CrossMark

Foreign languages and trade: evidence from a natural experiment

Jan Fidrmuc[1,2,3] · **Jarko Fidrmuc**[2,3,4,5,6]

Abstract Cultural factors and common languages are well-known determinants of trade. By contrast, the knowledge of foreign languages was not explored in the literature so far. We combine traditional gravity models with data on fluency in the main languages used in EU and candidate countries. We show that widespread knowledge of languages is an important determinant for foreign trade, with English playing an especially important role. The robustness of our results is confirmed by quantile regressions.

✉ Jan Fidrmuc
  Jan.Fidrmuc@brunel.ac.uk; jan@fidrmuc.net
  http://www.fidrmuc.net/

✉ Jarko Fidrmuc
  jarko.fidrmuc@zu.de; jarko.fidrmuc@wifo.ac.at

1 Department of Economics and Finance, Centre for Economic Development and Institutions (CEDI), Brunel University, UB8 3PH London, UK

2 Institute of Economic Studies, Charles University, Prague, Czech Republic

3 CESifo Munich, Munich, Germany

4 Zeppelin University Friedrichshafen, Am Seemooser Horn 20, 88045 Friedrichshafen, Germany

5 School of Economics, Henan University, Kaifeng, China

6 Austrian Institute of Economic Research (WIFO), Vienna, Austria

🖉 Springer

1 Introduction

Speaking the same language facilitates communication and makes transactions easier and more transparent. In this, the effect of language is similar to that of common culture, legal norms or units of measurement: Engaging in mutually beneficial exchange is possible without them, but it is generally more costly and the outcome is less predictable. The additional complexity inherent in transactions without a common language and the increased potential for errors and misunderstandings imply an increase in costs that may be large enough to prevent mutually beneficial transactions from occurring. Consequently, the ability to speak a foreign language should translate into positive individual economic payoffs. These gains will be embodied in better employment opportunities and higher wages, in addition to nonpecuniary benefits such as being able to visit foreign countries, meet new people and read foreign books or newspapers. The previous literature has found such individual gains to be potentially large.[1]

In this paper, however, we are interested in the economic returns to proficiency in foreign languages at the aggregate level rather than at the individual level. If enough people in two countries speak the same language, they will be able to communicate with each other more readily. Consequently, trade between these two countries will be easier, cheaper and more intensive. Hence, we should expect languages to foster bilateral trade. This observation, of course, is not new. In fact, most studies using the gravity model to analyze trade account for common official languages between countries (e.g., French is the official language of France, Belgium, Luxembourg, Switzerland, Canada and many former French and Belgian colonies). Such studies invariably find that sharing a common official language increases trade intensity. However, languages do not need to be formally recognized as official languages in both countries in order to foster trade. International commerce is increasingly conducted in English even if neither party of the transaction is coming from an English-speaking country.

While most gravity model analyses consider only official languages, Melitz (2008) goes a step further in considering also all indigenous or established languages spoken in a country. Furthermore, he also accounts for the fraction of the population speaking these languages. English, for example, is spoken in many of the former British colonies but often only a small fraction of the population speaks it. Chinese, similarly, is spoken in a number of South Asian countries (e.g., Singapore, Malaysia, Indonesia and Philippines) even though it does not have the status of official language in all of these nations. However, a crucial limitation of his data is that it only includes languages

---

[1] Most empirical studies focus on immigrants (e.g., Chiswick and Miller 2002, 2007; Grenier and Nadeau 2013) in whose case a positive return to the ability to speak the host-country language is to be expected. Ginsburgh and Prieto-Rodriguez (2006) estimate the returns to using a foreign language at work for native Europeans and find that the return depends on the relative *scarcity* of the foreign language (for instance, English has a much lower return in Denmark than in Spain).

that are indigenous or otherwise established in the country. Specifically, the *Ethnologue* database[2] used by Méliz collects information only on the languages spoken by primary speakers, which can be identified with native or established (ethnic minority) populations of each country (including those spoken by people who are bilingual or multilingual). The database, however, omits the languages spoken by secondary speakers who learn foreign languages not readily spoken in their own country. These abilities often facilitate economic interactions and trade especially. For example, trade relations between Greek and Swedish firms are most likely facilitated by the ability to speak English rather than speaking Greek or Swedish.

In contrast to Melitz, we consider not only native but also secondary speakers. We utilize a new and previously little used survey data set on the knowledge of languages in the member and candidate countries of the European Union. Importantly, the data contain detailed information on the respondents' native languages and also on up to three foreign languages that they can speak. These surveys are nationally representative, and therefore, they allow us to construct probabilities that two randomly chosen individuals from two different countries will be able to communicate with each other. We use such *communicative probabilities* to investigate the effect of languages on bilateral trade flows in Europe.

We find that greater density of linguistic skills actually translates into greater trade intensity. In the relatively homogenous sample of 15 EU countries, the average probability that two randomly chosen individuals from two different countries will be able to communicate in English with each other is 22 % (considering both native English speakers and those who speak English as a foreign language). This raises intra-EU15 trade, on average, by approximately one quarter. The effect of languages on trade is slightly weaker, but still strongly significant when we include all 29 member and candidate countries in the analysis. English plays a particularly important role while German, French and Russian, in contrast, produce weaker and more mixed results.

Causality between trade and language proficiency can, in principle, go either way. Countries whose residents can communicate easily are likely to trade more with each other, but residents of countries that trade a lot have also an incentive to learn each other's language. Given the wide-ranging separation between Eastern and Western Europe during the Cold War, trade between these two regions is unlikely to be subject to such an endogeneity. Correspondingly, the level of fluency in Western European languages is significantly lower in Eastern Europe than in Western Europe (while the opposite is true for the knowledge of Russian). This creates a situation akin to a natural experiment, which we use to test the robustness of our findings. We observe nearly the same effect of languages on trade when considering only trade flows between Eastern and Western Europe as in the unrestricted data set. Given that suitable instruments for language proficiency are difficult to identify, we find this last result particularly reassuring.

In the following section, we discuss briefly the available literature on the effect of languages on international trade. In Sect. 3, we introduce our data. Section 4 contains the empirical analysis. Sections 5 and 6 present sensitivity analysis using trade between

---

[2] See http://www.ethnologue.com/.

Eastern and Western Europe and median/quantile regressions, respectively. The final section summarizes and discusses our findings.

