Globalization and Gravity

By PANKAJ GHEMAWAT AND TAMARA DE LA MATA*

Gravity modeling in international economics focuses on trade in goods. But globalization is a multidimensional phenomenon that also includes international movements of capital, people, and information. Simple gravity models that consider the sizes of economies and geographic distance work well across a broad range of international interactions. So do the additional types of distance customarily included in augmented gravity models of trade. Countries that are relatively distant from each with respect to a particular type of flow also tend to be distant as far as other interactions are concerned. And geographic distance, in particular, is a fair proxy for augmented distance. (JEL: F01, F15).

* Ghemawat: IESE Business School, Universidad de Navarra. Arnús i Gari, 3-7, 08034, Barcelona, Spain and Stern School of Business, New York University. 44 West Fourth Street, New York, NY 10012, USA (e-mail: PGhemawat@iese.edu); de la Mata: IESE Business School, Universidad de Navarra. Arnús i Gari, 3-7, 08034, Barcelona, Spain and Universidad Autónoma de Madrid, Campus de Cantoblanco, 28049, Madrid, Spain (e-mail: tamara.delamata@uam.es). We are grateful to Steven Altman for helping oversee the compilation of the DHL Global Connectedness Index and to Pol Antras and seminar audiences at NYU, Rutgers, and the University of Toronto for helpful comments on an earlier draft. IESE Business School and the NYU Stern School of Business supported this research financially.

There are easily hundreds, if not thousands, of studies that use gravity models to study international interactions. The overwhelming majority, however, focuses on trade flows, particularly trade in goods. Consider, for instance, Head and Mayer's (2014) chapter on gravity equations in the *Handbook of International Economics:* the subsection on "Gravity models beyond trade in goods" occupies about one page out of a total of 64.

However, globalization is generally defined as being broader than trade in just products or even services: definitions typically also refer to international movements of capital, people, and information. This paper examines how well gravity modeling works in explaining patterns of globalization. Specifically, it uses a detailed dyadic dataset assembled for the DHL Global Connectedness Report (Ghemawat and Altman 2012) to compare how sensitive 11 different types of international interactions are to distances of various sorts.

We organize the remainder of the paper as follows. In Section I, we discuss the basic specification of the gravity model employed in this paper, describe the variables analyzed here, and briefly review prior work on them. In Section II, we present our baseline estimates. Section III examines the extent to which countries that are distant from each other with respect to a particular type of interaction also tend to be distant as far as other interactions are concerned. Section IV reexamines the robustness of the baseline analyses, which rely on OLS (ordinary least squares) estimation of log linear models, by treating more carefully observations for which the dependent variable is zero. Section V summarizes the results of the analysis and discusses some extensions to it.

I. The gravity model

The gravity model has been used widely to analyze the factors that influence the magnitude of the bilateral trade between countries or regions. According to the basic gravity model, the magnitude of these interactions increases with the size of both economies and decreases with the geographical distance that separates them. Discussions of the theoretical foundations of gravity models of trade flows include Anderson (1979), Bergstrandt (1985, 1989), Anderson and van Wincoop (2003), Chaney (2008), Helpman, Melitz, and Rubinstein (2008), and Olivero and Yotov (2012).

The first articulation of gravity models in economics is generally attributed to Carey, who writes, “gravitation is here, as everywhere else in the material world, in the direct ratio of the mass, and in the inverse one of the distance” (1858, 42–43). However, Carey offers no empirical evidence in support of his conjecture, and seems to have been unaware of the prior work of Desart (1846), who identifies a log linear relationship between passenger flows and distances on Belgian railways.[1] Another early application is Ravenstein’s (1885) work on flows of migrants to Britain. However, the big surge in applications of gravity models to spatial interactions, including international ones, did not take place until after World War II.

The gravity equation expressing bilateral trade as a function of distance is generally credited to Tinbergen (1962), who also notes the dependence of the magnitude of trade flows on political and socio-economic factors such as the existence of special trade agreements (e.g., the Benelux Union) and colonial ties (e.g., the British Commonwealth). Fifty years later, Head and Mayer’s (2014) chapter in the *Handbook of International Economics* identifies the most frequently used independent variables in gravity models of trade in goods and presents a meta-analysis that draws on 159 papers and more than 2,500 estimates from them that were published in leading journals between 2006 and 2012. Table 1 contains a summary.

[1] For further discussion, see Odlyzko (2014).

[Insert Table 1 Here]

We mostly follow Head and Mayer in terms of the independent variables employed and focus on extending the range of dependent variables examined, as we describe in more detail in the next section. The only independent variable in Table 1 that we drop is the one indicating whether two countries share a common currency. There are two reasons for this. First, it seems more pertinent to trade flows than to some of the other types of interactions we study in this paper. Second, Head and Mayer (2014) recognize that the results obtained in the literature for this variable have been the subject of some controversy, and are particularly sensitive to the methodology and sample used.

Instead of currency differences versus commonalities, we include a final independent variable related to economic distance: disparities in per capita (PC) income. We define this variable as the ratio between the maximum and the minimum average PC income in US$ using market exchange rates for each country-pair. Our interest in looking at this variable derives from the broad shift of economic activity that is taking place from advanced to emerging economies, reducing the dominance of international interactions between advanced economies (Spence 2011).

Although many different specifications have been employed in modeling the effects of these independent variables, recent studies tend to take on board Anderson and van Wincoop's observation (2003) that bilateral trade flows depend not only on bilateral trade barriers but also on trade barriers across all trading partners. In analyzing trade in goods—and some of the other interactions that we examine—it has become standard to control for these multilateral resistance terms by using country-specific fixed-effects. As Head and Mayer (2014) note in the context of trade in goods, including fixed effects does not involve strong structural assumptions about the underlying model, but it does mean that the determinants of overall export or import propensities can no longer be identified through gravity modeling. In other words, with fixed effects, overall export or import propensities are taken as given and it is their distribution across partners that is analyzed.

In keeping with what has been described as the multilateral resistance/fixed effects revolution in gravity modeling, we assume that there are country-specific unobservable characteristics (φ_o and φ_d for origin and destination countries, respectively) that must be taken into account to avoid biasing coefficient estimates. We also include a set of year-specific dummy variables (φ_t) to control for the specific characteristics of each year (which is important because our data include years before and after the onset of the global financial crisis in 2008). Finally, u_{odt} is a random error term that represents other, omitted influences on bilateral flows. Equation 1 summarizes the basic log linear gravity equation in which the intensity of the connection between country o and country d in year t (F_{odt}) is a function of a set of variables:

(1) $$\log(F_{odt}) = \beta_0 + \beta_1 Language_{od} + \beta_2 Colony_{od} + \beta_3 Agreement_{odt} + \beta_4 \log(Geographic\ distance)_{od} + \beta_5 Contiguity_{od} + \beta_6 \log(Income\ disparities)_{odt} + \beta_7 \log(Size)_{odt} + \varphi_o + \varphi_d + \varphi_t + u_{odt}$$

In order to estimate such gravity equations, we cluster errors by country pair to avoid potential heteroscedasticity problems and to relax the assumption of independence across observations for the same country-pair over the period we study. Furthermore, it is important to add that we analyze different types of interactions independently of each other. Other studies look at patterns of substitution/complementarity across (subsets) of the interactions that we examine; however, systematic analysis of this sort is beyond the scope of the present paper, although it does highlight some of the difficulties with past analyses of this sort.

In order to examine whether gravity applies generally to globalization, and not just to trade in goods, we analyze 11 types of international interactions that can be grouped into four categories: trade, capital, people, and information for the period 2005–2012. We generally rely on official statistical sources so that others can easily replicate our analyses.

A. Trade

We analyze both goods and services exports. We use the value of exports from the IMF Direction of Trade Statistics (DOTS) and the UN Comtrade database for trade in goods and the value of services exports from the UN Services Trade database. Services receive much less attention in the gravity modeling literature for reasons ranging from data limitations[2] to their relatively limited share of international trade (officially reported to be about 20 percent), which results in their often being treated as non-tradable. Exceptions include Kimura and Lee (2006) and Kandilov and Grennes (2012). Kimura and Lee (2006) use gravity modeling to analyze the impact of distance and other factors on bilateral services trade for OECD countries and conclude that the gravity model fits trade in services even better than it does trade in goods. And Kandilov and Grennes (2012) analyze the determinants of service offshoring and find that after controlling for trade costs unrelated to transportation (e.g., information and cultural barriers), the influence of geographic distance, although weakened by the controls, remains significant.

[2] The issuance of the first edition of the Manual on Statistics of International Trade in Services in 2002 marked a major improvement in this regard. It is available at: http://unstats.un.org/unsd/publication/Seriesm/Seriesm_86e.pdf.

B. Capital

In regard to capital interactions, we analyze outward FDI stocks (from the OECD database and national sources) and portfolio equity stocks and portfolio long-term debt, both from the IMF Coordinated Portfolio Investment Survey (CPIS).

There is some theoretical ambiguity about how geographic distance is likely to affect these and other capital interactions. Thus, Helpman, Melitz, and Yeaple (2004) develop a model of international trade and investment in which firms can choose to serve foreign countries by exporting or engaging in FDI. They predict that foreign markets will be served by exports instead of by FDI when trade frictions are lower. Thus, with horizontal FDI, one might expect to see a positive effect of distance as firms invest in remote markets that cannot cost-effectively be served through exports. Regarding portfolio equity, if investors prefer more heterogeneous portfolios to diversify risk, they might want to invest in more distant countries since that is likely to reduce the synchronicity of business cycles (Baxter and Kouparitsas 2005). But these effects may be outweighed by others—e.g., due to the differences between horizontal and vertical FDI (Egger 2008) or unfamiliarity that implies fixed costs that increase with distance—that point in the opposite direction. Additionally, it is also possible that the weightlessness of capital flows might simply attenuate (geographic) distance effects.

Empirical analyses of bilateral FDI actually find a large negative coefficient for geographic distance (Carr, Markusen, and Maskus 2001; Markusen and Maskus 2002; Blonigen, Davies, and Head 2003). Blonigen et al. (2007), using data on US outbound FDI activity, find that this negative effect survives the inclusion of terms to capture the spatial autocorrelation of FDI flows to alternative host destinations. And Kleinert and Toubal (2010), working with a dataset on sales of foreign affiliates of multinational firms, find a negative distance effect that offers more support for gravity equations derived from models of horizontal instead of vertical FDI. In regard to portfolio equity investments, Portes and Rey (2005) find a negative relationship that they associate with informational frictions. Daude and Fratzscher (2008) look across FDI stocks, portfolio investment (debt and equity), and bank loans and find that geographic distance has a larger negative effect on FDI stocks and loans than on portfolio equity and debt.

C. People

We analyze three types of people-related interactions: international migration based on bilateral migration stock data from the UN Population Division, incoming tertiary students from the UNESCO Institute for Statistics, and international tourist arrivals from UN World Tourism Organization. The general expectation about such movements of people is that geographic distance should dampen them.

Of the people-related interactions that we look at, migration has attracted the most attention, (e.g., Lewer and Van den Berg 2008; Mayda 2010; Grogger and Hanson 2011; Ortega and Peri 2013) and all these studies find a strong negative effect for geographic distance. Studies that focus on medium-term student flows are scarcer and generally concentrate on particular regions or countries e.g., the González, Mesanza, and Mariel (2011) analysis of the European Union’s Erasmus program and Bessey’s analysis of university students coming to Germany (2012), but do point in the same direction. Short-term flows of tourists have been studied by, among others, Eilat and Einav (2004) and Keum (2010), with the latter focusing on Korea. Both these studies find, once again, that geographic distance has a strong negative effect.

D. Information

Under this category, we analyze trade in printed publications using exports reported under code 49 of the Harmonized System from UN Comtrade; the number of minutes of outgoing international phone calls from Telegeography (the only proprietary data we rely on); and the number of patents registered by foreign inventors in a country's patent office from the World Intellectual Property Organization (WIPO) database.

Comprehensive analysis of informational interactions appears to be rarer than for the other categories we consider above. Related to our first informational variable, Meng (2011) estimates gravity models for China's international trade in various cultural goods and notes a generally negative impact of geographic distance, although distance-sensitivity for printed materials is estimated to be smaller than for other types of goods. For phone calls, Rietveld and Janssen (1990) examine calling patterns between the Netherlands and 27 other countries and find that the intensity of calls with East European countries is a fraction of that with other West European countries, and that international phone calls tend to follow patterns similar to international trade. Wong (2008), who uses Telegeography data, as we do, reaches similar conclusions about the parallels between trade flows and telephone calls. Finally, cross-border knowledge flows have been analyzed from more than one perspective. Keller (2002) studies the extent to which knowledge spillovers are global or local and concludes that the geographic distance between countries reduces the magnitude of productivity gains from each other's R&D spending. And Archontakis and Varsakelis (2011) use the same patent data as we do but focus on the international patenting activity of 27 OECD countries. They conclude that international patenting appears to be reduced by geographic distance, although the estimated effect is smaller than for most of the other international interactions we discuss and its statistical significance, and even its sign, fluctuate across country subsamples.

