The international platform of Ludwig-Maximilians University's Center for Economic Studies and the Ifo Institute

CESifo Working Papers

www.cesifo.org/wp

Foreign Languages and Trade

Jan Fidrmuc
Jarko Fidrmuc

CESifo Working Paper No. 4670
Category 8: Trade Policy
February 2014

CESifo *Working Paper No. 4670*

Foreign Languages and Trade

Abstract

Cultural factors and especially 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 a natural experiment of trade between Eastern and Western Europe.

JEL-Code: C230, F150, F400, Z100.

Keywords: gravity model, foreign trade, economics of languages, natural experiment, median regression, quantile regression.

*Jan Fidrmuc**
Department of Economics and Finance &
Centre for Economic Development and
Institutions (CEDI); Brunel University
UK – Uxbridge, UB8 3PH
Jan.Fidrmuc@brunel.ac.uk
jan@fidrmuc.net

Jarko Fidrmuc
Zeppelin University
Friedrichshafen / Germany
jarko.fidrmuc@zu.de
jarko.fidrmuc@wifo.ac.at

*corresponding author

January 2014
We appreciate the research assistance by Mareike Heimeshoff. We benefited from comments by Peter Neary, Beata Javorcik, Jacques Mélitz, Oded Galor, Assaf Razin, Marc-Andreas Muendler, Eric Rougier, Jenny Kragl, Michael Landesmann, Javier Ortega, Nicolas Sauter, Agnès Bénassy-Quéré, Mariusz Krawczyk, Yutaka Kurihara, Eiji Ogawa, participants in the European Economic Association conference in Glasgow, CESifo Delphi conference, Royal Economics Society conference in London and ISNIE conference in Berkeley, as well as seminar participants at Brunel University, City University London, WIIW, University of Munich, Hitotsubashi University Tokyo and the School of Slavonic and Eastern European Studies in London.

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 otherwise mutually beneficial transactions from occurring. Consequently, the ability to speak a foreign language should translate into positive individual economic payoffs, embodied in better employment opportunities and higher wages, in addition to non-pecuniary benefits such as the being able to visit foreign countries, study and live abroad, meet new people, read foreign books or newspapers and the like. Indeed, 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, in turn, more intensive. Hence, we should expect languages to foster bilateral trade. This observation, of course, is not new. Indeed, most studies using the gravity model to analyze trade account for common official languages between countries (for example, 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 need not be formally recognized as official languages in

[1] Most empirical studies focus on immigrants (e.g. Chiswick and Miller, 2002 and 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).

both countries in order to foster trade: international commerce is increasingly conducted in English, even if neither party to the transaction is from an English speaking country.

While most gravity-model analyses consider only official languages, Mélitz (2008) goes a step further: he considers also all indigenous or established languages spoken in a country and, furthermore, he accounts also for the fraction of the population speaking them. 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 enjoy an official-language status in all of them. However, a crucial limitation of his data is that it only includes languages that are indigenous or otherwise established in the country. Specifically, the *Ethnologue* database[2] that he uses collects information only on languages spoken by primary speakers, that is native or established (ethnic-minority) populations of each country (including those spoken by people who are bilingual or multilingual). The database, however, omits languages spoken by secondary speakers, that is those who learned them as foreign languages, although such language abilities often facilitate economic interactions and trade especially: these days, trade relations between, for example, Greek and Swedish firms are more likely to be facilitated by English rather than by either Greek or Swedish.

In contrast to Mélitz, we consider not only native but also secondary speakers. We utilize a new and previously little used survey data set on knowledge of languages in the member and candidate countries of the European Union. Importantly, the data contain detailed information not only on the respondents' native languages but 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 indeed translates into greater trade intensity. In the relatively homogenous sample of 15 EU countries, the average

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

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 speakers of English and those who speak it 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. German, French and Russian, in contrast, produce weaker and more mixed results.