2 Languages and trade

The gravity model (see Linder 1961; Linnemann 1966; Anderson and van Wincoop 2003; Helpman et al. 2009) relates trade between two countries to their aggregate supply and demand, transport and transaction costs and specific bilateral factors (e.g., free trade agreements) between them. It has proven to be an extremely popular tool for applied trade analysis. Models based on the gravity relation have been used to assess the impact of trade liberalization and economic integration, to discuss the so-called 'home bias' in trade (McCallum 1995) and to estimate the effects of currency unions on trade (Rose 2000).

Accounting for common official languages is a standard feature of gravity models. To this effect, the basic gravity equation is typically augmented to include a common-language dummy, alongside the other potential determinants of bilateral trade such as common border, landlocked dummy and indicators of shared colonial heritage.[3] Most studies, however, pay little attention to the effect of languages that they estimate. Rather, they account for common languages primarily to help disentangle their effect from the one related to preferential trade liberalization. For instance, several European languages have official status in two or more EU countries: English (UK, Ireland and Malta), German (Austria, Germany and Luxembourg), French (France, Belgium and Luxembourg), Dutch (Belgium and Netherlands), Swedish (Sweden and Finland) and Greek (Greece and Cyprus).[4] It is natural to expect that having the same official language fosters bilateral trade. Therefore, failure to account for the common-language effect would likely result in an upward-biased estimate of the trade effect of economic integration in the EU.

Cultural factors, which may promote a more efficient communication between countries, are often positively correlated with trade. Felbermayr and Toubal (2010) find that a measure of cultural proximity based on voting in the Eurovision Song Contest increases bilateral trade, especially trade in differentiated products. Using transaction data and the regression discontinuity design methodology, Egger and Lassmann (2013) find high trade effects of different native languages in the Swiss cantons. Some studies, such as Rauch and Trindade (2002), find that ethnic minorities help foster trade links between their current country of residence and the ancestral country.

While most studies do not specifically discuss the language effects, these are generally found to be highly important. Frankel and Rose (1998) find that two countries that share the same official language tend to have 1.8 times higher trade than two otherwise similar countries without a common language, an effect that is similar in magnitude to having a common border. Melitz and Toubal (2014) confirm this. Their estimated coefficient of 0.3–0.5 implies that two countries that share the same official language tend to have 1.3–1.6 times higher trade than two otherwise similar countries

---

[3] More recent studies often include these factors as fixed effects.

[4] In addition, Turkish is the official language of both Turkey and Cyprus.

without a common language. There have been several attempts to estimate the impact of language barriers in more detail. Anderson and van Wincoop (2004) report a tax equivalent of language barriers of about 7 %[5] while other information-related costs correspond to a tax equivalent of 6 %. This comes closer to the effect of tariff and nontariff barriers, which are estimated at a similar level of 8 %. The summary effect of all border-related trade barriers is estimated as equivalent to a 44 % tax. Ipshording and Otten (2013) go one step further, and instead of considering common official languages, they look at the linguistic distances in the context of a gravity model. They find that countries with similar languages trade significantly more with each other. However, the effect is relatively modest considering that the shift from the 25th to the 75th percentile of linguistic distance is associated with only a 4 % increase in trade on average.

The new trade theory with heterogeneous firms shed more light on the role of language-related costs in trade. Helpman et al. (2009) distinguish between extensive and intensive margin of trade. Their empirical results indicate that common languages are an important part of fixed costs related to market entry, thus influencing mainly the extensive margin of trade. In particular, common language between two countries increases the probability of bilateral trade by 10 %.

To the best of our knowledge, only a few studies focus specifically on the relationship between bilateral trade and languages: Hutchinson (2002), Melitz (2008), Fidrmuc and Fidrmuc (2009) and Melitz and Toubal (2014). Hutchinson considers the role of English in trade relations of a number of countries with the USA. Melitz (2008) goes beyond official languages and instead considers all indigenous or established languages spoken by at least 4 % of the population, in addition to official languages.[6] He finds that both categories of languages, which he labels as 'open-circuit' and 'direct communication'[7] languages, respectively, increase bilateral trade. Similarly, Fidrmuc and Fidrmuc (2009) and Melitz and Toubal (2014) find that the impact of common official languages is positive but smaller than the impact of common spoken languages.

As Melitz (2008) only considers indigenous or established languages, he fails to measure the effect of foreign languages.[8] Especially in Europe, the knowledge of foreign languages is widespread and such nonindigenous languages are likely to play an important role in facilitating trade and economic relations in general. So far, only Fidrmuc and Fidrmuc (2009) and Melitz and Toubal (2014) used the proficiency in foreign languages as a determinant of trade.

---

[5] Melitz (2008) presents somewhat higher estimates of language-related costs between 18 and 32 %.

[6] His analysis is based on the Ethnologue database (see http://www.ethnologue.com/), complemented by the CIA World Factbook.

[7] Open-circuit languages are those that either have official status or are spoken by at least 20 % of the population in both countries. Direct-communication languages are those that are spoken by at least 4 % of the population in each country. The former are measured using dummy variables, the latter as the probability that two randomly chosen individuals from either country can communicate directly in any direct-communication language.

[8] We make the distinction between indigenous and foreign languages. The former are those that are either native or widely spoken by the natives in a given country (as is the case, for example, with Swahili or Arabic in many countries of East and North Africa, respectively). Foreign languages, in contrast, are rarely spoken when two natives of the same country meet, even if both are proficient in the same foreign language.

3 Data

An important strength of our analysis is that we have detailed information on languages spoken in 29 European countries, including both native and foreign languages. We follow Fidrmuc and Fidrmuc (2009) and Melitz and Toubal (2014) in drawing upon a Eurobarometer survey[9] covering all member states and candidate countries of the European Union. The respondents[10] were asked to list their native languages, allowing multiple entries, and up to three other languages that they 'speak well enough in order to be able to have a conversation.' Additionally, the respondents were asked to rate their skills in each of these languages as basic, good or very good. In our analysis, we drop those with basic proficiency and include those who speak each language well, very well or as native speakers. The survey is nationally representative and therefore can be used to estimate the share of each country's population that speak each language. The languages included are all EU official languages, regional languages of Spain (Catalan, Basque and Galician), and selected non-EU languages (Arabic, Russian, Chinese, Hindi, Urdu, Gujarati, Bengali and Punjabi).

The trade data report bilateral trade flows among the 29 countries between 2001 and 2007. Choosing this period ensures that our estimates are not influenced by major events such as the transformational recessions afflicting the formerly communist countries during much of the 1990s, the recent entry of some of these countries to the Eurozone or the financial crisis and the associated trade collapse of 2008–2009 (see Levchenko et al. 2010; Eaton et al. 2011). The data were compiled from the IMF Direction of Trade Statistics and are expressed in US dollars, converted to euros at the current exchange rates. We furthermore use nominal GDP data, based on the IMF International Financial Statistics, converted to euros as well, and the distance between countries measured in terms of great circle distances between capital cities.