E. Independent Variables

As indicated in equation 1, we control for size by including the product (logged) of the GDPs or of the populations of both countries—the former for trade and capital and the latter for people and informational interactions. The other independent variables we employ, which are related to distance of various sorts, merit more discussion, as do the effects that they have previously been estimated to have on the range of dependent variables that we analyze. Given the number of dependent variables and the prior literature's concentration on one of them—trade in goods—we focus on establishing a baseline for further analysis by looking at the extent to which the distance variables commonly used in gravity models to explain trade in goods also apply to the other kinds of international interactions examined here.

On the geographic dimension, we include the two independent variables most commonly employed in the prior literature: geographical distance and sharing a border. For the former, we use the weighted distance variable from the CEPII dataset, which relies on city-level data to assess the geographic distribution of population (in 2004) in each country. The distance between two countries is measured in terms of bilateral distances between their biggest cities, with weights proportional to those cities' shares of overall country population.[3] We discussed the effect of geographic distance on each type of interaction in the previous subsection.

The positive effect of a common border has been tested and confirmed in the literature of trade in goods but results concerning other types of international

[3] Head and Mayer (2002) developed the formula to obtain this distance variable.

interactions are more equivocal. For services trade Kimura and Lee (2006) and Braymen and Briggs (2015) find limited or insignificant effects. Kleinert and Toubal (2010) confirm the positive and significant effect of sharing a border on FDI and find that adding this variable almost halves the estimated effect of distance on real affiliate sales. However, for portfolio equity, Portes and Rey (2005) recognize that in their dataset, adjacency is very collinear with regional bloc dummies and does not improve explanatory power. For people flows, Lewer and Van den Berg (2008) find a positive effect of contiguity that disappears when other factors such as the stock of past migrants are included. For OECD countries, Belot and Ederveen (2012) conclude that migration flows between neighboring countries are more than twice as large as flows between countries without a common border, but the effect of this variable is non-significant when the analysis is restricted to members of the European Union or European Economic Area. Eilat and Einav (2004) confirm the importance of sharing a border for tourism only for low income destination countries. This variable is typically not analyzed for information interactions. Wong (2008) and Picci (2010) are exceptions. The former finds a positive effect for phone calls and the latter for international patent applications.

Along the cultural dimension, we include another commonly used variable that Head and Mayer (2014) highlight: a dummy variable indicating whether two countries share the same official language. A meta-analysis of 81 academic articles by Egger and Lassmann (2012) confirms a strong positive effect of linguistic commonality on trade in goods. Kimura and Lee (2006) find a similar positive effect on trade in services. Oh, Selmier, and Lien (2011) find that linguistic commonality has a larger positive effect on FDI than on trade flows. For portfolio equity, Daude and Fratzscher (2008), Lane and Milesi-Ferretti (2008), and Aggarwal, Kearney and Lucey (2012), find a positive, significant effect, although linguistic commonality fails to achieve significance in Aviat and

Coeurdacier's (2007) analysis. For people flows, Belot and Ederveen (2012) show that cultural barriers, including linguistic commonality versus difference, explain the pattern of migration flows between OECD countries better than income and unemployment differentials and Eilat and Einav (2004) confirm the importance of sharing the same official language for international tourism. And for information flows, Wong (2008) finds a positive impact of linguistic commonality on telephone calls and Picci (2010) on bilateral collaboration in the context of international patent applications.

Along the administrative dimension, we include two variables that Head and Mayer (2014) flag: a dummy variable identifying whether two countries have ever had a colonial linkage as a measure of historical proximity, and a policy variable related to whether two countries have a free trade agreement or are both party to specific multilateral blocs (CUSA/NAFTA/EU). We modify the latter variable depending on the type of interaction being examined: a dummy variable that captures the effect of any trade agreement in force for goods and services exports; and a dummy variable that indicates common membership in certain regional blocs (Mercosur, ASEAN, CARICOM, EU, NAFTA, and GCC) for the other dependent variables. We extract both variables from the CEPII dataset and update them by using data from the World Trade Organization.

There is a large literature confirming the positive, strongly significant effects of colonial linkages on trade in goods. Thus, Head, Mayer, and Ries (2010) find that independence did substantially diminish post-colonial trade in goods—with particularly large, immediate reductions in the event of hostile separations—but that historical linkages of this sort continue to exert a significant, positive effect even decades after independence. Kandilov and Grennes (2012) argue that colonial linkages should matter less for trade in services than for trade in goods but nonetheless find a positive and significant impact in almost all the specifications they estimate. For capital interactions, Lane and Milesi-Ferretti

(2008) and Oh, Selmier, and Lien (2011) detect significant positive impacts on bilateral FDI and portfolio equity stocks respectively. For people flows, previous research is more ambiguous. Grogger and Hanson (2011) and Ortega and Peri (2013) find a strong, positive impact on immigrant stocks, but others estimate a significantly weaker effect (e.g., Kim and Cohen 2010) or even one that fades off into insignificance (Mayda 2010). Beine, Noël, and Ragot (2014) uncover a positive effect on university students studying abroad, and highlight the possibility that this effect might be due to the role played by networks of previous students that studied in the colonizing country. For information flows, the available empirical evidence is limited. Disdier el al. (2010) find a positive effect of colonial linkages on cultural goods, but this is linked more to the visual arts and cultural heritage goods than to printed publications. And Wong (2008) finds that the ties between the United Kingdom, France, and Spain and their former colonies significantly boost international phone calls.

Turning to trade agreements/regional blocs, the effects on trade in goods have generally been estimated to be positive and significant. Marchetti (2011) demonstrates that preferential trade agreements also have a positive effect on bilateral services trade. The expected effects on FDI are ambiguous and depend on the type of FDI considered. From a knowledge-capital model, which integrates both horizontal and vertical FDI, Bergstrand and Egger (2007) predict a negative effect of regional trade agreements on FDI, which they confirm empirically using UNCTAD data for 17 countries. In the same vein, Guerin (2006) finds a negative and significant effect for North-to-South FDI flows whereas Levy-Yeyati, Stein, and Daude (2003) study FDI flows in the Americas and find a significant positive effect. Similarly, Aristotelous and Fountas (1996), who study the determinants of US and Japanese FDI in the EU, find evidence on the positive effect of the Single European Act on FDI. For people and information flows, the available evidence is more limited, but Orefice, Lima, and Figueiredo (2014) find a positive relation between shared membership of a regional trade agreement and migration flows.

The final measure of distance that we study, the ratio of per capita (PC) income, was not included in the Head and Mayer (2014) list of canonical variables and has generally been subject to less gravity modeling. For trade flows the Linder hypothesis (Linder 1961) suggests that countries with similar income per capita should trade more intensively with one another, e.g., that our measure of differences in PC income should have a negative effect on trade. However, previous analyses of this in the context of trade in goods have yielded ambiguous results. Braymen and Briggs (2015) seems to be the only work that analyzes the effect of income differences on services trade. Their results confirm that dissimilar per capita income levels influence services trade positively. For FDI, the expected effect of income per capita are ambiguous and depend on the FDI model considered. Empirically, Fajgelbaum, Grossman, and Helpman (2015) find that horizontal FDI is more likely to occur between countries with similar per capita income levels—the so-called Lucas paradox (Lucas 1990).

This brief review of an extensive literature suggests that the international interactions that we study in this paper—beyond trade in goods—have all previously been subject to gravity modeling. The motivation for this paper is that prior work has a patchwork character, with varying specifications, coverage (in terms of time frame and countries considered), and estimation methods. Given our interest in exploring commonalities versus differences well across the range of interactions examined and, in particular, examining whether countries that are relatively distant from each other with respect to a particular type of interaction also tend to be distant as far as other interactions are concerned, greater consistency is required—which is what the analyses in the rest of this paper aim to provide. In the next two sections, we drop all zero flows and rely on OLS estimation of equation 1; a more sophisticated treatment of zero values is provided in Section IV.

II. OLS estimation of gravity models

It is useful to begin the presentation of the results of the estimation exercise by noting that the coefficient estimates that we obtain for goods exports seem generally consistent—for the variables that overlap—with the median coefficients estimated for structural gravity equations that Head and Mayer (2014) report (see Table 1). Having noted as much, we turn to a broader discussion of the results.

Table 2 shows the results of OLS estimates of equation 1 for all 11 types of international interactions described in the previous section. Despite the fact that we are using (almost) exactly the same explanatory variables across the interactions, the goodness of fit of the models ranges from 70 percent to nearly 90 percent (with goods exports, to which the explanatory variables were "tuned," situated at the median).

[Insert Table 2 Here]

Looking at the individual explanatory variables, the coefficients on common official language are positive and significant almost across the board—except in the case of portfolio long-term debt—and generally larger for people and informational interactions. The coefficients on the colony-colonizer link perform similarly. The coefficients on trade agreements/regional blocs, in contrast, perform much more poorly, often failing to have the predicted sign let alone achieving statistical significance. Of the geographic variables, the coefficients on geographic distance are negative and significant throughout and, with the exception of the one on patenting activity, range from close to -1 to nearly -2 on printed publication exports. Sharing a border works less well: it achieves significance only six times out of 11, with its strongest effects felt on migrants and tourists. And finally, the ratio of PC income is significant in 10 of 11 cases: the only interaction on which it doesn't have a significant effect is trade in goods (in keeping with the ambiguous results reported in past research). Differences in PC income generally depress international interactions, except in the case of service exports, migrants, and phone calls. Note that the latter two types of interactions are particularly closely related: thus, the top two destinations of international calls placed from the United States are Mexico (the largest source of first-generation immigrants) and India (the third largest).

While the overall goodness of fit we report in Table 2 is impressive, much of this is due to the fixed effects and the size variables. In order to gain a better sense of how much the distance variables add to explanatory power, we compare the adjusted R^2 of a similar set of regressions that include just the country-specific fixed effects and the size variables (column 1 in Table 3) with regressions that also include geographical distance (column 2 in Table 3),[4] as well as with the standard specification that includes all the distance variables analyzed in Table 1 (column 3 in Table 3).

The first thing to note from Table 3 is that the regressions that include just the fixed effects and the sizes of the country-pairs (column 1) are able to explain a large share of the variance, ranging from nearly one-half to more than four-fifths. This serves as a reminder of the problems with focusing on overall goodness of fit as an indicator of the importance of the distance variables. Second, adding in geographic distance helps explain about one-quarter of the remaining variance on average, although the precise shares range from 6 percent to 49 percent. Third, the addition of the remaining distance variables generally boosts overall explanatory

[4] The details of these regressions are available upon request.

power significantly less: in the case of trade in goods, in particular, the share of remaining variance explained goes up from only 29 percent to 31 percent.

[Insert Table 3 here]

These data might seem to suggest that at least in the case of trade in goods, augmenting geographic distance with the other distance variables we include in equation 1—as is common practice, according to Head and Mayer (2014)—makes little sense. But it is important to remember that several of the other distance variables are significantly correlated with geographic distance. Therefore, attributing all the additional variance accounted for by the specification in column 2 versus column 1 of Table 3 to geographic distance likely overstates its real contribution. A safer conclusion is that augmenting geographic distance with other distance variables seems to make (even) more sense for other international interactions than it does for trade in goods (where such augmentation is common).

III. Rankings and augmented distances across interactions

In addition to considering the distance variables equation by equation, it is possible to look at their joint implications across the 11 different equations. One approach is to look at the extent to which if two countries are predicted to be close to each other with respect to one type of international interaction, they are also predicted to be close to each other with respect to the other types of interactions studied.