Causality between trade and language proficiency can, in principle, go either way: counties whose residents can communicate easily are likely to trade more with each other, but residents of countries that trade a lot also have 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. We utilize this natural experiment to test the robustness of our findings, and observe nearly the same effect of languages on trade as in the unrestricted dataset. 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 section 3, we introduce our data. Section 4 contains the empirical analysis. Sections 5 and 6 present sensitivity analysis using trade between 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, and Helpman et al., 2009), relates trade between two countries to their aggregate supply and aggregate demand, transport and transaction costs and specific bilateral factors (e.g. free trade agreements) between them. It has proven 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 effect of 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 more efficient communication between countries, are often found to be 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 transactions data and the regression discontinuity design methodology, Egger and Lassmann (2013) find high trade effects of different native languages in 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, they are generally found to be highly important. Frankel and Rose (2002) 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. There have been several attempts to estimate the impact of language barriers in more details. Anderson and van Wincoop (2004) report a tax

[3] More recent studies often include these factors as fixed effects.
[4] In addition, Turkish is the official language of both Turkey and Cyprus.

equivalent of language barriers at about 7 percent[5] while other information-related costs correspond to a tax equivalent of 6 percent. This comes close to the effect of tariff and non-tariff barriers which are estimated at a similar level of 8 percent; the summary effect of all border-related trade barriers is estimated as equivalent to a 44 percent tax. Ipshording and Otten (2013) go one step further and instead of common official languages, consider linguistic distance in the context of a gravity model. They find that countries with languages that are more similar trade significantly more with each other, although the effect is relatively modest: moving from 25th to 75th percentile of linguistic distance is associated with trade being higher by only 4 percent 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 per cent.

To the best of our knowledge, only two studies focus specifically on the relationship between bilateral trade and languages: Hutchinson (2002) and Mélitz (2008). Hutchinson considers the role of English in trade relations of selected countries with the USA. Mélitz 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 'open-circuit' and 'direct communication'[7] languages, respectively, increase bilateral trade. Nevertheless, as he only considers indigenous or established languages, he fails to measure the effect of foreign languages. Especially in Europe, knowledge of foreign languages is wide-spread and such non-indigenous languages are likely to play an important role in facilitating trade and economic relations in general.

---

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

3 Data

An important strength of our analysis is that we have detailed information on languages spoken in 29 European countries: including both native as well as foreign languages. The data draw upon a Eurobarometer survey[8] that was carried out in the late 2005 in all member states and candidate countries of the European Union. The respondents[9] were asked to list their mother's tongue, 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 skill 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 we can use it to estimate the share of each country's population that speaks 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 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% have it as their native language or speak it well or very well (Figure 1). Seven EU countries (Cyprus, Denmark, Malta,

---

[8] Special Eurobarometer 243 (EB64.3), Europeans and their languages, European Commission. See http://ec.europa.eu/public_opinion/archives/ebs/ebs_243_sum_en.pdf for detailed information.

[9] 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,

Netherlands, Sweden as well as Ireland and the UK) have majority of their populations 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% (Figure 2 through Figure 4).[10] 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 a 100% proficiency rate in any single country, not even in the country where it is native; this is because of immigrants and/or minorities who do not possess sufficiently good linguistic skills in the host-country language.

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 language or set of languages f as the product of the average proficiency rates, ω_{fi} and ω_{fj}, in the two countries (see, for example, Alesina et al., 2003, and Mélitz, 2008):

$$P_{fij} = \omega_{fi}\omega_{fj} \,. \qquad (1)$$

In doing so, we make no distinction between those who are native speakers of the language 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. This yields 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 Luxemburgers 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

---

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

Danes and 20% of Finns) and Hungarian (spoken by 7% of Romanians and 16% of Slovaks).[11]

English is clearly the most likely conduit for inter-country communication: 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 at 15%. In several cases, the probability that English may serve as the communication language 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 non-negligible for specific pairs of countries.[12]

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 each type of individual (as identified by their linguistic skills) is counted only once.

4 Gravity Model with Languages

We estimate the following gravity equation (all variables are in logarithms):

[11] Results obtained with these languages are available upon request.
[12] 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 case of Finland and Denmark (1%).

$$T_{ijt} = \theta_{ijt} + \beta_1 \left(Y_{it} + Y_{jt}\right) + \beta_2 D_{ij} + \beta_3 B_{ij} + \beta_4 F_{ij} + \beta_5 EU_{ij} + \beta_6 EMU_{ij} + \sum_f^F \delta_f P_{fij} + \varepsilon_{ijt} \quad (2)$$

where T_{ijt} corresponds to the size of bilateral trade between country i and country j at time t, Y_{it} and Y_{jt} stand for the nominal GDP in countries i and j at time t, and D_{ij} is the 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 geographic adjacency, B, for the former federations in East 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. Finally, P_{fij} are indicators of languages d and f, respectively, specific to each pair of countries, which are discussed below.