The figures on language skills are interesting in their own right. English is the language spoken by the largest number of Europeans: 33 % declare it as their native language or speak it well or very well (Fig. 1). Seven EU countries (Cyprus, Denmark, Malta, Netherlands, Sweden as well as Ireland and the UK) show that the majority of their populations are proficient in English, and only two countries (Hungary and Turkey) have proficiency rates below 10 %. German is spoken by 22 %, French by 17 % and Russian by 4 % (Fig. 2 through Fig. 4).[11] Unlike English, these three languages are mainly spoken in their native countries or (in case of Russian) in countries that have large minorities of native speakers. Note that no language attains 100 % proficiency rate in any single country, not even in the countries where it is native. This mainly relates to the fact that immigrants and/or minorities do not possess sufficiently good linguistic skills in the host-country language.

---

[9] Special Eurobarometer 243 (EB64.3), Europeans and their languages, European Commission, which was carried out in the late 2005. See http://ec.europa.eu/public_opinion/archives/ebs/ebs_243_sum_en.pdf for detailed information.

[10] The survey included only citizens of an EU member or candidate country, although the respondents are not necessarily nationals of the country in which they were interviewed.

[11] The shares of those speaking Italian, Spanish and Polish are 12, 10 and 7 %, respectively.

Fig. 1 Proficiency in English (native, very good or good proficiency)
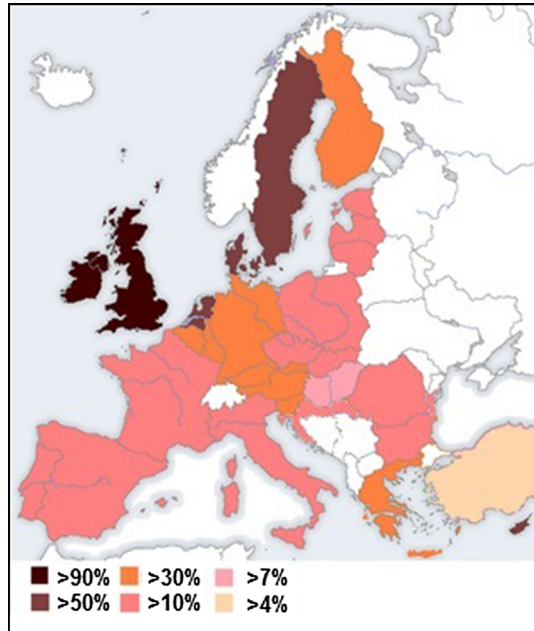
Fig. 2 Proficiency in French (native, very good or good proficiency)
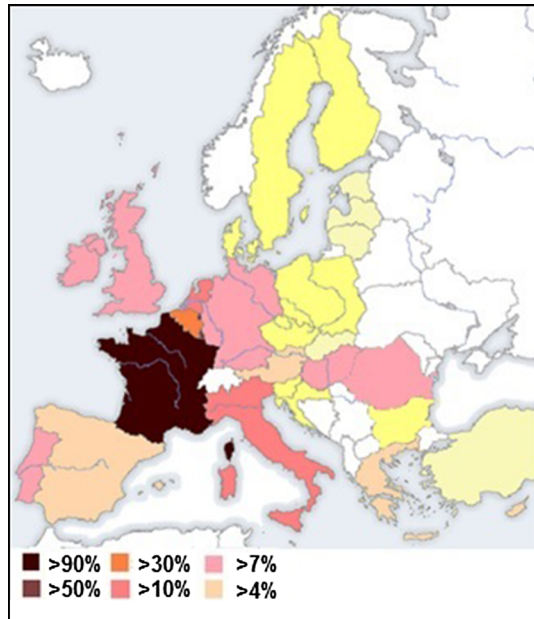
Rather than using the proficiency rates alone, we estimate the probability $P_{f,ij}$ that two randomly chosen individuals from countries i and j will be able to communicate in a language or a set of languages f as the product of the average proficiency rates, ω_{fi} and ω_{fj}, in the two countries (see, e.g., Alesina et al. 2003; Melitz 2008):

Fig. 3 Proficiency in German (native, very good or good proficiency)

Fig. 4 Proficiency in Russian (native, very good or good proficiency)

$$P_{fij} = \omega_{fi}\omega_{fj}. \tag{1}$$

In so doing, we make no distinction between those who are native speakers of the language considered and those who speak it as a foreign language (except that we

require that the respondent's self-assessed proficiency, if not native, is good or very good).

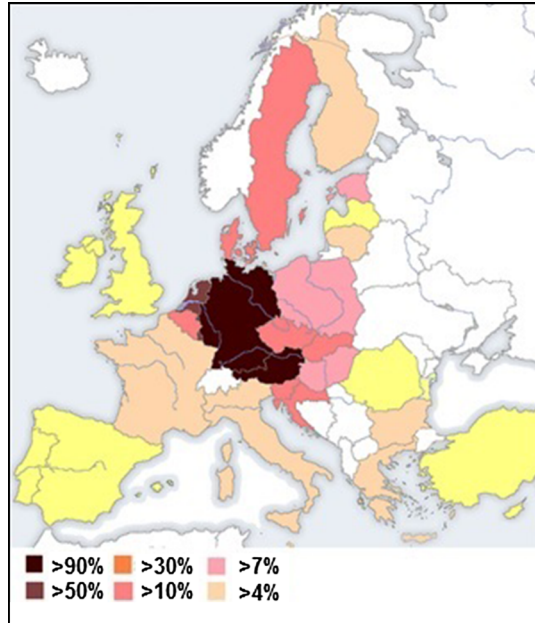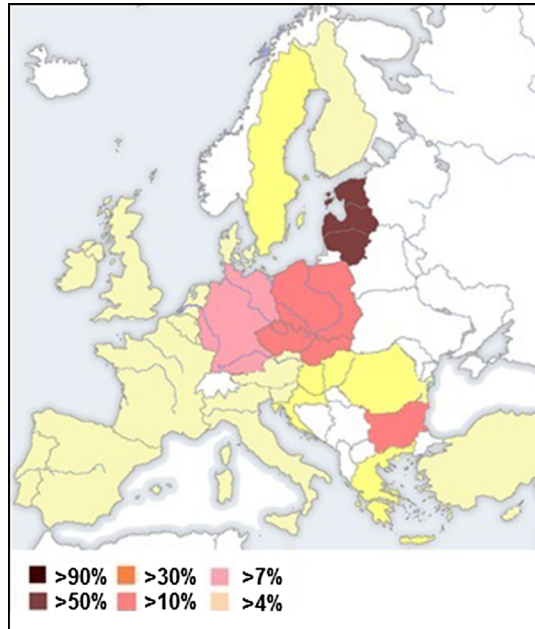Our data contain information on proficiency in 32 languages. However, it is obvious that only a relatively small subset of them can realistically serve as conduits of inter-country communication. We impose the requirement that conduit languages should be spoken by at least 10 % of the population in at least three different countries. There are four such languages: English, German, French and Russian, the last being spoken mainly in the new member countries and also in Germany (8 % of population). Note that this relatively strict definition leaves out Italian, which, outside of Italy, is spoken by 3–5 % of Austrians, Belgians, French and Luxembourgers and 7–9 % of Croats and Slovenes. Similarly, Spanish, although spoken widely outside the EU, has relatively small linguistic constituencies in Europe—between 2–7 % of Austria, Denmark, France, Germany, Netherlands and Portugal—and therefore, it is not included. Lowering the threshold to 4 % would add these two languages and also Swedish (spoken by 8 % of Danes and 20 % of Finns) and Hungarian (spoken by 7 % of Romanians and 16 % of Slovaks).[12]