Simple examination of the US's top five partners for each type of interaction—see Table 4—suggests that one should indeed expect such concordance. Canada, the 11th largest economy in the world,[5] shows up among the top five partners for

[5] Based on IMF World Economic Outlook Database, October 2014 using 2012 GDP in US$ (current prices).

all 11 interactions, Great Britain, the 6th largest, for nine of them, and Mexico, the 15th largest, for six of the 11. In contrast, China, the second largest economy in the world, figures among the top five partners just three times. It is worth adding that in ten out of 11 cases, the top five partners account for more than half of the US's international interactions: service exports are the exception, with 49 percent directed at the top five. And the United States is not idiosyncratic in the sense of having particularly narrow interactions with others: it actually ranks second out of 140 countries—after the United Kingdom—on the DHL Global Connectedness Index for 2012 in terms of the global breadth of its interactions.[6]

[Insert Table 4 Here]

To check whether such patterns apply across all countries, we compute for each type of interaction, *f*, the deviation (difference) between the shares that each partner represents in the total interactions of that type for each country and the share that each partner accounts for in total international interactions of that type (equation 2). This yields a measure of the importance of the linkages between two countries relative to the share of the partner in the flows of all other countries.

$$\text{(2)} \qquad \text{Deviation}^{f}_{od} = (F^{f}_{od}/\textstyle\sum_{d}F^{f}_{od})-(\sum_{o}F^{f}_{od}/\sum_{o,d}F^{f}_{od})$$

Table 5 shows the pairwise correlations of these deviation vectors. They turn out to be positive in all cases as well as statistically significant. More than three quarters of the correlations exceed 0.5 and the median correlation is 0.64. The five that exceed 0.8 are highlighted in bold in the table. The highest correlation is between tourist arrivals and services exports, and reflects, in part, an accounting relationship: inbound tourism is a component of service exports in the balance of

[6] This breadth ranking is based on 9 of the 11 interactions that we are studying here—portfolio long-term debt and patent citations are excluded from the DHL Global Connectedness Index for 2012.

payments. Service exports are also highly correlated with portfolio long-term debt, and with goods exports as well (the sixth highest correlation in the table). Other particularly strong correlations are between immigrants, tourists, and phone calls. University students, however, exhibit low correlations with those three variables and account, in fact, for ten of the 11 lowest correlations in the table.

[Insert Table 5 Here]

Note that such correlations complicate attempts to sort out causality. Consider, for instance, Portes and Rey's (2005) finding that cross-border portfolio equity flows are highly correlated with cross-border telephone calls. They aver that since equity trading is unlikely to boost telephone calls, the causality is likely to run the other way around: "the geography of information is the main determinant of the pattern of international transactions" (269). What this ignores of course is the possibility that common underlying factors—the distance-related variables—might underlie both these types of international interactions as well as, based on the evidence we present, the other variables studied in this paper.

Next, we look at the extent to which the coefficients on different types of distances we estimate in Section II imply that countries that are relatively close or distant with respect to one type of interaction are similarly situated with respect to other types of interactions. Such similarity would render distance-based modeling more attractive than if the coefficients imply very different patterns across the 11 interactions (in which case, relative distance would be a contingent notion rather than a broad one). We compute an "augmented distance" measure that considers the impact and actual values of all distance-related variables in the standard gravity equation for each type of interaction (based on the estimates in Table 2). We obtain a vector with 97*(96-1) = 9312 elements containing, for each country pair, an augmented distance measure implied by the coefficients estimates of each type of flow, that we calculate as follows:[7]

(3) $ADistance_{od}=1/(MLanguage_{od}*MColony_{od}*MAgreement_{od}*MGeographic\ distance_{od}*MContiguity_{od}*MIncome\ disparities_{od})$

(4) $Augmented\ Distance_{od} = ADistance_{od} * \sum_{od} Geographic\ distance_{od} / \sum_{od} ADistance_{od}$

where $MLanguage_{od}=\exp(\alpha_1 Language_{od})$; $MColony_{od}=\exp(\alpha_2 Colony_{od})$; $MAgreement_{od}=\exp(\alpha_3 Agreement_{od})$[8]; $MGeographic\ distance_{od}=\exp(\alpha_4 \log(Geographic\ distance_{od}))$; $MContiguity_{od}=\exp(\alpha_5 Contiguity_{od})$ and $MIncome\ disparities_{od}=\exp(\alpha_6 \log(\max(pc\ gcp_{ot}, pc\ gcp_{dt})/\min(pc\ gcp_{ot}, pc\ gcp_{dt}))$.

Equation (3) describes an inverse augmented distance obtained as the inverse of the product of the impact multiplier considering the coefficients estimates from the gravity equation and the similarities or differences of each country pair. Then, we rescale the augmented distance, multiplying the inverse of the product of all the multipliers ($ADistance_{od}$) by the second term in equation (4), $\sum_{od} Geographic\ distance_{od} / \sum_{od} ADistance_{od}$, making the augmented distance a measure comparable to the geographical distance.

Figure 1a shows the correlation between augmented distances implied by coefficients for goods trade estimates and distances implied by estimates for other types of interactions (Table B1 in the appendix contains the complete matrix of correlations). Again, the correlations are high: they exceed 0.9 for five of the ten

[7] The augmented distance measure we describe in this section is based on the values of the independent variables as in 2012. Note that the only independent variables that vary with time are the regional bloc or trade agreement variable and the ratio of PC income.

[8] We calculate this using the Regional Bloc or the Trade Agreement variable, depending on the variable considered in the original regressions. For goods and services exports, the Trade Agreement dummy variable has been employed, while for the rest of the flows, the Regional Bloc dummy has been used.

other types of interactions and are particularly elevated for people and informational interactions, except for patenting activity. We find correlations of about 0.8 for the latter and for services exports, FDI outward stock, and portfolio equity assets; for portfolio long-term debt, the correlation level falls to its lowest level, 0.6. All the correlations are significant at the 1 percent level. Finally, Figure 1b shows the correlation between simple geographical distance and the augmented distances implied by coefficients for all types of flows, yielding very similar results. Geographic distance is generally a good proxy for augmented distance, especially for “weighty” interactions. It works a bit less well for weightless interactions (with the exception of phone calls) although most of the correlations remain quite high. According to Table 3, patenting activity is the interaction with the lowest increase in the goodness of fit when we include geographical distance. This is consistent with the correlation shown for patenting activity below, since most innovation activity takes place in large, wealthy countries, so fixed effects and the size variable capture a large share of the total variation in this regard.

[Insert Figure 1a Here]

[Insert Figure 1b Here]

IV. Robustness to zero values

Until now, we have simply discarded zero values in using OLS to estimate log linear models. While customary, this is empirically not very satisfactory: values of zero account for 30 percent or more of all flows—up to a high of 77 percent for patent citations—except for trade in goods, which comes in at 12 percent and phone calls, where we are working from Telegeography’s subsample of countries, which reports no zero values. Censoring such observations implies that small values for residual terms in such regressions are more likely to be censored and also induces a correlation between the residuals and independent variables within the subset of non-censored observations. Therefore, the expected value of the residual will not be zero and OLS will yield biased estimates.

Santos Silva and Tenreyro (2006) further discuss the econometric limitations of lognormal transformation—even when the incidence of zero values is limited, it yields biased and inconsistent estimations in the presence of heteroscedasticity—and propose a modified Poisson model of interactions. Then, the multiplicative gravity relationship that forms the baseline for the present paper can be written as an exponential function, as in equation 1, which summarizes the basic gravity equation in which the intensity of the connection between country o and country d in year t (F_{odt}) is a function of a set of variables:

(5) $F_{odt} = \exp(\beta_0 + \beta_1 \text{Language}_{od} + \beta_2 \text{Colony}_{od} + \beta_3 \text{Agreement}_{odt} + \beta_4 \ln(\text{Geographic distance})_{od} + \beta_5 \text{Contiguity}_{od} + \beta_6 \ln(\text{Income disparities})_{odt} + \beta_7 \ln(\text{Size})_{odt} + \varphi_o + \varphi_d + \varphi_t)\, \epsilon_{odt}$

where $\epsilon_{odt} = \exp((1\text{-}\sigma)\, u_{odt})$.

In words, there is a positive probability of zero interactions between a particular country-pair that shifts in line with structural parameters, but with the (conditional) distribution of strictly positive values remaining invariant (and taking on a particular functional form).

Table 6 summarizes the results of estimating standard gravity models using the Poisson Pseudo-Maximum-Likelihood (PPML) procedure that Santos Silva and Tenreyro (2006) propose and implement.

[Insert Table 6 Here]

Turning to the coefficient estimates themselves, perhaps the biggest headline is that distance effects continue broadly to be significant, but are a bit weaker than under OLS. Looking in the aggregate at the six distance-related variables across the 11 types of international interactions, these are now significant more than two-thirds of the time (71 percent) at least at the 10 percent level, versus four-fifths of the time under OLS. The net difference is accounted for equally by colonial linkages, which are now significant across seven rather than ten out of 11 interactions, and per capita income disparities, which exhibit the same overall pattern, but also include two signs that flip and achieve significance. This suggests some particular uncertainty around predictions about this "new" variable that does not appear in Head and Mayer's (2014) canonical set. Additionally, in the overwhelming majority of the instances in which the coefficients do retain their sign and statistical significance, the magnitudes of the coefficient estimates are smaller more often than not. In particular, estimated sensitivity to geographic distance decreases across the board, often significantly, although it always retains a significant negative sign. Reduced distance-sensitivity is also evident in Santos Silva and Tenreyro's (2006) PPML versus OLS estimates.

Looking interaction by interaction, the most changes in coefficient significance, and even sign—on four of the six distance-related variables—occur with patenting (for which values are zero half of the time) and PC income (changing the sign from negative to positive). The second largest number of such changes, three each, are accounted for by service exports and portfolio equity assets, which trailed only patenting and tourism in terms of the percentage of zero values. One or two changes per interaction is more common. And at the other extreme, outbound FDI and outbound phone calls experience no such change whatsoever, although they do see estimated distance-sensitivities (which remain statistically significant) decreasing substantially in absolute terms for three of six distance-related variables.

The PPML estimates we present in Table 6 reproduce the analysis that follows Table 3's presentation of the OLS estimates. The percentage points of remaining variance explained by expanding from geographic to augmented distance goes up more with PPML (ten times out of 11, with the exception of outward FDI) than with OLS, and the average increment is nearly twice as large, implying that the case for expanding beyond geographical to augmented distance is even clearer. (The counterpart of Table 3 based on PPML estimates appears in Table B2 in the Appendix.) And Figure 2, a PPML counterpart to OLS-based Figure 1b, frames the same point differently: the usability of geographic distance as a proxy for augmented distance does decrease significantly with the switch from OLS to PPML: a bit for patenting, and rather more for outbound FDI and portfolio long-term debt (pushing them to lower levels than any seen in Figure 1b). That said, geographic distance retains a large degree of influence; only for portfolio long-term debt (which was already associated with the weakest effect in Table B2) does the correlation with geographic distance drop below 0.5.

[Insert Figure 2 Here]

V. Conclusions and discussion

Gravity modeling in international economics tends to focus on trade, particularly trade in goods. How well does gravity work when we look at a broad range—11, to be precise—of cross-border interactions?

This paper provides some basic descriptive analysis that aims to address that question. It finds, first of all, that simple gravity models that focus on the sizes of economies and the geographic distance between them work fairly well across the range of interactions examined—i.e., gravity applies generally to globalization, not just to trade. Geographic distance, in particular, seems to dampen—at different rates—all types of interactions, including weightless ones. Second, the additional distance variables customarily included in augmented gravity models of trade boost explanatory power appreciably in the context of other types of interactions—i.e., apply across the board. Since the boost is smallest for trade in goods, the case for augmenting geographic distance with other distance-related variables can be made at least as strongly for other types of interactions. Third, countries that are relatively distant from each other with respect to a particular type of interaction also tend to be distant as far as other types are concerned, in terms of actual interactions and based on the coefficients from the regression analysis. These results, derived from OLS estimation, weaken a bit but remain broadly valid when zero values are dealt with in a more sophisticated way.

Of course, this paper makes but a start at the systematic analysis of multiple types of international interactions. It stops short of undertaking other obvious analyses: looking at additional explanatory variables and customizing them by interaction, including more types of interactions, engaging in network as well as dyadic analysis, trying to sort out causality despite the typically high correlations between different types of interactions, and so on.

VI. Figures

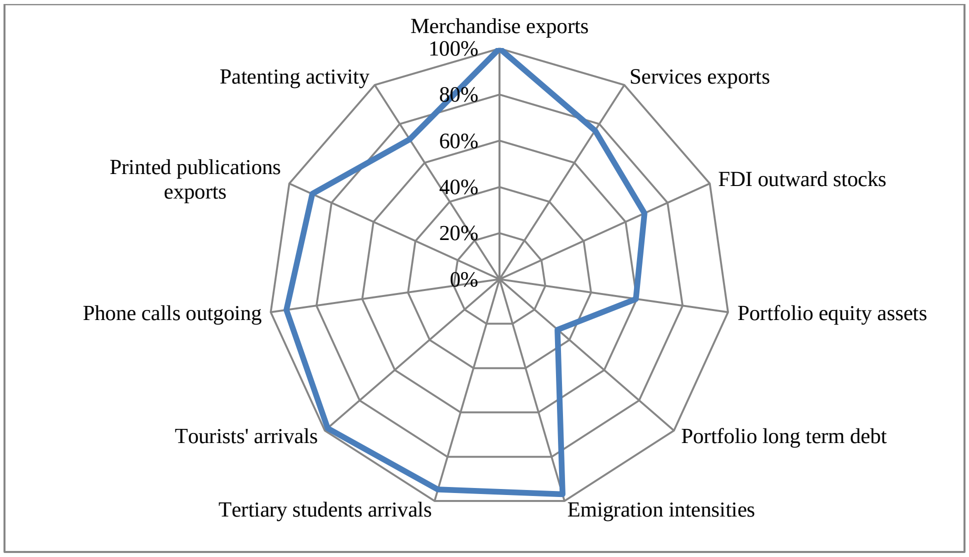

FIGURE 1A. CORRELATION BETWEEN DISTANCES IMPLIED BY COEFFICIENTS FOR TRADE IN GOODS ESTIMATES AND DISTANCES IMPLIED BY ESTIMATES OF OTHER FLOWS BASED ON OLS ESTIMATES.