We follow Baldwin's and Taglioni's (2006) critique of common approaches to estimating the gravity model. Firstly, we define trade volume as the average of logs of exports and imports, instead of the log of average of exports and imports. This precludes a possible bias 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 discussion of the so called gold medal error in Baldwin and Taglioni, 2006). Thirdly, we include country specific time dummies, which stand for both time-invariant and time-variable country specific factors.[13] These effects may include several unobservable factors like institutions, which would bias our results otherwise. Baldwin and Taglioni (2011) argue that this approach leads to unbiased results also when trade includes a high share of intermediate products.

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 section 3). These measure the probability that two randomly chosen inhabitants of

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

country *i* and *j* can communicate in a specific language. Importantly, in computing the probabilities, we make no distinction whether the individuals are native speakers of the language or whether one or both of them speaks it as a foreign language. Clearly, language can facilitate trade even when one or both parties to the transaction speak an acquired rather than native language. The communication probability is thus a better indicator of communication costs than language dummies used in the previous literature, which typically only account for official languages.

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. Columns (1) through (3) of Table 1 presents the results obtained with the 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 more than 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 languages to 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 2.9-fold increase in trade over what can be ascribed only to economic factors and geography. The proficiency in English applies also to 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. 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, German raises trade by approximately 5% (based on the estimates in column 2). The effect of cumulative probability, finally, is also strongly significant and positive in all three regression specifications; the coefficient estimate is approximately half that for English.

The results obtained with the wider data set covering the whole of EU29 are broadly similar, despite some noticeable differences (columns 4 to 6). Besides English, French and German, the analysis now includes also Russian. We also add a dummy variable for countries which arose from the break-up of the former federations in Eastern Europe (Czechoslovakia, Yugoslavia and the USSR) and a dummy for the membership in the EU (to distinguish the 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 members and 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 Block countries. Somewhat surprisingly, communication probability in French seems to have a significantly negative effect on trade in this sample. 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 is presented in the fact that the 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 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 population with a good or very good proficiency in Russian in the new member states stands now 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 formerly 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 is to use instrumental variables to remove the endogeneity bias. Finding suitable and valid instruments, however, is notoriously difficult task.[14] An alternative possibility is to find a suitable natural experiment. This approach has become widely used in economics (Wolpin and Rosenzweig, 2000, Angrist and Pischke, 2010). The political foundations of the communist countries' education systems represent such an experiment because it created a long-term divergence in language skills between Eastern and Western Europe. In this section, therefore, we use the variation in foreign language skills between Western and Eastern European countries to analyze the impact of language skills on trade. To this effect, we restrict the sample only 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

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

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

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

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, 2011). Further columns include the language proficiency 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 the English proficiency, but its size of slightly above 2 is not surprisingly high. Finally, French and Russian proficiency 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 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.

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 so that the estimated 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 underestimated effect of languages. We analyze these factors in this section by means of median and quantile regressions. The median regression is frequently used when standard OLS regression may be biased by outliers. While the least squares regression considers the sum of the squared residuals, which gives much weight to outliers, the median regression finds the regression line that equates the number of positive and negative residuals. This property makes the 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 (2000) argue that inference in quantile regressions is more robust than in ordinary regression. While the concept of quantile regression is now frequently used in economics, especially in labor and family economics (see the literature survey by Koenkeer and Hallock, 2001), it has found little application in trade analysis so far (see Wagner, 2006).

We estimate the following linear model for the τ^{th} conditional quantile, Q, of bilateral trade volume, T,

