fernandez-Huertas, J., Ortega, F., 2011. Immigration policies and the Ecuadorian exodus. World Bank Econ. Rev. 25 (1), 55–76.
Bertoli, S., Fernandez-Huertas, J., Ortega, F., 2013. Crossing the border: self-selection, earnings, and individual migration decisions. J. Dev. Econ. 101, 75–91.
Bloom, N., Draca, M., van Reenen, J., 2011. Trade induced technical change? The impact of Chinese imports on innovation and information technology. CEP Discussion Paper No. 1000.
Borjas, G.J., Freeman, R.B., Katz, L.F., 1992. On the labor market effects of immigration and trade. In: Borjas, G.J., Freeman, R.B. (Eds.), Immigration and the Workforce: Economic Consequences for the United States and Source Areas. National Bureau of Economic Research, pp. 213–244.
Borjas, G.J., Freeman, R.B., Katz, L.F., 1997. How much do immigration and trade affect labor market outcomes? Brook. Pap. Econ. Act. 1997 (1), 1–90.
Broda, C., Weinstein, D.E., 2006. Globalization and the gains from variety. Q. J. Econ. 121 (2), 541–585.
Clark, X., Hatton, T.J., Williamson, J., 2007. Explaining U.S. immigration 1971–1998. Rev. Econ. Stat. 89 (2), 359–373.
Clemens, M., 2011. Economics and emigration: trillion-dollar bills on the sidewalk? J. Econ. Perspect. 25 (3), 83–106.
Cohen, D., Soto, M., 2007. Growth and human capital: good data, good results. J. Econ. Growth 12 (1), 51–76.
Comin, D.A., Easterly, W., Gong, E., 2010. Was the wealth of nations determined in 1000 BC? Am. Econ. J. Macroecon. 2 (3), 65–97.
di Giovanni, Julian, Levchenko, Andrei A., 2009. Trade openness and volatility. The Review of Economics and Statistics, 91(3). MIT Press, pp. 558–585.
di Giovanni, J., Levchenko, A., Ortega, F., 2012. A global view of cross-border migration. CReAM Discussion Paper Series 1218.
Diamond, J., 1997. Guns, Germs and Steel: The Fates of Human Societies. WW Norton & Company, NY 14.
Do, Q., Levchenko, A., 2007. Comparative advantage, demand for external finance, and financial development. J. Financ. Econ. 86 (3), 796–834 (December).
Docquier, F., Özden, Ç., Peri, G., 2010. The Wage Effects of Immigration and Emigration. Working Paper 16646. National Bureau of Economic Research.
Dustmann, C., Frattini, T., Preston, I.P., 2013. The effect of immigration along the distribution of wages. Rev. Econ. Stud. 80 (1), 145–173.
Eaton, J., Kortum, S., 1996. Trade in ideas: patenting and productivity in the OECD. J. Int. Econ. 40 (3), 251–278.
Feenstra, R.C., Lipsey, R.E., Deng, H., Ma, A.C., Mo, H.Y., 2005. World Trade Flows: 1962–2000. Working Paper 11040. National Bureau of Economic Research.
Feyrer, J., 2009a. Trade and income—exploiting time series in geography. Working Paper 14910. National Bureau of Economic Research.
Feyrer, J., 2009b. Distance, trade, and income—the 1967 to 1975 closing of the Suez Canal as a natural experiment. National Bureau of Economic Research Working Paper 15557.
Frankel, J.A., Romer, D., 1999. Does trade cause growth? Am. Econ. Rev. 89 (3), 379–399.
Galor, O., Mountford, A., 2008. Trading population for productivity: theory and evidence. Rev. Econ. Stud. 75 (4), 1143–1179.
Garcia-Montalvo, J., Reynal-Querol, M., 2005a. Ethnic diversity and economic development. J. Dev. Econ. 76 (2), 293–323 (April).
Garcia-Montalvo, J., Reynal-Querol, M., 2005b. Ethnic polarization, potential conflict and civil war. Am. Econ. Rev. 95 (3), 796–816 (June).
Gauthier-Loiselle, M., Hunt, J., 2010. How much does immigration boost innovation? Am. Econ. J. Macroecon. 2 (2), 31–56.
Gonzalez, L., Ortega, F., 2010. How do very open economies absorb large immigration flows? Evidence from Spanish regions. Labour Econ. 18 (1), 57–70.
Grogger, J., Hanson, G.H., 2011. Income maximization and the selection and sorting of international migrants. J. Dev. Econ. 95 (1), 42–57.
Grossman, G.M., Helpman, E., 1991. Trade, knowledge spillovers, and growth. Eur. Econ. Rev. 35 (2), 517–526.
Hall, R.E., Jones, C.I., 1999. Why do some countries produce so much more output per worker than others? Q. J. Econ. 114 (1), 83–116.
Hanson, G.H., 2009. The economic consequences of the international migration of labor. Ann. Rev. Econ. 1 (1), 179–208.
Head, K., Mayer, T., Ries, J., 2010. The erosion of colonial trade linkages after independence. J. Int. Econ. 81 (1), 1–14.
Hunt, J., 2011. Which immigrants are most innovative and entrepreneurial? Distinctions by entry visa. J. Labor Econ. 29 (3), 417–457.

Jaffe, A., Trajtenberg, M., 2002. Patents, Citations and Innovations: A Window on the Knowledge Economy. MIT Press, Cambridge, MA.
Johnson, R., Noguera, G., 2012. Accounting for intermediates: production sharing and trade in value added. J. Int. Econ. 86 (2), 224–236.
Kerr, W., Lincoln, W.F., 2010. The supply side of innovation: H-1B visa reforms and U.S. ethnic invention. J. Labor Econ. 28 (3), 473–508.