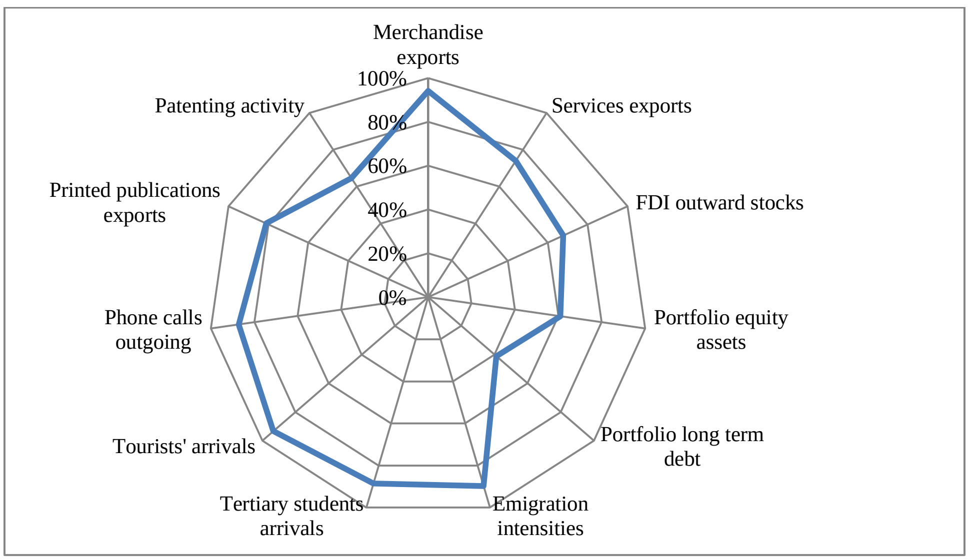

FIGURE 1B. CORRELATION BETWEEN GEOGRAPHICAL DISTANCE AND DISTANCES IMPLIED BY COEFFICIENTS FOR EACH TYPE OF FLOW BASED ON OLS ESTIMATES.

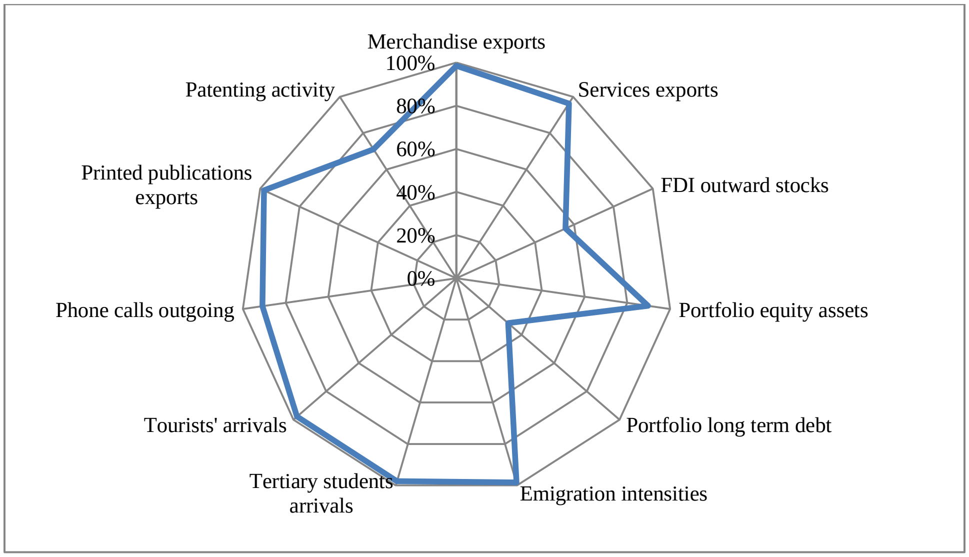

FIGURE 2. CORRELATION BETWEEN GEOGRAPHICAL DISTANCE AND AUGMENTED DISTANCES IMPLIED BY COEFFICIENTS FOR EACH TYPE OF FLOW BASED ON PPML ESTIMATES

VII. Tables

TABLE 1— COMPARISON BETWEEN OUR ESTIMATES FOR MERCHANDISE EXPORTS AND THE MEDIANS OF THE ESTIMATES OF THE STRUCTURAL GRAVITY MODELS.

Variables	Our estimates (OLS)	Our estimates (PPML)	Medians for the structural gravity models in Head and Mayer (2014) – Table 4
Common currency	-	-	0.98
Common official language	0.765	0.149	0.33
Colonial link	0.717	0.13	0.84
RTA/ Regional Bloc	0.314	0.530	0.28
Distance (logged)	-1.510	-0.748	-1.14
Share a common border	0.532	0.513	0.52
Ratio of PC income (logged)	-0.0177	0.0371	-
GDPs (logged)	0.571[a]	0.528[a]	0.86/0.67[b]

[a] Product of GDPs.

[b] Origin GDP / Destination GDP.

TABLE 2—GENERAL GRAVITY MODEL. OLS ESTIMATION

	Goods exports	Services exports	FDI outward stocks	Portfolio equity stocks	Portfolio long-term debt
Panel A					
Common official language	0.765***	0.380***	0.755***	0.556***	0.0856
	(0.0619)	(0.100)	(0.143)	(0.145)	(0.120)
Colonial linkage	0.717***	0.695***	1.266***	0.512**	0.176
	(0.136)	(0.160)	(0.195)	(0.233)	(0.182)
Trade agreement	0.314***	0.135			
	(0.0607)	(0.0965)			
Regional bloc			0.101	0.207	0.740***
			(0.171)	(0.143)	(0.130)
Distance	-1.510***	-0.974***	-1.210***	-1.011***	-0.925***
(logged)	(0.0373)	(0.0559)	(0.0756)	(0.0714)	(0.0574)
Share a common border	0.532***	0.131	-0.0140	0.388*	-0.0604
	(0.158)	(0.170)	(0.225)	(0.218)	(0.178)
Ratio of pc income	-0.0177	0.430***	-0.377***	-0.401***	-0.559***
(max / min) – logged	(0.0157)	(0.0422)	(0.0794)	(0.0603)	(0.0484)
Product of gdps	0.571***	0.576***	0.226**	0.498***	0.243***
(logged)	(0.0388)	(0.0652)	(0.0933)	(0.113)	(0.0858)
Constant	15.34***	15.49***	14.35***	6.570***	7.584***
	(0.582)	(1.127)	(1.237)	(1.592)	(1.140)
Observations	63,384	10,726	13,514	16,206	16,967
Adjusted R-squared	0.777	0.881	0.751	0.766	0.762

Notes: Robust standard errors are in parentheses, clustered by country-pair. Origin and destination specific fixed effects are included.

Source: Author calculations.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

TABLE 2 (CONT.)—GENERAL GRAVITY MODEL. OLS ESTIMATION

	Int. students arrivals [a]	Emigration	Intl. tourists arrivals	Outgoing phone calls	Printed publications exports	Intl. patenting activity
Panel B						
Common official language	1.176***	1.162***	0.954***	0.990***	1.854***	0.638***
	(0.0832)	(0.0754)	(0.0911)	(0.0848)	(0.0838)	(0.0710)
Colonial linkage	1.428***	1.183***	0.445*	0.713***	1.420***	0.750***
	(0.160)	(0.155)	(0.241)	(0.122)	(0.162)	(0.107)
Regional bloc	0.164*	-0.793***	0.561***	-0.0553	0.135	-0.367***
	(0.0994)	(0.111)	(0.175)	(0.0953)	(0.115)	(0.0983)
Distance	-1.149***	-1.401***	-	-1.132***	-1.917***	-0.274***
(logged)	(0.0451)	(0.0378)	(0.0584)	(0.0504)	(0.0474)	(0.0426)
Share a common	0.241	1.239***	0.759***	0.421***	0.236	0.467***
border	(0.152)	(0.163)	(0.226)	(0.105)	(0.162)	(0.116)
Ratio of pc income	-0.114***	0.110***	-	0.183***	-0.266***	-0.0607**
(max / min) – logged	(0.0280)	(0.0222)	(0.0316)	(0.0244)	(0.0214)	(0.0249)
Product of populations	0.215	-0.311***	1.747***	0.932***	0.463	2.593***
(logged)	(0.313)	(0.0622)	(0.325)	(0.181)	(0.388)	(0.386)
Constant	17.11***	14.12***	31.92***	15.11***	12.64***	23.91***
	(2.570)	(0.731)	(2.714)	(1.516)	(2.372)	(4.064)
Observations	18,837	38,352	10,752	13,752	36,894	11,004
Adjusted R-squared	0.715	0.802	0.879	0.893	0.701	0.820

Notes: Robust standard errors are in parentheses, clustered by country-pair. Origin and destination specific fixed effects are included.

Source: Author calculations.

[a] Tertiary Students has been analyzed based on data up to 2010 data since there are not more recent data available.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

TABLE 3—PERCENTAGE OF VARIANCE EXPLAINED BY DISTANCE VARIABLES

	C1[a]	C2[b]	C3[c]	C4[d]=[C2-C1]/[1-C1]	C5[e]=[C3-C1]/[1-C1]
Goods exports	0.679	0.771	0.777	29%	31%
Services exports	0.827	0.869	0.881	24%	31%
FDI outward stocks	0.661	0.732	0.751	21%	27%
Portfolio equity assets	0.705	0.758	0.766	18%	21%
Portfolio long term debt	0.672	0.747	0.762	23%	27%
Intl. Tertiary students	0.539	0.672	0.715	29%	38%
Emigration intensities	0.630	0.767	0.802	37%	46%
Intl. Tourists’ arrivals	0.735	0.865	0.879	49%	54%
Outgoing phone calls	0.793	0.867	0.893	36%	48%
Printed publication exports	0.478	0.664	0.701	36%	43%
Patenting activity	0.787	0.800	0.820	6%	15%

Source: Author calculations.

[a] R-squared of a gravity model with size variables (gdp/population) and fixed effects.

[b] R-squared of a gravity model with geographic distance, size variables (gdp/population), and fixed effects.

[c] R-squared of a gravity model with standard set of variables and fixed effects.

[d] % of the remaining variance explained by geographical distance.

[e] of the remaining variance explained by all distance variables.

TABLE 4—TOP 5 US PARTNERS

Ranking	1	2	3	4	5
Goods exports	CAN	MEX	CHN	JPN	UK
Services exports	UK	CAN	JPN	DEU	MEX
FDI outward stocks	UK	NLD	CAN	LUX	IRL
Portfolio equity assets	UK	JPN	CAN	CHE	FRA
Portfolio long – term debt	UK	CAN	AUS	NLD	FRA
Education	CHN	IND	KOR	JPN	CAN
Emigration	MEX	CAN	UK	DEU	AUS
Tourists	CAN	MEX	UK	JPN	DEU
Outgoing phone calls	MEX	CAN	IND	COL	CHN
Publications exports	CAN	UK	MEX	JPN	AUS
Patents	JPN	DEU	KOR	CAN	UK

Source: Author calculations.

TABLE 5—PAIRWISE CORRELATIONS

	(1)	(2)	(3)	(4)	(5)
Panel A					
Goods exports (1)	1				
Services exports (2)	0.7953	1			
FDI outward stocks (3)	0.6803	0.7580	1		
Portfolio equity assets (4)	0.6042	0.7037	0.6891	1	
Portfolio long term debt (5)	0.7407	**0.8646**	0.6813	0.7193	1
Education (6)	0.3842	0.3185	0.3411	0.2107	0.1772
Emigration (7)	0.6471	0.7602	0.6264	0.5914	0.6771
Tourists (8)	0.7864	**0.8832**	0.7726	0.6711	0.7387
Outgoing phone calls (9)	0.7288	**0.8156**	0.7451	0.6228	0.6396
Printed publications (10)	0.5233	0.7209	0.6185	0.4702	0.5654
Patents (11)	0.6666	0.6785	0.5027	0.6128	0.6383

Source: Author calculations.