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

The ease of communication is measured with English proficiency, i.e. based on the specification (1) in Table 1. For computational reasons, we are not able to include country and time effects. The OLS estimation of equation (3) confirms the robustness of the previous results. Table 3 reports the results for the 10^{th}, 25^{th}, 75^{th} and 90^{th} percentiles in addition to the median regression, while details for each 5^{th} percentile are given in Figure 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 the English proficiency is similar to the transport cost elasticity, which underlines the importance of foreign language proficiency for the reduction of the transaction costs. Figure 5 shows that increasing language proficiency has large significant effects at the very beginning of the scale and at relatively high level of proficiency. Thus, both the countries with relatively low and high communicative probabilities tend to display 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 that share 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 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 endogeneity of language skills. Therefore, we utilize a convenient natural experiment embodied in trade between Eastern and Western Europe. Until the early 1990s, the trade between the two parts of Europe was severely restricted due to the Cold War. It is therefore unlikely that the broad segments of the population in East and West possess linguistic skills that are motivated by the economic benefits of East-West trade. The analysis restricted to East-West trade (which thus omits trade flows within East or West) yields results that are very similar to those obtained with the unrestricted sample of all EU countries. This suggests that the resulting 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. Our findings suggest that gains of similar magnitude could be realized by improving linguistic skills. 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. Crucially, while adopting a common currency is costly because a country must give up its national currency and autonomy over monetary policy, improving linguistic skills in foreign languages does not require abandoning national languages. Indeed, small countries like the

Scandinavian countries with high shares of 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 to 60 percent higher than what can be ascribed to economic and geographic factors. Substantial gains thus are possible at relatively little cost.

References

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

Anderson JE, van Wincoop E (2003) Gravity with gravitas: A solution to the border puzzle. American Economic Review 93: 170-192.

Anderson, J.E., and van Wincoop, E. (2004). Trade costs. Journal of Economic Literature 42, 691–751.

Angrist, J. D., Pischke, J.-S. 2010. The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. Journal of Economic Perspectives 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 (2006) Gravity chains: estimating bilateral trade flows when parts and components trade is important. Working Paper No. 1401, European Central Bank, Frankfurt.

Baltagi, BH (2005) Econometric Analysis of Panel Data, Wiley and Sons, Chichester, 3rd edition.

Bussière M, Fidrmuc J, Schnatz B (2005) Trade Integration of Central and South Eastern European Countries: Lessons from a Gravity Model. Working Paper No. 545, European Central Bank, Frankfurt.

Bussière, M., Fidrmuc, J., Schnatz, B. (2008) EU Enlargement and Trade Integration: Lessons from a Gravity Model. Review of Development Economics, 12, 3, 501-515.

Chiswick, B.R., Miller, P.W. (2002), Immigrant Earnings: Langauge Skills, Linguistic Concentrations and the Business Cycle," Journal of Population Economics 15, 31-57.

Chiswick, B.R., Miller, P.W. (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 (2003) An econometric view on the estimation of gravity models and the calculation of trade potentials. World Economy 25: 297-312.

Egger, P. and 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.

Feenstra RC (2002) Border effect and the gravity equation: Consistent method for estimation. Scottish Journal of Political Economy 49: 491-506.

Felbermayr, G.J., and Toubal, F. (2010) Cultural proximity and trade. *European Economic Review* 54, 279-293.

Grenier, G., and 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.

Ginsburgh, V., Prieto-Rodriguez, J. (2006), Returns to Foreign Languages of Native Workers in the EU, mimeo.

Glick, R, Rose AK (2002) Does a Currency Union Affect Trade? The Time-Series Evidence. European Economic Review 46: 1125–1151.

Helpman, E., Melitz, M. J., Rubinstein, Y. (2009), Estimating Trade Flows: Trading Partners and Trading Volumes, Quarterly Journal of Economics 123 (2), 441-487.

Hutchinson, W., (2002) Does ease of communication increase trade? Commonality of language and bilateral trade. Scottish Journal of Political Economy 49, 544-556.

Ipshording, I.E. and S. Otten (2013). “The Costs of Babylon: Linguistic Distance in Applied Economics.” *Review of International Economics* 21(2), 354-369.

Koenker, Roger and Bassett, Gilbert (1978) Regression Quantiles. Econometrica. 46(1), 33–50.

Koenker, R. and Hallock, K. F. (2006) Quantile Regression. Journal of Economic Perspectives 15 (4), 143–156.

Levchenko, A. A., Lewis, L. T. , Tesar, L. L. (2010). The Collapse of International Trade during the 2008–09 Crisis: In Search of the Smoking Gun. IMF Economic Review, 58, 214-253.

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

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

McCallum, J. (1995) National borders matter: Canada-U.S. regional trade patterns, American Economic Review 85: 615-623.

Mátyás, L. (1997) Proper econometric specification of the gravity model, World Economy 20: 363-368.

Mátyás, L. (1998) The gravity model: Some econometric considerations, World Economy 21: 397-401.