English is clearly the most likely conduit for inter-country communication (Melitz 2014): The average communicative probability for the 29 countries is 17 % (22 % for the EU15). Even excluding Ireland and the UK, this probability remains very high, 15 %. In several cases, the probability that English may serve as the language of communication exceeds 50 % (e.g., for Netherlands–Sweden and Netherlands–Denmark). In turn, there are only few bilateral pairs which display probabilities below 10 %: most of these are countries with Romance languages.

German and French lag far behind English, with 5 and 3 % average communicative probabilities, respectively (or 7 and 5 % in the EU15). Nevertheless, there are some cases where the communicative probabilities are relatively high: For example, the probability that a Dutchman and a Dane will be able to speak German with one another is 16 %. For all remaining languages, the average communicative probability is essentially zero, although it is often nonnegligible for specific pairs of countries.[13]

Finally, we compute a cumulative communicative probability that considers those who speak English, French or German as the three most widely spoken languages. Constructing such a probability over a set of languages is not trivial: Adding up the respective probabilities would result in some pairs of countries with overall communicative probability exceeding 1, as some individuals speak two, three or even more languages at the same time. We take care therefore that the speakers of each language are counted only once.

---

[12] Results obtained with these languages are available upon request.

[13] The less obvious examples include Russian between Germany and Bulgaria (2 %), Polish between Poland and Lithuania (13 %), Hungarian for Slovakia and Romania (1 %), Italian in case of Malta and Slovenia (3 %), Czech and Slovak between the Czech and Slovak Republics (22 % for Czech and 16 % for Slovak), and Swedish in the case of Finland and Denmark (1 %).

4 Gravity model with languages

We estimate the following gravity equation,

$$T_{ijt} = \beta_1 \left(Y_{it} + Y_{jt} \right) + \beta_2 D_{ij} + \beta_3 B_{ij} + \beta_4 F_{ij} + \beta_5 EU_{ij} + \beta_6 \text{EMU}_{ij}$$
$$+ \sum_{f=1}^{F} \delta_f P_{fij} + \theta_{it} + \theta_{jt} + \varepsilon_{ijt}, \tag{2}$$

where T_{ijt} corresponds to the size of bilateral trade (in logs) between country i and country j at time t, Y_{it} and Y_{jt} stand for the log of nominal GDP in countries i and j at time t, and D_{ij} is the log of distance between them proxying for transport costs. The income elasticity of foreign trade, β_1, is expected to be positive, while the transport cost elasticity, β_2, should be negative. We also include dummy variables for geographical adjacency, B, for the former federations in Eastern Europe, F (these are Czechoslovakia, Yugoslavia and the Soviet Union), as well as for EU and EMU membership. These variables are all expected to have positive effects on trade. P_{fij} is the probability that two random individuals, one from country i and one from j, can communicate in language f. The construction of this communicative probability is discussed in detail below. Finally, θ_{it} and θ_{jt} denote time-specific country effects, and ε is the residual term. Importantly, the time-specific country effects account for any country-specific time-invariant and time-varying heterogeneity, including history, institutions, and culture. Not accounting for these unobservable factors would otherwise bias our results (see Baltagi et al. 2003; Baldwin and Taglioni 2006).

We take heed of Baldwin's and Taglioni's (2006) critique of common approaches to estimating the gravity model. Firstly, we define trade volume as the average of the logs of exports and imports, and not as the log of the average of exports and imports. This precludes a possible bias that would occur if trade flows are systematically unbalanced, which is commonly observed between countries of the European Union. Secondly, we include trade flows and GDP in nominal terms (but converted to euros using contemporaneous exchange rates). This reflects the fact that gravity models can be derived from expenditure functions of consumers (see the discussion of the so-called gold medal error in Baldwin and Taglioni 2006). Thirdly, we include time-varying country effects, as discussed above.[14]

In addition to the standard core variables of gravity models, we control for the ease of communication between countries. In particular, we include communicative probabilities for English, French, German and Russian (constructed as explained in Sect. 3). These measure the probability that two randomly chosen inhabitants of country i and j can communicate in a specific language. Importantly, in computing the probabilities, we make no distinction as to whether the individuals are native speakers of the language or whether one or both of them speak it as a foreign language. Clearly, language can facilitate trade even when one or both parties to the transaction speak an acquired

---

[14] Note that effects for country groups such as free trade areas, contingency, and monetary unions are not covered by time-varying country effects. Alternative specifications with simple time-invariant country effects (Mátyás 1997, 1998) or as a standard OLS, which are also popular in the literature, are available upon request.

rather than their native language. The communicative probability is thus a better indicator of communication costs than language dummies used in the previous literature, which typically only account for official languages. Moreover, the communicative probability reflects actual language proficiencies, as opposed to looking only at official languages: In countries with sizeable ethnic minority or foreign-born populations, a nonnegligible share of inhabitants is not proficient in the official language.

We start with an analysis of trade flows among the EU15 countries because they constitute a relatively homogenous group of countries with regard to their economic, historical and cultural characteristics. Thus, our approach is similar to that of Melitz and Toubal (2014), who discuss the difference between official and spoken languages: If language proficiency is significant in the homogenous country sample, then communication abilities work beyond history, culture and trust.