TABLE 5 (CONT.)—PAIRWISE CORRELATIONS

	(6)	(7)	(8)	(9)	(10)
Panel B					
Education (6)	1				
Emigration (7)	0.2929	1			
Tourists (8)	0.3104	**0.8228**	1		
Outgoing phone calls (9)	0.4542	0.7675	**0.8642**	1	
Printed publications (10)	0.3155	0.5615	0.6279	0.5960	1
Patents (11)	0.2160	0.6012	0.7229	0.5522	0.4440

Source: Author calculations.

TABLE 6—GENERAL GRAVITY MODEL. PPML ESTIMATION.

	Goods exports	Services exports	FDI outward stocks	Portfolio equity stocks	Portfolio long-term debt
Panel A					
Common official language	0.149**	0.313***	0.390***	0.450***	0.523***
	(0.0663)	(0.0989)	(0.137)	(0.155)	(0.117)
Colonial linkage	0.130	0.276***	0.380***	0.201	0.107
	(0.0896)	(0.105)	(0.140)	(0.147)	(0.137)
Trade agreement	0.530***	0.416***			
	(0.0637)	(0.108)			
Regional bloc			0.0268	0.583**	1.098***
			(0.237)	(0.248)	(0.139)
Distance (logged)	-0.748***	-0.571***	-0.589***	-0.254***	-0.200***
	(0.0347)	(0.0537)	(0.0951)	(0.0882)	(0.0668)
Share a common border	0.513***	0.247**	0.306	0.338	0.0960
	(0.0663)	(0.0992)	(0.191)	(0.233)	(0.108)
Ratio of pc income	0.0371	0.0701	-0.419***	-0.442*	-0.250***
(max/min)-logged	(0.0252)	(0.0671)	(0.0984)	(0.251)	(0.0910)
Product of gdps	0.528***	0.388***	0.382***	0.443***	0.629***
(logged)	(0.0321)	(0.0653)	(0.143)	(0.127)	(0.0998)
Constant	14.01***	11.67***	14.94***	0.415	2.865***
	(0.564)	(1.370)	(0.977)	(1.441)	(1.091)
Observations	73,200	21,007	22,328	31,984	32,072
R-squared	0.923	0.884	0.767	0.807	0.898

Notes: Robust standard errors are in parentheses, clustered by country-pair. Origin and destination specific fixed effects are included.

Source: Author calculations.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

TABLE 6 (CONT.)—GENERAL GRAVITY MODEL. PPML ESTIMATION

	Int. students arrivals [a]	Emigration	Intl. tourists arrivals	Outgoing phone calls	Printed publications exports	Intl. patenting activity
Panel B						
Common official language	1.494***	0.688***	0.176	0.646***	1.031***	0.755***
	(0.155)	(0.154)	(0.147)	(0.100)	(0.190)	(0.109)
Colonial linkage	0.623***	1.043***	0.870***	0.524***	0.573***	-0.0841
	(0.171)	(0.186)	(0.316)	(0.102)	(0.190)	(0.102)
Regional bloc	-0.0918	-0.632***	0.653***	0.111	0.387	-0.0121
	(0.153)	(0.205)	(0.151)	(0.129)	(0.239)	(0.164)
Distance	-1.202***	-1.390***	-	-1.018***	-1.041***	-0.204***
(logged)	(0.0751)	(0.0990)	(0.104)	(0.0652)	(0.177)	(0.0597)
Share a common	-0.134	0.914***	0.669***	0.467***	0.420**	0.181
border	(0.153)	(0.173)	(0.233)	(0.110)	(0.192)	(0.168)
Ratio of pc income	-0.0951	-0.117*	0.121***	0.251***	0.0269	0.132**
(max / min) – logged	(0.0626)	(0.0620)	(0.0443)	(0.0378)	(0.0597)	(0.0624)
Product of populations	1.791***	-0.163	2.147***	1.327**	0.696	0.722
(logged)	(0.644)	(0.160)	(0.488)	(0.612)	(0.487)	(0.658)
Constant	27.91***	17.72***	34.15***	17.66***	11.78***	4.033
	(5.168)	(1.449)	(4.131)	(4.977)	(4.159)	(3.473)

Observations	28,398	60,656	24,576	13,752	69,888	48,128
Adjusted R-squared	0.843	0.875	0.899	0.907	0.757	0.972

Notes: Robust standard errors are in parentheses, clustered by country-pair. Origin and destination specific fixed effects are included.

Source: Author calculations.

[a] Tertiary Students has been analyzed based on data up to 2010 data since there are not more recent data available.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

References

Aggarwal, Raj, Colm Kearney, and Brian Lucey. 2012. "Gravity and Culture in Foreign Portfolio Investment." *Journal of Banking & Finance* no. 36 (2):525–538. doi: 10.1016/j.jbankfin.2011.08.007.

Anderson, James E. 1979. "A Theoretical Foundation for the Gravity Equation." *American Economic Review* no. 69 (1):106–116.

Anderson, James E., and Eric van Wincoop. 2003. "Gravity with Gravitas: A Solution to the Border Puzzle." *American Economic Review* no. 93 (1):170–192.

Archontakis, Fragiskos, and Nikos C. Varsakelis. 2011. "US Patents Abroad: Does Gravity Matter?" *The Journal of Technology Transfer* no. 36 (4):404–416. doi: 10.1007/s10961-010-9156-9.

Aristotelous, Kyriacos, and Stilianos Fountas. 1996. "An Empirical Analysis of Inward Foreign Direct Investment Flows in the EU with Emphasis on the Market Enlargement Hypothesis." *Journal of Common Market Studies* no. 34 (4):571–583.

Aviat, Antonin, and Nicolas Coeurdacier. 2007. "The Geography of Trade in Goods and Asset Holdings." *Journal of International Economics* no. 71 (1):22–51. doi: 10.1016/j.jinteco.2006.02.002.

Baxter, Marianne, and Michael A. Kouparitsas. 2005. "Determinants of Business Cycle Comovement: A Robust Analysis." *Journal of Monetary Economics* no. 52: 113–157. doi:10.1016/j.jmoneco.2004.08.002.

Beine, Michel, Romain Noël, and Lionel Ragot. 2014. "Determinants of the International Mobility of Students." *Economics of Education Review* no. 41:40–54. doi: 10.1016/j.econedurev.2014.03.003.

Belot, Michèle, and Sjef Ederveen. 2012. "Cultural Barriers in Migration between OECD Countries." *Journal of Population Economics* no. 25 (3):1077–

1105. doi: 10.1007/s00148-011-0356-x.

Bergstrand, Jeffrey H. 1985. “The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence.” *Review of Economics & Statistics* no. 67 (3):474–481.

———. 1989. “The Generalized Gravity Equation, Monopolistic Competition, and the Factor-Proportions Theory in International Trade.” *Review of Economics & Statistics* no. 71 (1):143–153.

Bergstrand, Jeffrey H., and Peter Egger. 2007. “A Knowledge-and-Physical-Capital Model of International Trade Flows, Foreign Direct Investment, and Multinational Enterprises.” *Journal of International Economics* no. 73 (2):278–308. doi: 10.1016/j.jinteco.2007.03.004.

Bessey, Donata. 2012. “International Student Migration to Germany.” *Empirical Economics* no. 42 (1):345–361. doi: 10.1007/s00181-010-0417-0.

Blonigen, Bruce A., Ronald B. Davies, and Keith Head. 2003. “Estimating the Knowledge-Capital Model of the Multinational Enterprise: Comment.” *American Economic Review* no. 93 (3):980–994.

Blonigen, Bruce A., Ronald B. Davies, Glen R. Waddell, and Helen T. Naughton. 2007. “FDI in Space: Spatial Autoregressive Relationships in Foreign Direct Investment.” *European Economic Review* no. 51 (5):1303–1325. doi: 10.1016/j.euroecorev.2006.08.006.

Braymen, Charles, and Kristie Briggs. 2015. “Income Differences in Services Trade Partners.” *Applied Economics Letters* no. 22 (8):593–596. doi: 10.1080/13504851.2014.959651.

Carey, H. C. 1858. *Principles of Social Science*. Vol. 1. New York: J.B. Lippincott & Co. https://www.questia.com/read/1345027/principles-of-social-science.

Carr, David L., James R. Markusen, and Keith E. Maskus. 2001. “Estimating the Knowledge-Capital Model of the Multinational Enterprise.” *American*

Economic Review no. 91 (3):693–708.

Chaney, Thomas. 2008. "Distorted Gravity: The Intensive and Extensive Margins of International Trade." *American Economic Review* no. 98 (4):1707–1721. doi: 10.1257/aer.98.4.1707.

Daude, Christian, and Marcel Fratzscher. 2008. "The Pecking Order of Cross-Border Investment." *Journal of International Economics* no. 74 (1):94–119. doi: 10.1016/j.jinteco.2007.05.010.

Desart, Henri-Guillaume. 1846. *Chemin De Fer Direct De Bruxelles Vers Gand, Par Alost, En Communication Avec Les Stations Diverses De La Capitale*. Bruxelles: E. Devroye.

Disdier, Anne-Célia, SilvioH T. Tai, Lionel Fontagné, and Thierry Mayer. 2010. "Bilateral Trade of Cultural Goods." *Review of World Economics* no. 145 (4):575–595. doi: 10.1007/s10290-009-0030-5.

Egger, Peter. 2008. "On the Role of Distance for Outward FDI." *Annals of Regional Science* no. 42 (2):375–389. doi: 10.1007/s00168-007-0166-y.

Egger, Peter H., and Andrea Lassmann. 2012. "The Language Effect in International Trade: A Meta-Analysis." *Economics Letters* no. 116 (2):221–224. doi: 10.1016/j.econlet.2012.02.018.

Eilat, Yair, and Liran Einav. 2004. "Determinants of International Tourism: A Three-Dimensional Panel Data Analysis." *Applied Economics* no. 36 (12):1315–1327. doi: 10.1080/0003684042000180897.

Fajgelbaum, Pablo, Gene M. Grossman, and Elhanan Helpman. 2015. "A Linder Hypothesis for Foreign Direct Investment." *Review of Economic Studies* no. 82 (1):83–121.

Ghemawat, Pankaj, Steven A. Altman. 2012. "DHL Global Connectedness Index 2012." Deutsche Post DHL, http://www.dhl.com/en/about_us/logistics_insights/studies_research/global_connectedness_index/global_connectedness_index_2012/gci_results.html.

Grogger, Jeffrey, and Gordon H. Hanson. 2011. “Income Maximization and the Selection and Sorting of International Migrants.” *Journal of Development Economics* no. 95 (1):42–57. doi: 10.1016/j.jdeveco.2010.06.003.

Guerin, Selen Sarisoy. 2006. “The Role of Geography in Financial and Economic Integration: A Comparative Analysis of Foreign Direct Investment, Trade and Portfolio Investment Flows.” *World Economy* no. 29 (2):189–209. doi: 10.1111/j.1467-9701.2006.00777.x.

Head, Keith, and Thierry Mayer. 2002. “Illusory Border Effects: Distance Mismeasurement Inflates Estimates of Home Bias in Trade.” CEPII Working paper 2002-1, http://www.cepii.fr/anglaisgraph/workpap/pdf/2002/wp02-01.pdf.

———. 2014. “Gravity Equations: Workhorse, Toolkit, and Cookbook.” In *Handbook of International Economics*, edited by Gita Gopinath, Elhanan Helpman, and Kenneth S. Rogoff, 131–195.

Head, Keith, Thierry Mayer, and John Ries. 2010. “The Erosion of Colonial Trade Linkages after Independence.” *Journal of International Economics* no. 81 (1):1–14. doi: 10.1016/j.jinteco.2010.01.002.

Helpman, Elhanan, Marc J. Melitz, and Stephen R. Yeaple. 2004. “Export Versus FDI with Heterogeneous Firms.” *American Economic Review* no. 94 (1):300–316. doi: 10.1257/000282804322970814.

Helpman, Elhanan, Marc Melitz, and Yona Rubinstein. 2008. “Estimating Trade Flows: Trading Partners and Trading Volumes.” *Quarterly Journal of Economics* no. 123 (2):441–487.

Kandilov, Ivan T., and Thomas Grennes. 2012. “The Determinants of Service Offshoring: Does Distance Matter?” *Japan and the World Economy* no. 24 (1):36–43. doi: 10.1016/j.japwor.2011.12.001.

Keller, Wolfgang. 2002. “Geographic Localization of International Technology Diffusion.” *American Economic Review* no. 92 (1):120–142. doi:

10.1257/000282802760015630.