Melitz, J (2008), “Language and Foreign Trade,” European Economic Review 52 (4), 667-699.

OECD (2006) Science Competencies for Tomorrow’s World, Programme for International Student Assessment, OECD, Paris.

Rauch, JE, Trindade, V (2002), “Ethnic Chinese Networks in International Trade,” Review of Economics and Statistics 84 (1), 116–130.

Wolpin, K. I., Rosenzweig, M. R. (2000) Natural “Natural Experiments” in Economics. Journal of Economic Literature 38(4), pp. 827-874.

Figure 1: Proficiency in English (native, very good or good proficiency)

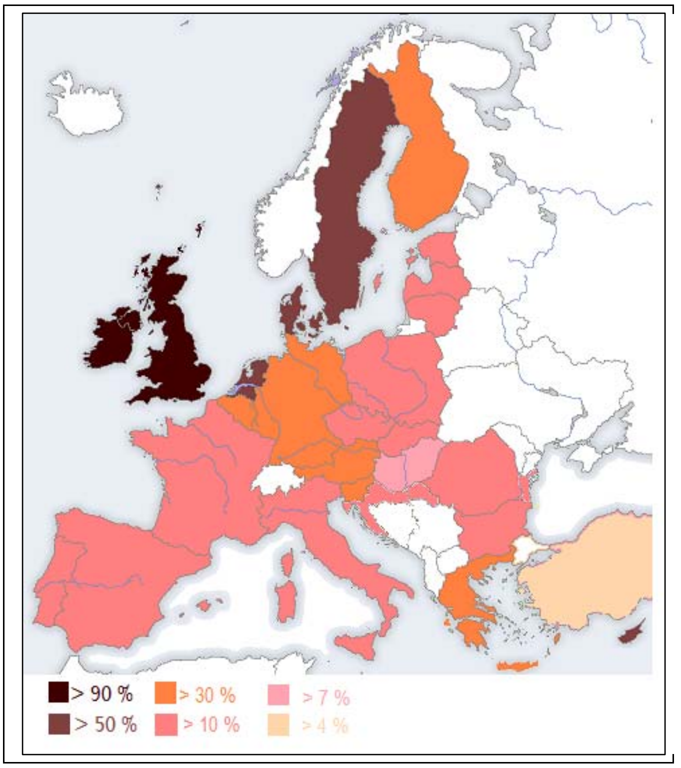


Figure 2: Proficiency in French (native, very good or good proficiency)

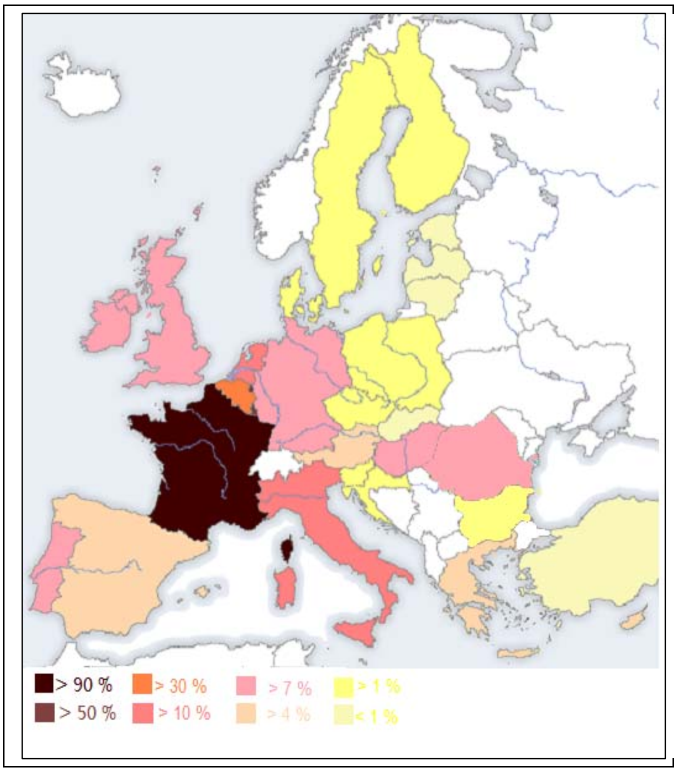


Figure 3: Proficiency in German (native, very good or good proficiency)