Columns (1) through (3) of Table 1 present the results obtained with various alternative ways of controlling for bilateral language relations between countries. The standard gravity model variables (in the top part of the table) are all significant and have the expected signs. Trade increases with the economic size of countries and falls with distance. Sharing a common border reduces transaction costs and correspondingly increases trade. Those EU countries that use the euro trade over 1.5 times more with each other than with otherwise similar countries outside the Eurozone. This is similar to estimates currently available in the literature (see Baldwin 2006, for a literature survey).

A traditional formulation of the gravity model would feature official language dummies. We replace these with communicative probabilities to fully account for the effect of languages, whether native or foreign. In this way, our specification allows us to observe how languages affect trade also between countries in which they do not have an official status, as long as they are sufficiently widely spoken. Column (1) accounts only for communicative probability in English. The ability to communicate in English has a positive and strongly significant effect on trade. To quantify this effect, one has to take account of the communicative probability. For example, the communicative probability for the UK and Ireland is 0.97, which translates into a 2.9-fold increase in trade over what can be ascribed only to economic factors and geography. Proficiency in English also affects trade between other countries: For example, it increases trade between the Netherlands and Sweden by three quarters, while Dutch trade with the UK is more than doubled. With the average English communicative probability being 22 % in the EU15, the ability to communicate in English increases trade by approximately one quarter.

In column (2), we add communicative probabilities in French and German, and in column (3), we replace individual languages with the cumulative communicative probability that considers all three languages simultaneously. The English communicative probability remains significant also after controlling for other languages. Of these, only German appears to foster trade, but its coefficient estimate is much smaller than that for English. However, again, when interpreting the point estimates, one must bear in mind the relative strength of the various languages: the average communicative probability is substantially higher for English (22 %) than for German (7 %, respectively). Therefore, on average, proficiency in German raises trade by approximately 5 % (based on the estimates in column 2).

Table 1 Trade effects of foreign languages, EU15 and EU29

Variable	(1) EU15	(2) EU15	(3) EU15	(4) EU29	(5) EU29	(6) EU29
Intercept	15.705***	15.568***	15.659***	18.027***	17.981***	18.016***
	(46.069)	(43.873)	(44.080)	(101.297)	(100.013)	(100.553)
GDP	0.915***	0.920***	0.923***	0.916***	0.924***	0.916***
	(42.706)	(42.348)	(42.525)	(79.985)	(80.472)	(79.903)
Distance	−0.767***	−0.757***	−0.759***	−1.047***	−1.051***	−1.043***
	(−29.601)	(−27.774)	(−27.323)	(−50.708)	(−50.409)	(−50.162)
Contiguity	0.541***	0.519***	0.466***	0.384***	0.405***	0.361***
	(16.761)	(15.568)	(13.014)	(9.902)	(10.231)	(9.190)
Former fed.				2.401***	1.880***	2.426***
				(26.099)	(15.301)	(26.252)
EU				0.038	0.031	0.058
				(0.813)	(0.664)	(1.236)
EMU	0.460***	0.454***	0.408***	0.188***	0.193***	0.165***
	(9.224)	(9.027)	(7.787)	(4.925)	(4.935)	(4.292)
English	1.078***	1.101***		0.653***	0.622***	
	(9.349)	(9.476)		(5.523)	(5.280)	
French		0.048			−0.479***	
		(0.445)			(−2.759)	
German		0.241***			0.051	
		(3.654)			(0.411)	
Russian					1.642***	
					(6.159)	
Cumulative[a]			0.573***			0.336***
			(7.741)			(3.739)
N	1470	1470	1470	5634	5634	5634
Adj. R^2	0.968	0.968	0.967	0.926	0.926	0.925

[a]Cumulative probability that two inhabitants of the country pair can communicate in English, French or German (reflecting knowledge of two or all three languages). Time-varying country effects are not reported. t-statistics based on bootstrapped standard errors are given in parentheses
***, ** and * significance at 1, 5 and 10 %, respectively

However, the interpretation of the effects of different languages is not straight-forward, if the control group is not clearly defined. Intuitively, the control group for assessing each foreign language should be the hypothetical situation where no foreign language is available, which is never the case. Especially when considering the less widely spoken languages such as German and French, the control group explicitly includes those able to communicate in English (and any other foreign languages). Therefore, we include the cumulative probability for English, French and German, which is compared to population not being able to communicate in any major language. The effect of this cumulative probability is also strongly significant and positive. The coefficient estimate is approximately half that for English.

The results obtained with the wider data set covering the whole of the EU29 are broadly similar, despite some noticeable differences (columns 4–6 of Table 1). Besides English, French and German, the analysis now also includes Russian. We also add a dummy variable for countries which emerged from the breakup of the former federations in Eastern Europe (Czechoslovakia, Yugoslavia and the USSR) and a dummy for membership in the EU (to distinguish member states from candidates). The English communicative probability again has a strongly positive effect on trade. The coefficient estimate is lower than that obtained for the EU15, which is not entirely surprising given the much lower levels of English proficiency in the new member and the candidate countries. Among the remaining languages, only Russian appears to have a significantly positive effect on trade: Besides capturing the effect of proficiency in Russian, this may also reflect the legacy of greater economic cooperation among the former Soviet Bloc countries. Somewhat surprisingly, communicative probability in French seems to have a significantly negative effect on trade in this sample. This surprising result may be due to heterogeneity[15] in the broad EU29 country sample, which is addressed in our discussion of the natural experiment in language education in Eastern Europe in the next section. As stated above, when interpreting this coefficient, one must also take into account that the relevant control group includes those able to communicate in English, German, Russian and any other foreign language. Reassuringly, the cumulative communicative probability remains significant and positive.

5 Natural experiment of East–West political differences in language education

A potential problem with the preceding results may arise due to the fact that bilateral trade intensity and the knowledge of foreign languages may be endogenous. People have an incentive to learn languages which they can subsequently use in their job, business or social life. For example, only a negligible fraction of the European population speaks Latin despite many cultural, academic and historical reasons to learn it. Furthermore, knowledge of languages which are not used frequently is likely to diminish after some time. Thus, the share of the population with a good or very good proficiency in Russian in the new member states now stands at between 10 and 20 % (and is only 1.4 % in Hungary), despite the long tradition of obligatory and rather extensive teaching of Russian in the former communist countries. Therefore, although we find evidence of a positive correlation between language proficiency and trade flows, we cannot convincingly interpret this correlation as a causal effect of languages on trade.