Keum, Kiyong. 2010. "Tourism Flows and Trade Theory: A Panel Data Analysis with the Gravity Model." *Annals of Regional Science* no. 44 (3):541–557. doi: 10.1007/s00168-008-0275-2.

Kim, Keuntae, and Joel E. Cohen. 2010. "Determinants of International Migration Flows to and from Industrialized Countries: A Panel Data Approach Beyond Gravity." *International Migration Review* no. 44 (4):899–932. doi: 10.1111/j.1747-7379.2010.00830.x.

Kimura, Fukunari, and Hyun-Hoon Lee. 2006. "The Gravity Equation in International Trade in Services." *Review of World Economics* no. 142 (1):92–121. doi: 10.1007/s10290-006-0058-8.

Kleinert, Jörn, and Farid Toubal. 2010. "Gravity for FDI." *Review of International Economics* no. 18 (1):1–13. doi: 10.1111/j.1467-9396.2009.00869.x.

Lane, Philip R., and Gian Maria Milesi-Ferretti. 2008. "International Investment Patterns." *Review of Economics and Statistics* no. 90 (3):538–549. doi: 10.1162/rest.90.3.538.

Levy-Yeyati, Eduardo Levy, Ernesto Stein, and Christian Daude. 2003. "Regional Integration and the Location of FDI." Inter-American Development Bank, Research Department working paper 492, http://www.econstor.eu/bitstream/10419/87947/1/idb-wp_492.pdf.

Lewer, Joshua J., and Hendrik Van den Berg. 2008. "A Gravity Model of Immigration." *Economics Letters* no. 99 (1):164–167. doi: 10.1016/j.econlet.2007.06.019.

Linder, Staffan Burenstam. 1961. *An Essay on Trade and Transformation*. Stockholm, New York: Almqvist & Wiksell; J. Wiley.

Lucas, Robert B., Jr. 1990. "Why Doesn't Capital Flow from Rich to Poor Countries?" *American Economic Review* no. 80 (2):92–96.

Marchetti, Juan A. 2011. "Do Economic Integration Agreements Lead to Deeper Integration of Services Markets?" In *International Handbook on the Economics of Integration: Factor Mobility, Agriculture, Environment and Quantitative Studies*, edited by Miroslav N. Jovanović, 435–454. Cheltenham, UK; Northampton, MA: Edward Elgar.

Markusen, James R., and Keith E. Maskus. 2002. "Discriminating among Alternative Theories of the Multinational Enterprise." *Review of International Economics* no. 10 (4):694–707.

Mayda, Anna Maria. 2010. "International Migration: A Panel Data Analysis of the Determinants of Bilateral Flows." *Journal of Population Economics* no. 23 (4):1249–1274. doi: 10.1007/s00148-009-0251-x.

Meng, Yinghua. 2011. "Determined Factors of International Trade in Cultural Goods of China: A Panel Data Analysis." In *Applied Economics, Business and Development*, edited by Qingyuan Zhou, 429–435. Springer Berlin Heidelberg.

Odlyzko, Andrew. 2014. "The Forgotten Discovery of Gravity Models and the Inefficiency of Early Railway Networks." SSRN Working Paper 2490241. doi: http://dx.doi.org/10.2139/ssrn.2490241.

Oh, Chang Hoon, W. Travis Selmier, and Donald Lien. 2011. "International Trade, Foreign Direct Investment, and Transaction Costs in Languages." *The Journal of Socio-Economics* no. 40 (6):732–735. doi: 10.1016/j.socec.2011.08.003.

Olivero, María Pía, and Yoto V. Yotov. 2012. "Dynamic Gravity: Endogenous Country Size and Asset Accumulation." *Canadian Journal of Economics* no. 45 (1):64–92. doi: 10.1111/j.1540-5982.2011.01687.x.

Orefice, Gianluca, Luiz Lima, and Erik Figueiredo. 2014. "Migration and Regional Trade Agreement: A (New) Gravity Estimation." CEPII Research Center Working Paper 2014-13, http://www.cepii.fr/PDF_PUB/wp/2014/wp2014-13.pdf.

Ortega, Francesc, and Giovanni Peri. 2013. "The Effect of Income and Immigration Policies on International Migration." *Migration Studies* no. 1 (1):47–74. doi: 10.1093/migration/mns004.

Picci, Lucio. 2010. "The Internationalization of Inventive Activity: A Gravity Model Using Patent Data." *Research Policy* no. 39 (8):1070–1081. doi: 10.1016/j.respol.2010.05.007.

Portes, Richard, and Hélène Rey. 2005. "The Determinants of Cross-Border Equity Flows." *Journal of International Economics* no. 65 (2):269–296. doi: 10.1016/j.jinteco.2004.05.002.

Ravenstein, E. G. 1885. "The Laws of Migration." *Journal of the Statistical Society of London* no. 48 (2):167–235. doi: 10.2307/2979181.

Rietveld, Piet, and Leon Janssen. 1990. "Telephone Calls and Communication Barriers." *Annals of Regional Science* no. 24 (4):307–318.

González, Carlos Rodríguez, Ricardo Bustillo Mesanza, and Petr Mariel. 2011. "The Determinants of International Student Mobility Flows: An Empirical Study on the Erasmus Programme." *Higher Education* no. 62 (4):413–430. doi: 10.1007/s10734-010-9396-5.

Santos Silva, J. M. C., and Silvana Tenreyro. 2006. "The Log of Gravity." *Review of Economics & Statistics* no. 88 (4):641–658.

Spence, Michael. 2011. *The Next Convergence: The Future of Economic Growth in a Multispeed World*. New York: Farrar, Straus and Giroux.

Tinbergen, Jan. 1962. *Shaping the World Economy: Suggestions for an International Economic Policy*. New York: Twentieth Century Fund.

Wong, Wei-Kang. 2008. "Comparing the Fit of the Gravity Model for Different Cross-Border Flows." *Economics Letters* no. 99 (3):474–477. doi: 10.1016/j.econlet.2007.09.018.

Globalization and Gravity

By PANKAJ GHEMAWAT AND TAMARA DE LA MATA

Online Appendix

APPENDIX A. Sample and Data Sources

I. List of countries

Albania, Algeria, Angola, Argentina, Australia, Austria, Bangladesh, Belgium, Bolivia, Brazil, Bulgaria, Burkina Faso, Cameroon, Canada, Chile, China, Colombia, Congo, Democratic Republic of the Congo, Costa Rica, Côte d'Ivoire, Cyprus, Denmark, Dominican Republic, Ecuador, Egypt, Ethiopia, Finland, France, Gabon, Gambia, Germany, Ghana, Greece, Guatemala, Guinea-Bissau, Guyana, Haiti, Honduras, Hong Kong SAR (China), Hungary, India, Indonesia, Iran, Ireland, Israel, Italy, Jamaica, Japan, Kenya, Republic of Korea, Luxembourg, Madagascar, Malawi, Malaysia, Mali, Mauritania, Mauritius, Mexico, Morocco, Mozambique, Netherlands, New Zealand, Nicaragua, Niger, Nigeria, Norway, Pakistan, Panama, Paraguay, Peru, Philippines, Poland, Portugal, Romania, El Salvador, Saudi Arabia, Senegal, Sierra Leone, Singapore, Spain, Sri Lanka, Sudan, Sweden, Switzerland, Syria, Thailand, Trinidad and Tobago, Tunisia, Turkey, Uganda, United Kingdom, Uruguay, USA, Venezuela, Zambia, Zimbabwe.

II. Data sources

A. Dependent variables.

Merchandise Exports

Direction of Trade Statistics (DOTS) from IMF. Goods, Value of Exports. US$. (http://elibrary-data.imf.org/QueryBuilder.aspx?key=19784661&s=322).

UN Comtrade data. Exports reported by exporters, US$. (http://comtrade.un.org/db/).

Some countries report data on USSR, Czechoslovakia, Belgium-Luxembourg, and Yugoslavia, instead of on the current countries. In the cases of the countries that report flows with these group of countries and not with the current countries separately (Mexico and Sudan), we use the UN Comtrade data.

Interpolation or repetition is not required for the IMF data. No need to fill gaps with zeroes, since the original data is already treated. Country-pairs with no exports reported in the UN Comtrade data are considered as zero, with the exception of 2012 data, when the data was incomplete at the time it was downloaded. Then, if one country reports data with one partner in 2011 and not in 2012, we repeat this bilateral flow in 2012, assuming that this is a missing value and we fill rest of the gaps with zeroes in the UN Comtrade data.

Services Exports

UN Services Trade, Services Exports, Total Exports, EBOPS, US$. Data reported by exporters.

Gaps for those countries that report data are filled with zeroes, with the exception of 2012 data, when the data was incomplete at the time it was downloaded. Then, if one country reports data with one partner in 2011 and not in 2012, we repeat this bilateral flow in 2012, assuming that this is a missing value and we fill gaps with zeroes in the UN Services data.

UN ServiceTrade data are compiled according to the recommendations contained in the *Manual on Statistics of International Trade in Services*. According to these recommendations, there are certain categories of services, such as insurance or merchanting, for which negative flows can be observed: i.e., in *merchanting*, if the goods are resold for less than the original cost of purchase—that is, the merchant takes a loss on the sale—then a negative export of merchanting services would be recorded. (Box 6 in MSITS 2002.) Depending on the importance of the bilateral trade in those services categories, you may observe a negative value at a higher level, as well (total EBOPS). These negative flows have been dropped from the analysis.

FDI Outward Stocks

OECD (2012), *OECD.Stat*, (database). US$, millions. doi: 10.1787/data-00285-en. (Accessed on 09 December 2013).

This dataset is complemented with other sources:

Brazil: Columbia FDI Profiles (http://www.vcc.columbia.edu/content/columbia-fdi-profiles).

China: National Bureau of Statistics of China (http://www.stats.gov.cn/tjsj/ndsj/2013/indexeh.htm).

Singapore: Department of Statistics Singapore (http://www.singstat.gov.sg/pubn/business.html#sia; http://www.singstat.gov.sg/statistics/browse_by_theme/investment.html). Millions of S$ converted in US $ using the exchange rate: USD/SGD = 0.79320.

Hong Kong SAR (China): Different reports available at the Government of the Hong Kong Special Administrative Region (http://www.censtatd.gov.hk/products_and_services/products/publications/statistical_report/national_income_and_bop/index_cd_B1040003_dt_latest.jsp). Billions of HKD converted in US $ using the exchange rate: USD/HKD = 0.12903.

Interpolation and repetition is done. FDI Data from the OECD distinguish between zeroes and missing values. For data coming from different sources we follow the rule of filling gaps with zeroes when the coverage of the bilateral dataset is higher than 90%, we fill gaps with zeroes and when it is below the threshold of 90%, we consider gaps as missing values. Gaps are filled with zeroes for the FDI Outward stocks from Brazil, Singapore, and Hong. FDI Outward stocks are considered in absolute value.

Portfolio equity assets (IMF – CPIS)

Cross-Economy Tables comprise, in matrix form, data from the individual economy tables of residents' holdings of securities issued by nonresidents (reported data) and the derived data for nonresidents' holdings of securities issued by residents (derived data). The geographic breakdown of the reported data is limited to the CPIS participating economies, while the geographic breakdown of the derived data covers all economies that issue securities that are held by CPIS participating economies. The cross economy metadata presented below are based on information provided by economies that participated in the CPIS Metadata Survey. (http://cpis.imf.org/ Table 8.1. US $, Millions).

Zeroes are identified in the original source. We try to fill missing values with interpolation. Portfolio equity assets are considered in absolute value.

Portfolio long term debt (IMF – CPIS)

Cross-Economy Tables comprise, in matrix form, data from the individual economy tables of residents' holdings of securities issued by nonresidents (reported data) and the derived data for nonresidents' holdings of securities issued by residents (derived data). The geographic breakdown of the reported data is limited to the CPIS participating economies, while the geographic breakdown of the derived data covers all economies that issue securities that are held by CPIS participating economies. The cross economy metadata presented below are based on information provided by economies that participated in the CPIS Metadata Survey. http://cpis.imf.org/ Table 8.2A. US $, Millions. Table 8.2.A: Geographic Breakdown of Total Portfolio Investment Assets: long-term debt securities.

Zeroes are identified in the original source. We try to fill missing values with interpolation. Portfolio long term debt is considered in absolute value.

Students arrivals

Data based on students' mobility by country of origin, 2005–2012. Data incomplete since 2010– we don't use the data available for 2011 and 2012.

Zeroes are identified in the original source. Interpolation and repetition is done for the complete dataset to fill gaps. Source: UNESCO Institute for Statistics. (http://stats.uis.unesco.org/unesco/TableViewer/tableView.aspx).