La Porta, R., Lopez-de-Silanes, F., Shleifer, A., Vishny, R., 1999. The quality of government. J. Law Econ. Org. 15 (1), 222–279.
Levchenko, A., Zhang, J., 2012. Comparative advantage and the welfare impact of European integration. National Bureau of Economic Research Working Paper 18061.
Lewis, E., 2003. Local, open economies within the US: how do industries respond to immigration? Federal Reserve Bank of Philadelphia Working Paper 04-01.
Llull, J., 2011. Understanding international migration: evidence from a new dataset of bilateral stocks (1960–2000). Manuscript .
Lucas, R.E., 2009. Trade and the diffusion of the industrial revolution. Am. Econ. J. Macroecon. 1 (1), 1–25.
Manacorda, M., Manning, A., Wadsworth, J., 2012. The impact of immigration on the structure of wages: theory and evidence from Britain. J. Eur. Econ. Assoc. 10 (1), 120–151.
Matsuyama, K., 1992. A simple model of sectoral adjustment. Rev. Econ. Stud. 59 (2), 375–388.
Mayda, A.M., 2007. International Migration Flows: An Analysis of the Forces and Constraints at Work. Manuscript Georgetown University.
Mayda, A.M., 2010. International migration: a panel data analysis of the determinants of bilateral flows. J. Popul. Econ. 23 (4), 1249–1274.
Noguer, M., Siscart, M., 2005. Trade raises income: a precise and robust result. J. Int. Econ. 65 (2), 447–460.
Ortega, F., 2005. Immigration policy and skill upgrading. J. Public Econ. 89 (9), 1841–1863.
Ortega, F., 2010. Immigration, citizenship, and the size of government. B.E. J. Econ. Anal. Policy 10 (1).
Ortega, F., Peri, G., 2009. The causes and effects of international migrations: evidence from OECD countries 1980–2005. National Bureau of Economic Research Working Paper 14833.
Ortega, F., Peri, G., 2012. The role of income and immigration policies in attracting international migrants. IZA Discussion Paper 6655.
Ortega, F., Peri, G., 2013. The effect of income and immigration policies on international migration. Migr. Stud. 1, 1–28.
Ottaviano, G.I.P., Peri, G., 2006. The economic value of cultural diversity: evidence from US cities. J. Econ. Geogr. 6 (1), 9–44.
Ottaviano, G.I.P., Peri, G., 2012. Rethinking the effects of immigration on wages. J. Eur. Econ. Assoc. 10 (1), 152–197.
Pedersen, P.J., Pytlikova, M., Smith, N., 2006. Migration into OECD countries 1990–2000. In: Parson, C.A., Smeeding, T.M. (Eds.), Immigration and the Transformation of Europe. Cambridge University Press.
Peri, G., 2012. The effect of immigration on productivity: evidence from U.S. states. Rev. Econ. Stat. 94 (1), 348–358.
Peri, G., Sparber, C., 2009. Task specialization, immigration, and wages. Am. Econ. J. Appl. Econ. 1 (3), 135–169.
Putterman, L., Weil, D.N., 2010. Post-1500 population flows and the long-run determinants of economic growth and inequality. Q. J. Econ. 125 (4), 1627–1682.
Raphael, S., Ronconi, L., 2007. The effect of labor market competition with immigrants on the wages and employment of natives: what does existing research tell us?. Manuscript available at http://gsppi.berkeley.edu/faculty/sraphael/du-bois-review-january-2007.pdf.
Richardson, D.J., 1995. Income inequality and trade: how to think, what to conclude. J. Econ. Perspect. 9 (3), 33–55.
Rivera-Batiz, L., Romer, P.M., 1994. Economic integration and endogenous growth: an addendum. Q. J. Econ. 109 (1), 307–308.
Rodriguez, F., Rodrik, D., 2001. Trade policy and economic growth: a skeptic's guide to the cross-national evidence. NBER Macroeconomics Annual 2000, vol. 15, pp. 261–338.
Rodrik, D., 2000. How far will international economic integration go? J. Econ. Perspect. 14 (1), 177–186.
Rodrik, D., Subramanian, A., Trebbi, F., 2004. Institutions rule: the primacy of institutions over geography and integration in economic development. J. Econ. Growth 9, 131–165.
Sachs, J.D., Warner, A., 1995. Economic reform and the process of global integration. Brook. Pap. Econ. Act. 1995 (1), 1–118.
Silva, J.M.C.S., Tenreyro, S., 2006. The log of gravity. Rev. Econ. Stat. 88 (4), 641–658.
Stark, O., Helmenstein, C., Prskawetz, A., 1997. A brain gain with a brain drain. Econ. Lett. 55 (2), 227–234.
Stock, J.H., Yogo, M., 2005. Testing for weak instruments in linear IV regression. In: Stock, J.H., Andrews, D.W.K. (Eds.), Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg, Chapter 5. Cambridge University Press.
Taylor, A.M., 1997a. Peopling the pampa: on the impact of mass migration to the River Plate, 1870–1914. Explor. Econ. Hist. 34 (1), 100–132.
Taylor, A.M., 1997b. Growth and convergence in the Asia-Pacific region: on the role of openness, trade and migration. In: Lloyd, P.J., Williams, L.S. (Eds.), International Trade and Migration in the APEC Region. Oxford University Press, Oxford.
Taylor, A.M., Williamson, J.G., 1997. Convergence in the age of mass migration. Eur. Rev. Econ. Hist. 1 (1), 27–63.
Weil, D.N., 2005. Economic Growth. Pearson Addison, Wesley, New York.
Weil, D.N., 2007. Accounting for the effect of health on economic growth. Q. J. Econ. 122 (3), 1265–1306.