The standard solution for removing the endogeneity bias is to use instrumental variables. Finding suitable and valid instruments, however, is a notoriously difficult task.[16] An alternative possibility is to find a suitable natural experiment. This approach has become widely used in economics (Wolpin and Rosenzweig 2000; Angrist and

---

[15] In particular, language skills in French are practically negligible in Eastern Europe.

[16] We used instrumental variables in an earlier version of this paper, with similar, though somewhat mixed, results. These are available upon request.

Pischke 2010). The political foundations of the communist countries' education systems represent such an experiment, because they created a long-term divergence in language skills between Eastern and Western Europe. In other words, while in the west, knowledge of English is widespread in part because of economic and other benefits that this brings at the individual level, this is much less the case in the east (where Russian played a similar role until the end of the 1980s). Therefore, we use the variation in foreign language skills between Western and Eastern European countries to analyze the impact of language skills on trade. Figures 1, 2, 3 and 4 show that the level of fluency in English in Eastern Europe is significantly lower than in North Western Europe and about comparable to South Western Europe. Similarly, the level of fluency in French is negligible in all Eastern European countries with the exception of Romania. By contrast, Eastern European countries show relatively good language skills in German, but these are still lower than those in North Western Europe. Finally, only the fluency in Russian is higher than in other European regions. The inspection of detailed data on language fluency for different age groups show that the adjustment of language skills is very slow and significant differences in language skills are likely to persist during the next decades.

To this effect, we restrict the sample to country pairs consisting of one Western and one Eastern European country. While this is a highly heterogeneous sample, imposing this restriction ensures that language skills are not correlated with the other possible determinants of trade, including geographical and cultural factors. Correspondingly, we can estimate equation (2') by standard OLS

$$T_{ijt} = \beta_1 \left(Y_{it} + Y_{jt} \right) + \beta_2 D_{ij} + \beta_3 B_{ij} + \sum_{f}^{F} \delta_f P_{fij} + \theta_{it} + \theta_{jt} + \varepsilon_{ijt}, \qquad (2')$$

where all variables (we exclude dummies for former federations EU, and EMU, which are not applicable for this subsample) and parameters are defined as above.

Table 2 presents the results. The first column presents the estimates of the core gravity model excluding the language variables. The results show that income and distance elasticities are very close to the previous estimates for the EU15 sample and to those presented in the literature (Baldwin and Taglioni 2014). Further columns include language proficiencies in English, French, German and Russian. The results confirm the importance of English. Similar to the EU15 sample, the coefficient for English proficiency is close to 1. German is also important. Its effect is even higher than that for English proficiency, but its size of slightly above 2 is not surprisingly high. Finally, French and Russian proficiencies are insignificant, which confirms that these languages are not playing an important role for East–West trade. These effects are confirmed if we include all language variables, as shown in column (6). Finally, the last column presents the results for the overall language proficiency defined in the same way as before (that is, the overall proficiency either in English, German or French). This coefficient is close to 1. Thus, the natural experiment of East–West language education division due to the different political orientation confirms that language skills have an important impact on trade which cannot be attributed to other underlying factors such as cultural or geographical proximity.

Table 2 Trade effects of foreign languages, natural experiment (East–West trade)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Intercept	16.290***	16.155***	16.270***	15.850***	16.287***	15.708***	15.958***
	(25.230)	(24.998)	(24.894)	(22.508)	(24.853)	(21.891)	(24.412)
GDP	0.947***	0.961***	0.948***	0.945***	0.947***	0.958***	0.962***
	(35.496)	(35.203)	(35.800)	(35.772)	(35.522)	(35.415)	(35.604)
Distance	−0.839***	−0.850***	−0.835***	−0.783***	−0.839***	−0.791***	−0.829***
	(−10.860)	(−11.037)	(−10.657)	(−9.161)	(−10.683)	(−9.082)	(−10.815)
Contiguity	0.485***	0.490***	0.488***	0.478***	0.485***	0.488***	0.486***
	(4.242)	(4.236)	(4.238)	(4.293)	(4.217)	(4.261)	(4.240)
English		0.944***				0.928***	
		(3.563)				(3.473)	
French			−2.490			−2.536	
			(−0.838)			(−0.857)	
German				2.164***		2.073***	
				(3.334)		(3.222)	
Russian					0.306	0.916	
					(0.126)	(0.377)	
Cumulative[a]							1.097***
							(4.354)
N	2006	2006	2006	2006	2006	2006	2006
Adjd R^2	0.868	0.869	0.868	0.869	0.868	0.87	0.869

[a]Cumulative probability that two inhabitants of the country pair can communicate in English, French or German (reflecting knowledge of two or all three languages). Time-varying country effects are not reported. t-statistics based on bootstrapped standard errors are given in parentheses
***, ** and * significance at 1, 5 and 10 %, respectively

6 Sensitivity analysis: quantile regression

The previous results may be sensitive to outliers. For example, there may be pairs of countries that have particularly high bilateral trade and high communicative probability in English or another language, such that the gain from foreign languages is overestimated. Or, on the contrary, we may have pairs of countries with relatively low bilateral trade despite high communicative probability, resulting in an underestimated effect of languages. We analyze these factors in this section by means of median and quantile regressions. Median regression is frequently used when standard OLS regression may be biased by outliers. While least squares regression considers the sum of the squared residuals, which gives a lot of weight to outliers, median regression finds the regression line that equates the number of positive and negative residuals. This property makes median regression more robust to influential observations. Koenker and Bassett (1978) generalized this concept to quantile regression, in which selected quantiles of the conditional distribution of the dependent variable are expressed as functions of observed explanatory variables. Koenker and Hallock (2006) argue that inference in quantile regressions is more robust than in ordinary regression. While

Table 3 Trade effects of English proficiency, quantile regression

	OLS	Q10	Q25	Q50	Q75	Q90
GDP	0.902***	0.975***	0.924***	0.898***	0.867***	0.784***
	(100.865)	(64.615)	(79.550)	(65.858)	(58.032)	(51.469)
Distance	−0.792***	−0.762***	−0.798***	−0.782***	−0.786***	−0.760***
	(−27.331)	(−13.174)	(−18.621)	(−24.396)	(−16.407)	(−10.306)
Contiguity	0.678***	0.574***	0.520***	0.706***	0.776***	0.585***
	(14.396)	(6.431)	(7.702)	(8.684)	(9.467)	(6.053)
EMU	0.243***	0.409***	0.280***	0.176***	0.274***	0.359***
	(7.387)	(6.936)	(5.358)	(4.532)	(5.188)	(5.307)
English	0.769***	0.893***	0.598***	0.834***	0.623***	0.985***
	(9.185)	(4.103)	(3.314)	(8.840)	(5.368)	(4.886)
Intercept	15.982***	14.151***	15.438***	15.974***	16.684***	17.771***
	(64.680)	(34.148)	(41.468)	(56.708)	(43.809)	(28.145)
N	1470	1470	1470	1470	1470	1470

Robust t-statistics using bootstrapped (1000 replications) standard errors are given in parentheses

***, ** and * denote significance at 1, 5 and 10 %, respectively

the concept of quantile regression is now frequently used in economics, especially in labor and family economics (see the literature survey by Koenker and Hallock 2006), it has found little application in trade analysis so far (see Wagner 2006).