Emigration

United Nations Population Division (Department of Economic and Social Affairs). Trends in International Migrant Stock: The 2013 revision (United Nations database, POP/DB/MIG/Stock/Rev.2013). International migrant stock by destination and origin - Estimates of the total number of international migrants by country that refer to 1 July of the reference year.
(http://esa.un.org/unmigration/TIMSO2013/migrantstocks2013.htm?msdo).

Empty cells are considered as missing values when the coverage for immigration is <90% of the total (World column), otherwise, gaps are filled with zeroes. Interpolation and repetition is done using just the 2000 and 2013 UN Data. Then we keep the data for the period 2005–2012 for consistency with the rest of the variables.

Tourism Arrivals

Compendium of Tourism Statistics from the UNWTO: 2005–2012. When countries report from country-pairs as Spain and Portugal, Czech Republic and Slovakia, they are removed from the analysis.

Missing values have been interpolated and repeated to obtain a complete dataset from 2005 to 2012. Gaps are filled with zeroes.

Printed Publications Exports

UNComtrade data – HS 49 code. Exports reported by exporters, US$. (http://comtrade.un.org/db/). Interpolation is done to fill gaps. If one country reports data with one partner in 2011 and not in 2012, we repeat this bilateral flow in 2012, assuming that this is a missing value and we fill rest of the gaps with zeroes.

Outgoing Phone calls

Minutes of international phone calls (VoIP and TDM phone calls) from Telegeography. Before interpolation, when coverage is <0.7 or >1.05, we set the flows as missing for the country in that year for all the partners. Also, when for one country-pair, it is reported zero minutes of phone calls, this is converted in a missing value. Then, interpolation and repetition is done. In this case, gaps are not filled with zeroes, because this dataset is a sample.

Patenting activity

Total patent applications (direct and PCT national phase entries) in patent offices by country of origin. Source: WIPO. Gaps are filled with zeroes.

B. Independent variables

Common official language

Dummy that takes the value 1 if both countries share the same official language and 0 otherwise. Source: CEPII (http://www.cepii.fr/CEPII/en/bdd_modele/bdd.asp).

Colonial linkage

Dummy that takes the value 1 if both countries have ever had a colonial linkage and 0 otherwise. In the original dataset, Spain and the USA seem to have a colonial linkage, while we set this country pair as never having had a colonial linkage. Similarly, Spain and Dominican Republic is set as never have a colonial linkage in our dataset. Source: CEPII.

Trade agreement

Dummy variable that takes the value of 1 when the countries have a trade agreement in force. Source: CEPII updated up to 2012 with data from WTO (www.wto.org).

Regional bloc

Dummy variable that takes the value of 1 when both countries belong to the same regional bloc of the following: MERCOSUR, ASEAN, CARICOM, EU, NAFTA, and GCC. Source: CEPII updated up to 2012 with data from WTO.

Distance (logged)

Bilateral distances between the 25 biggest cities of the two countries, those inter-city distances being weighted by the share of the city in the overall country's population. Source: CEPII.

Share a common border

Dummy that takes the value 1 if the countries share a common border and 0 otherwise. Source: CEPII.

Ratio of pc income (max/min) – logged

Logarithm of the ratio of the maximum and the minimum per capita income (maximum/minimum) (current US $). Source: International Monetary Fund (IMF). *World Economic Outlook (WEO) Database* April 2014 version. (http://www.imf.org/external/pubs/ft/weo/2014/01/weodata/index.aspx).

Product of gdps (logged)

Logarithm of the product of the gdps (current US $). Source: IMF WEO April 2014.

Product of populations (logged)

Logarithm of the product of the populations. Source: IMF WEO April 2014.

APPENDIX B. Tables.

TABLE B.1.— MATRIX OF CORRELATIONS BETWEEN THE COMPOSITE DISTANCES ACCORDING TO THE COEFFICIENT ESTIMATES OBTAINED FOR DIFFERENT KIND OF FLOWS AND GEOGRAPHICAL DISTANCE (OLS ESTIMATES).

	(1)	(2)	(3)	(4)	(5)
Panel A					
Goods exports (1)	-				
Services exports (2)	0.765	-			
FDI outward stocks	0.689	0.154	-		
Portfolio equity assets	0.596	0.052	**0.988**	-	
Portfolio long term	0.333	-0.164	0.852	**0.915**	-
Emigration (6)	**0.970**	0.870	0.545	0.439	0.159
Students (7)	**0.947**	0.597	0.842	0.766	0.491
Tourists (8)	**0.984**	0.677	0.777	0.691	0.429
Outgoing phone calls	**0.932**	**0.927**	0.446	0.340	0.065
Printed publications	0.891	0.449	0.898	0.824	0.577
Patents (11)	0.720	0.421	0.737	0.692	0.431
Geographical distance	**0.942**	0.740	0.678	0.609	0.411

Source: Author calculations.

TABLE B.1.(CONT.)— MATRIX OF CORRELATIONS BETWEEN THE COMPOSITE DISTANCES ACCORDING TO THE COEFFICIENT ESTIMATES OBTAINED FOR DIFFERENT KIND OF FLOWS AND GEOGRAPHICAL DISTANCE (OLS ESTIMATES).

	(6)	(7)	(8)	(9)	(10)	(11)
(Panel B						
Emigration (6)	-					
Students (7)	0.89	-				
Tourists (8)	**0.93**	**0.97**	-			
Outgoing phone calls	**0.98**	0.83	0.88	-		
Printed publications	0.80	**0.95**	**0.94**	0.71	-	
Patents (11)	0.70	0.85	0.75	0.66	0.75	-
Geographical distance	0.89	0.88	**0.93**	0.87	0.81	0.64

Source: Author calculations.

TABLE B.2. –PERCENTAGE OF VARIANCE EXPLAINED BY GRAVITY VARIABLES.

	C1[a]	C2[b]	C3[c]	C4[d]=[C2-C1]/[1-C1]	C5[e]=[C3-C1]/[1-C1]
Goods exports	0.470	0.885	0.923	78%	85%
Services exports	0.708	0.851	0.884	49%	60%
FDI outward stocks	0.714	0.728	0.767	5%	19%
Portfolio equity	0.639	0.750	0.807	31%	47%
Portfolio long term	0.729	0.853	0.898	46%	62%
Intl. Tertiary	0.690	0.811	0.843	39%	49%
Emigration	0.460	0.853	0.875	73%	77%
Intl. Tourists'	0.299	0.858	0.899	80%	86%
Outgoing phone	0.635	0.879	0.907	67%	75%
Printed publication	0.286	0.725	0.757	61%	66%
Patenting activity	0.947	0.959	0.972	23%	47%

Source: Author calculations.

[a] R-squared of a gravity model with size variables (gdp/population) and fixed effects.

[b] R-squared of a gravity model with geographic distance, size variables (gdp/population) and fixed effects.

[c] R-squared of a gravity model with standard set of variables and fixed effects.

[d] % of the remaining variance explained by geographical distance.

[e] % of the remaining variance explained by all distance variables.

TABLE B.3.— MATRIX OF CORRELATIONS BETWEEN THE COMPOSITE DISTANCES ACCORDING TO THE COEFFICIENT ESTIMATES OBTAINED FOR DIFFERENT KIND OF FLOWS AND GEOGRAPHICAL DISTANCE (PPML ESTIMATES).

	(1)	(2)	(3)	(4)	(5)
Panel A					
Goods exports (1)	-				
Services exports (2)	**0.985**	-			
FDI outward stocks	0.470	0.484	-		
Portfolio equity assets	0.839	**0.902**	0.460	-	
Portfolio long term	0.450	0.499	**0.936**	0.587	-
Emigration (6)	**0.921**	**0.919**	0.399	0.827	0.372
Students (7)	0.895	**0.918**	0.400	0.882	0.416
Tourists (8)	**0.905**	0.874	0.225	0.741	0.185
Outgoing phone calls	0.820	0.826	0.042	0.771	0.063
Printed publications	**0.915**	**0.938**	0.420	**0.901**	0.436
Patents (11)	0.580	0.657	-0.085	0.763	0.027
Geographical distance	**0.959**	**0.924**	0.431	0.790	0.378

Source: Author calculations.

TABLE B.3. (CONT.)— MATRIX OF CORRELATIONS BETWEEN THE COMPOSITE DISTANCES ACCORDING TO THE COEFFICIENT ESTIMATES OBTAINED FOR DIFFERENT KIND OF FLOWS AND GEOGRAPHICAL DISTANCE (PPML ESTIMATES).

	(6)	(7)	(8)	(9)	(10)	(11)
(Panel B						
Emigration (6)	-					
Students (7)	**0.98**	-				
Tourists (8)	**0.94**	0.89	-			
Outgoing phone calls	0.89	0.89	**0.93**	-		
Printed publications	**0.98**	**0.99**	**0.90**	0.89	-	
Patents (11)	0.67	0.74	0.64	0.83	0.74	-
Geographical distance	**0.94**	0.89	**0.95**	0.84	**0.91**	0.55

Source: Author calculations.

Disclosure Statement

Globalization and Gravity

By PANKAJ GHEMAWAT and TAMARA DE LA MATA*

IESE Business School and the NYU Stern School of Business supported this research financially. The author declares that he has no relevant or material financial interests that relate to the research described in this paper.

Pankaj Ghemawat

October 20, 2015

Disclosure Statement

Globalization and Gravity

By PANKAJ GHEMAWAT and TAMARA DE LA MATA*

IESE Business School and the NYU Stern School of Business supported this research financially. The author declares that he has no relevant or material financial interests that relate to the research described in this paper.

Tamara de la Mata

October 20, 2015

Globalization and Gravity

By PANKAJ GHEMAWAT AND TAMARA DE LA MATA

Online Appendix

APPENDIX A. Sample and Data Sources

I. List of countries

Albania, Algeria, Angola, Argentina, Australia, Austria, Bangladesh, Belgium, Bolivia, Brazil, Bulgaria, Burkina Faso, Cameroon, Canada, Chile, China, Colombia, Congo, Democratic Republic of the Congo, Costa Rica, Côte d'Ivoire, Cyprus, Denmark, Dominican Republic, Ecuador, Egypt, Ethiopia, Finland, France, Gabon, Gambia, Germany, Ghana, Greece, Guatemala, Guinea-Bissau, Guyana, Haiti, Honduras, Hong Kong SAR (China), Hungary, India, Indonesia, Iran, Ireland, Israel, Italy, Jamaica, Japan, Kenya, Republic of Korea, Luxembourg, Madagascar, Malawi, Malaysia, Mali, Mauritania, Mauritius, Mexico, Morocco, Mozambique, Netherlands, New Zealand, Nicaragua, Niger, Nigeria, Norway, Pakistan, Panama, Paraguay, Peru, Philippines, Poland, Portugal, Romania, El Salvador, Saudi Arabia, Senegal, Sierra Leone, Singapore, Spain, Sri Lanka, Sudan, Sweden, Switzerland, Syria, Thailand, Trinidad and Tobago, Tunisia, Turkey, Uganda, United Kingdom, Uruguay, USA, Venezuela, Zambia, Zimbabwe.

II. Data sources

A. Dependent variables.

Merchandise Exports

Direction of Trade Statistics (DOTS) from IMF. Goods, Value of Exports. US$. (http://elibrary-data.imf.org/QueryBuilder.aspx?key=19784661&s=322).

UN Comtrade data. Exports reported by exporters, US$. (http://comtrade.un.org/db/).

Some countries report data on USSR, Czechoslovakia, Belgium-Luxembourg, and Yugoslavia, instead of on the current countries. In the cases of the countries that report flows with these group of countries and not with the current countries separately (Mexico and Sudan), we use the UN Comtrade data.

Interpolation or repetition is not required for the IMF data. No need to fill gaps with zeroes, since the original data is already treated. Country-pairs with no exports reported in the UN Comtrade data are considered as zero, with the exception of 2012 data, when the data was incomplete at the time it was downloaded. Then, if one country reports data with one partner in 2011 and not in 2012, we repeat this bilateral flow in 2012, assuming that this is a missing value and we fill rest of the gaps with zeroes in the UN Comtrade data.

Services Exports

UN Services Trade, Services Exports, Total Exports, EBOPS, US$. Data reported by exporters.

Gaps for those countries that report data are filled with zeroes, with the exception of 2012 data, when the data was incomplete at the time it was downloaded. Then, if one country reports data with one partner in 2011 and not in 2012, we repeat this bilateral flow in 2012, assuming that this is a missing value and we fill gaps with zeroes in the UN Services data.

UN ServiceTrade data are compiled according to the recommendations contained in the *Manual on Statistics of International Trade in Services*. According to these recommendations, there are certain categories of services, such as insurance or merchanting, for which negative flows can be observed: i.e., in *merchanting*, if the goods are resold for less than the original cost of purchase—that is, the merchant takes a loss on the sale—then a negative export of merchanting services would be recorded. (Box 6 in MSITS 2002.) Depending on the importance of the bilateral trade in those services categories, you may observe a negative value at a higher level, as well (total EBOPS). These negative flows have been dropped from the analysis.

FDI Outward Stocks

OECD (2012), *OECD.Stat*, (database). US$, millions. doi: 10.1787/data-00285-en. (Accessed on 09 December 2013).

This dataset is complemented with other sources:

Brazil: Columbia FDI Profiles (http://www.vcc.columbia.edu/content/columbia-fdi-profiles).

China: National Bureau of Statistics of China (http://www.stats.gov.cn/tjsj/ndsj/2013/indexeh.htm).

Singapore: Department of Statistics Singapore (http://www.singstat.gov.sg/pubn/business.html#sia; http://www.singstat.gov.sg/statistics/browse_by_theme/investment.html). Millions of S$ converted in US $ using the exchange rate: USD/SGD = 0.79320.

Hong Kong SAR (China): Different reports available at the Government of the Hong Kong Special Administrative Region (http://www.censtatd.gov.hk/products_and_services/products/publications/statistical_report/national_income_and_bop/index_cd_B1040003_dt_latest.jsp). Billions of HKD converted in US $ using the exchange rate: USD/HKD = 0.12903.

Interpolation and repetition is done. FDI Data from the OECD distinguish between zeroes and missing values. For data coming from different sources we follow the rule of filling gaps with zeroes when the coverage of the bilateral dataset is higher than 90%, we fill gaps with zeroes and when it is below the threshold of 90%, we consider gaps as missing values. Gaps are filled with zeroes for the FDI Outward stocks from Brazil, Singapore, and Hong. FDI Outward stocks are considered in absolute value.

Portfolio equity assets (IMF – CPIS)

Cross-Economy Tables comprise, in matrix form, data from the individual economy tables of residents' holdings of securities issued by nonresidents (reported data) and the derived data for nonresidents' holdings of securities issued by residents (derived data). The geographic breakdown of the reported data is limited to the CPIS participating economies, while the geographic breakdown of the derived data covers all economies that issue securities that are held by CPIS participating economies. The cross economy metadata presented below are based on information provided by economies that participated in the CPIS Metadata Survey. (http://cpis.imf.org/ Table 8.1. US $, Millions).

Zeroes are identified in the original source. We try to fill missing values with interpolation. Portfolio equity assets are considered in absolute value.

Portfolio long term debt (IMF – CPIS)

Cross-Economy Tables comprise, in matrix form, data from the individual economy tables of residents' holdings of securities issued by nonresidents (reported data) and the derived data for nonresidents' holdings of securities issued by residents (derived data). The geographic breakdown of the reported data is limited to the CPIS participating economies, while the geographic breakdown of the derived data covers all economies that issue securities that are held by CPIS participating economies. The cross economy metadata presented below are based on information provided by economies that participated in the CPIS Metadata Survey. http://cpis.imf.org/ Table 8.2A. US $, Millions. Table 8.2.A: Geographic Breakdown of Total Portfolio Investment Assets: long-term debt securities.

Zeroes are identified in the original source. We try to fill missing values with interpolation. Portfolio long term debt is considered in absolute value.

Students arrivals

Data based on students' mobility by country of origin, 2005–2012. Data incomplete since 2010– we don't use the data available for 2011 and 2012.

Zeroes are identified in the original source. Interpolation and repetition is done for the complete dataset to fill gaps. Source: UNESCO Institute for Statistics. (http://stats.uis.unesco.org/unesco/TableViewer/tableView.aspx).

Emigration

United Nations Population Division (Department of Economic and Social Affairs). Trends in International Migrant Stock: The 2013 revision (United Nations database, POP/DB/MIG/Stock/Rev.2013). International migrant stock by destination and origin - Estimates of the total number of international migrants by country that refer to 1 July of the reference year.
(http://esa.un.org/unmigration/TIMSO2013/migrantstocks2013.htm?msdo).

Empty cells are considered as missing values when the coverage for immigration is <90% of the total (World column), otherwise, gaps are filled with zeroes. Interpolation and repetition is done using just the 2000 and 2013 UN Data. Then we keep the data for the period 2005–2012 for consistency with the rest of the variables.

Tourism Arrivals

Compendium of Tourism Statistics from the UNWTO: 2005–2012. When countries report from country-pairs as Spain and Portugal, Czech Republic and Slovakia, they are removed from the analysis.

Missing values have been interpolated and repeated to obtain a complete dataset from 2005 to 2012. Gaps are filled with zeroes.

Printed Publications Exports

UNComtrade data – HS 49 code. Exports reported by exporters, US$. (http://comtrade.un.org/db/). Interpolation is done to fill gaps. If one country reports data with one partner in 2011 and not in 2012, we repeat this bilateral flow in 2012, assuming that this is a missing value and we fill rest of the gaps with zeroes.

Outgoing Phone calls

Minutes of international phone calls (VoIP and TDM phone calls) from Telegeography. Before interpolation, when coverage is <0.7 or >1.05, we set the flows as missing for the country in that year for all the partners. Also, when for one country-pair, it is reported zero minutes of phone calls, this is converted in a missing value. Then, interpolation and repetition is done. In this case, gaps are not filled with zeroes, because this dataset is a sample.

Patenting activity

Total patent applications (direct and PCT national phase entries) in patent offices by country of origin. Source: WIPO. Gaps are filled with zeroes.

B. Independent variables

Common official language

Dummy that takes the value 1 if both countries share the same official language and 0 otherwise. Source: CEPII (http://www.cepii.fr/CEPII/en/bdd_modele/bdd.asp).

Colonial linkage

Dummy that takes the value 1 if both countries have ever had a colonial linkage and 0 otherwise. In the original dataset, Spain and the USA seem to have a colonial linkage, while we set this country pair as never having had a colonial linkage. Similarly, Spain and Dominican Republic is set as never have a colonial linkage in our dataset. Source: CEPII.

Trade agreement

Dummy variable that takes the value of 1 when the countries have a trade agreement in force. Source: CEPII updated up to 2012 with data from WTO (www.wto.org).

Regional bloc

Dummy variable that takes the value of 1 when both countries belong to the same regional bloc of the following: MERCOSUR, ASEAN, CARICOM, EU, NAFTA, and GCC. Source: CEPII updated up to 2012 with data from WTO.

Distance (logged)

Bilateral distances between the 25 biggest cities of the two countries, those inter-city distances being weighted by the share of the city in the overall country's population. Source: CEPII.

Share a common border

Dummy that takes the value 1 if the countries share a common border and 0 otherwise. Source: CEPII.

Ratio of pc income (max/min) – logged

Logarithm of the ratio of the maximum and the minimum per capita income (maximum/minimum) (current US $). Source: International Monetary Fund (IMF). *World Economic Outlook (WEO) Database* April 2014 version. (http://www.imf.org/external/pubs/ft/weo/2014/01/weodata/index.aspx).

Product of gdps (logged)

Logarithm of the product of the gdps (current US $). Source: IMF WEO April 2014.

Product of populations (logged)

Logarithm of the product of the populations. Source: IMF WEO April 2014.

APPENDIX B. Tables.

TABLE B.1.— MATRIX OF CORRELATIONS BETWEEN THE COMPOSITE DISTANCES ACCORDING TO THE COEFFICIENT ESTIMATES OBTAINED FOR DIFFERENT KIND OF FLOWS AND GEOGRAPHICAL DISTANCE (OLS ESTIMATES).

	(1)	(2)	(3)	(4)	(5)
Panel A					
Goods exports (1)	-				
Services exports (2)	0.765	-			
FDI outward stocks	0.689	0.154	-		
Portfolio equity assets	0.596	0.052	**0.988**	-	
Portfolio long term	0.333	-0.164	0.852	**0.915**	-
Emigration (6)	**0.970**	0.870	0.545	0.439	0.159
Students (7)	**0.947**	0.597	0.842	0.766	0.491
Tourists (8)	**0.984**	0.677	0.777	0.691	0.429
Outgoing phone calls	**0.932**	**0.927**	0.446	0.340	0.065
Printed publications	0.891	0.449	0.898	0.824	0.577
Patents (11)	0.720	0.421	0.737	0.692	0.431
Geographical distance	**0.942**	0.740	0.678	0.609	0.411

Source: Author calculations.

TABLE B.1.(CONT.)— MATRIX OF CORRELATIONS BETWEEN THE COMPOSITE DISTANCES ACCORDING TO THE COEFFICIENT ESTIMATES OBTAINED FOR DIFFERENT KIND OF FLOWS AND GEOGRAPHICAL DISTANCE (OLS ESTIMATES).

	(6)	(7)	(8)	(9)	(10)	(11)
(Panel B						
Emigration (6)	-					
Students (7)	0.89	-				
Tourists (8)	**0.93**	**0.97**	-			
Outgoing phone calls	**0.98**	0.83	0.88	-		
Printed publications	0.80	**0.95**	**0.94**	0.71	-	
Patents (11)	0.70	0.85	0.75	0.66	0.75	-
Geographical distance	0.89	0.88	**0.93**	0.87	0.81	0.64

Source: Author calculations.

TABLE B.2. –PERCENTAGE OF VARIANCE EXPLAINED BY GRAVITY VARIABLES.

	C1[a]	C2[b]	C3[c]	C4[d]=[C2-C1]/[1-C1]	C5[e]=[C3-C1]/[1-C1]
Goods exports	0.470	0.885	0.923	78%	85%
Services exports	0.708	0.851	0.884	49%	60%
FDI outward stocks	0.714	0.728	0.767	5%	19%
Portfolio equity	0.639	0.750	0.807	31%	47%
Portfolio long term	0.729	0.853	0.898	46%	62%
Intl. Tertiary	0.690	0.811	0.843	39%	49%
Emigration	0.460	0.853	0.875	73%	77%
Intl. Tourists’	0.299	0.858	0.899	80%	86%
Outgoing phone	0.635	0.879	0.907	67%	75%
Printed publication	0.286	0.725	0.757	61%	66%
Patenting activity	0.947	0.959	0.972	23%	47%

Source: Author calculations.

[a] R-squared of a gravity model with size variables (gdp/population) and fixed effects.

[b] R-squared of a gravity model with geographic distance, size variables (gdp/population) and fixed effects.

[c] R-squared of a gravity model with standard set of variables and fixed effects.

[d] % of the remaining variance explained by geographical distance.

[e] % of the remaining variance explained by all distance variables.

TABLE B.3.— MATRIX OF CORRELATIONS BETWEEN THE COMPOSITE DISTANCES ACCORDING TO THE COEFFICIENT ESTIMATES OBTAINED FOR DIFFERENT KIND OF FLOWS AND GEOGRAPHICAL DISTANCE (PPML ESTIMATES).

	(1)	(2)	(3)	(4)	(5)
Panel A					
Goods exports (1)	-				
Services exports (2)	**0.985**	-			
FDI outward stocks	0.470	0.484	-		
Portfolio equity assets	0.839	**0.902**	0.460	-	
Portfolio long term	0.450	0.499	**0.936**	0.587	-
Emigration (6)	**0.921**	**0.919**	0.399	0.827	0.372
Students (7)	0.895	**0.918**	0.400	0.882	0.416
Tourists (8)	**0.905**	0.874	0.225	0.741	0.185
Outgoing phone calls	0.820	0.826	0.042	0.771	0.063
Printed publications	**0.915**	**0.938**	0.420	**0.901**	0.436
Patents (11)	0.580	0.657	-0.085	0.763	0.027
Geographical distance	**0.959**	**0.924**	0.431	0.790	0.378

Source: Author calculations.

TABLE B.3. (CONT.)— MATRIX OF CORRELATIONS BETWEEN THE COMPOSITE DISTANCES ACCORDING TO THE COEFFICIENT ESTIMATES OBTAINED FOR DIFFERENT KIND OF FLOWS AND GEOGRAPHICAL DISTANCE (PPML ESTIMATES).

	(6)	(7)	(8)	(9)	(10)	(11)
(Panel B						
Emigration (6)	-					
Students (7)	**0.98**	-				
Tourists (8)	**0.94**	0.89	-			
Outgoing phone calls	0.89	0.89	**0.93**	-		
Printed publications	**0.98**	**0.99**	**0.90**	0.89	-	
Patents (11)	0.67	0.74	0.64	0.83	0.74	-
Geographical distance	**0.94**	0.89	**0.95**	0.84	**0.91**	0.55

Source: Author calculations.