We estimate the following linear model for the τth conditional quantile, Q, of bilateral trade volume, T,

$$Q_\tau\left(T_{ijt}\right) = \alpha_\tau + \beta_{\tau 1}\left(Y_{it} + Y_{jt}\right) + \beta_{\tau 2}D_{ij} + \beta_{\tau 3}B_{ij} + \beta_{\tau 4}EMU_{ij} + \beta_{\tau 5}P_{\text{eng},ij} + \varepsilon_{ijt}.$$

$$(3)$$

The ease of communication is measured with English proficiency, i.e., based on specification (1) in Table 1. For computational reasons, we are not able to include time-varying country effects.[17] The OLS estimation of equation (3) confirms the robustness of the previous results. Table 3 reports the results for the 10th, 25th, 75th and 90th percentiles in addition to the median regression, while details for each fifth percentile are given in Fig. 5. We can see that the effects of some gravity variables differ considerably between the individual percentiles. The income elasticity declines as bilateral trade increases. In contrast, the transport cost elasticity (proxied by distance) and the effects of geographical contiguity are relatively constant for all quartiles, although distance elasticity is higher for lower quantiles. The EMU has the lowest effect around the median, which indicates that the EMU effect can be influenced by outliers.

The effect of English proficiency is similar to that of transport cost elasticity, which underlines the importance of foreign language proficiency for the reduction in trans-

---

[17] The treatment of the time and cross-sectional variation is relatively complex in quantile estimations for panel data. Canay (2011) discusses a simple approach for estimation of quantile panel estimations, but as he notes, this approach is less appropriate for panels with a short time dimension, as analyzed here. Therefore, these results have to be interpreted carefully.
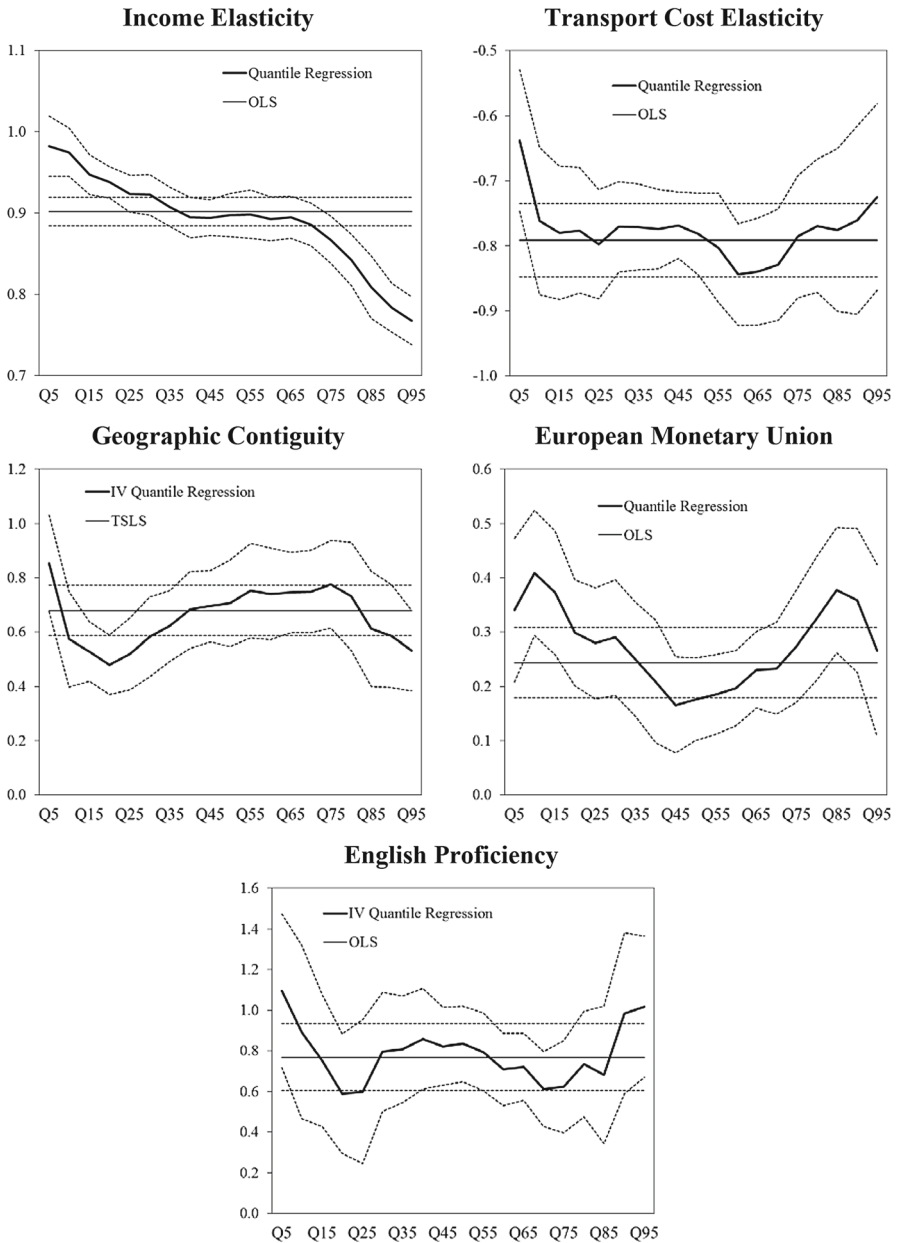
Income Elasticity

Transport Cost Elasticity

Geographic Contiguity

European Monetary Union

English Proficiency

Fig. 5 OLS Regression and quantile regression estimates. *Note*: bootstrapped standard errors are used for 95 % confidence bands

action costs. Figure 5 shows that increasing language proficiency has large significant effects at the very beginning of the scale and at a relatively high level of proficiency. Thus, both the countries with relatively low and high communicative probabilities tend to display a greater return to foreign languages.

7 Conclusions

The fact that language proficiency has a strong impact on trade flows is well understood: Numerous previous papers have found that countries sharing the same official language tend to trade significantly more with each other. We argue that the effect of languages is not limited to official tongues. Clearly, the ability to communicate in a particular language can have an effect on trade flows between two countries as long as it is spoken widely enough in both countries, irrespective of whether it holds an official language status in either or both.

Our findings suggest that English plays an especially important role in facilitating foreign trade. This is not surprising, given that it is the most widely spoken foreign language at present. Our results show that there is a strong positive relationship between bilateral trade and the probability that two randomly chosen individuals from two countries will be able to communicate in English. Of course, it is possible that this positive relationship is due to the endogeneity of language skills. Therefore, we utilize a convenient natural experiment embodied in trade between Eastern and Western Europe. Until the early 1990s, trade between the two parts of Europe was severely restricted due to the Cold War. It is therefore unlikely that broad segments of the population in the east and west possess linguistic skills the acquisition of which was motivated by the economic benefits of east–west trade. The analysis restricted to east–west trade (which thus omits trade flows within the East or West) yields results that are very similar to those obtained with the unrestricted sample of all EU countries. This suggests that the endogeneity bias is unlikely to be very important.

In the past few decades, the prospect of increased trade has become a powerful argument in favor of deepening European integration, although the actual growth of trade has remained much below the initial expectations. Our findings suggest that significant gains could be realized by improving linguistic skills. Moreover, a part of previous trade growth in Europe was possibly not due to European integration policies (e.g., including monetary integration) but a side effect of the increasing foreign language proficiency of the European population. However, more research would be needed to shed light on the different factors of past trade developments.

Foreign language acquisition is not a costless investment, but the gains from foreign language education go beyond its trade effects: Further benefits are likely to accrue in the labor markets, science and education, as well as in the social sphere. Small countries such as the Scandinavian countries with high shares of the population speaking several foreign languages are especially well equipped to benefit from the trade-enhancing effect of languages. Indeed, our results suggest that if all European countries had Scandinavian levels of proficiency in English, trade would be some 30–60 % higher than what can be ascribed to economic and geographical factors.

References

Alesina A, Devleeschauwer A, Easterly W, Kurlat S, Wacziarg R (2003) Fractionalization. J Econ Growth 8:155–194

Anderson JE, van Wincoop E (2003) Gravity with gravitas: a solution to the border puzzle. Am Econ Rev 93:170–192

Anderson JE, van Wincoop E (2004) Trade costs. J Econ Lit 42:691–751

Angrist JD, Pischke J-S (2010) The credibility revolution in empirical economics: how better research design is taking the con out of econometrics. J Econ Perspect 24(2):3–30

Baldwin R (2006) The euro's trade effects. Working Paper No. 594, ECB, Frankfurt

Baldwin R, Taglioni D (2006) Gravity for dummies and dummies for gravity equations. Working Paper No. 12516, NBER, Cambridge, MA

Baldwin R, Taglioni D (2014) Gravity chains: estimating bilateral trade flows when parts and components trade is important. J Bank Financ Econ 2:61–82

Baltagi P, Egger P, Pfaffermayr M (2003) A generalized design for bilateral trade flow models. Econ Lett 80:391–397

Canay IA (2011) A simple approach to quantile regression for panel data. Econom J 14:368–386

Chiswick BR, Miller PW (2002) Immigrant earnings: language skills linguistic concentrations and the business cycle. J Popul Econ 15:31–57

Chiswick BR, Miller PW (2007) Modelling immigrants' language skills, IZA DP 2974, Institute for the Study of Labor (IZA), Bonn

Eaton J, Kortum S, Neiman B, Romalis J (2011) Trade and the global recession. NBER Working Papers 16666, National Bureau of Economic Research

Egger P, Lassmann A (2013) The causal impact of common native language on international trade: evidence from a spatial regression discontinuity design. Discussion Papers 9441, CEPR London

Felbermayr GJ, Toubal F (2010) Cultural proximity and trade. Eur Econ Rev 54:279–293

Fidrmuc J, Fidrmuc J (2009) Foreign languages and trade. Discussion Paper No 7228, C.E.P.R. London

Frankel JA, Rose AK (1998) The endogeneity of optimum currency criteria. Econ J 108:1009–1025

Ginsburgh V, Prieto-Rodriguez J (2006) Returns to foreign languages of native workers in the EU, mimeo

Grenier G, Nadeau S (2013) English as the Lingua Franca and the economic value of other languages: the case of the language of work of immigrants and non-immigrants in the Montreal Labour Market. Venice Summer Institute Workshop, Venice

Helpman E, Melitz MJ, Rubinstein Y (2009) Estimating trade flows: trading partners and trading volumes. Q J Econ 123(2):441–487

Hutchinson W (2002) Does ease of communication increase trade? Commonality of language and bilateral trade. Scott J Polit Econ 49:544–556

Ipshording IE, Otten S (2013) The costs of Babylon: linguistic distance in applied economics. Rev Int Econ 21(2):354–369

Koenker R, Bassett G (1978) Regression quantiles. Econometrica 46(1):33–50

Koenker R, Hallock KF (2006) Quantile regression. J Econ Perspect 15(4):143–156

Levchenko AA, Lewis LT, Tesar LL (2010) The collapse of international trade during the 2008–09 Crisis. In search of the smoking gun. IMF Econ Rev 58:214–253

Linder S (1961) An essay on trade and transformation. Almqvist and Wiksells, Uppsala

Linnemann H (1966) An econometric study of international trade flows. North Holland, Amsterdam

Mátyás L (1997) Proper econometric specification of the gravity model. World Econ 20:363–368

Mátyás L (1998) The gravity model: some econometric considerations. World Econ 21:397–401

McCallum J (1995) National borders matter: Canada–U.S. regional trade patterns. Am Econ Rev 85:615–623

Melitz J (2008) Language and foreign trade. Eur Econ Rev 52(4):667–699

Melitz J (2014) English as a global language. CEPR Discussion Papers No. 10102, C.E.P.R. London

Melitz J, Toubal F (2014) Native language, spoken language, translation and trade. J Int Econ 93:351–363

Rauch JE, Trindade V (2002) Ethnic Chinese networks in international trade. Rev Econ Stat 84(1):116–130

Rose AK (2000) One money, one market: the effect of common currencies on trade. Econ Policy 15(30):7–46

Wagner J (2006) Export intensity and plant characteristics: what can we learn from quantile regression? Review World Econ (Weltwirtschaftliches Archiv) 142(1):195–203

Wolpin KI, Rosenzweig MR (2000) Natural "natural experiments" in economics. J Econ Lit 38(4):827–874