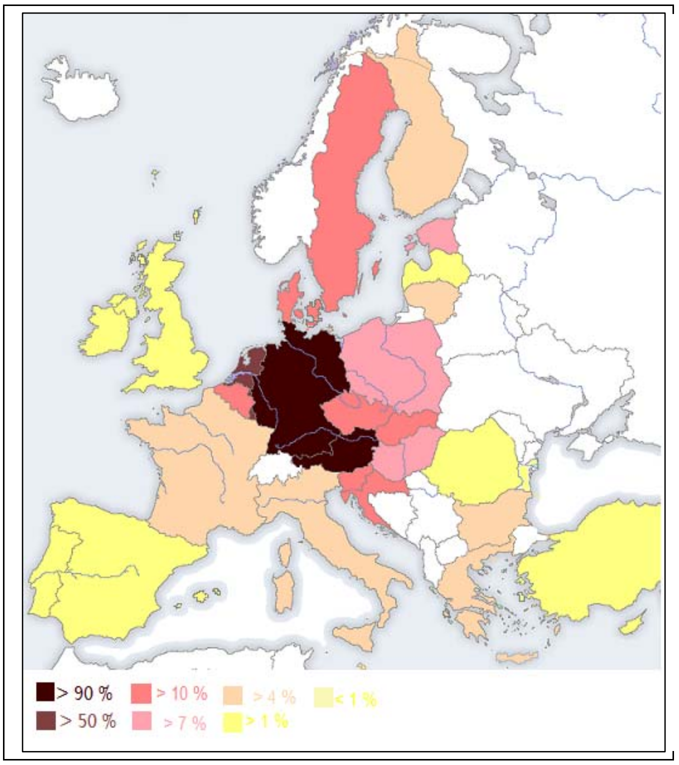


Figure 4: Proficiency in Russian (native, very good or good proficiency)

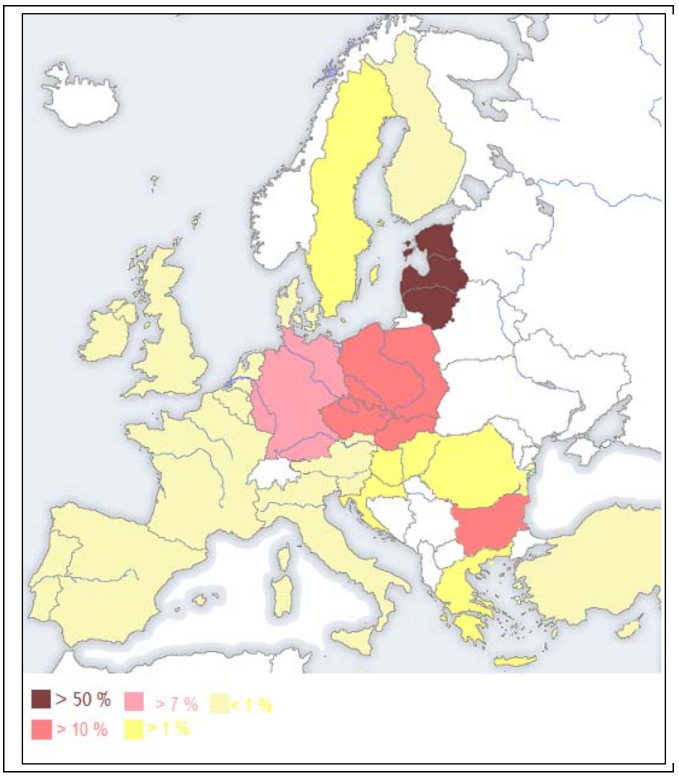


Table 1: Trade Effects of Foreign Languages, EU15 and EU29

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***
	(45.150)	(43.115)	(44.016)	(101.297)	(100.013)	(100.553)
GDP	0.915***	0.920***	0.923***	0.916***	0.924***	0.916***
	(41.323)	(41.048)	(40.782)	(79.985)	(80.472)	(79.903)
Distance	-0.767***	-0.757***	-0.759***	-1.047***	-1.051***	-1.043***
	(-28.906)	(-27.493)	(-27.746)	(-50.708)	(-50.409)	(-50.162)
Contiguity	0.541***	0.519***	0.466***	0.384***	0.405***	0.361***
	(16.398)	(15.007)	(12.322)	(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***
	(8.908)	(8.668)	(7.522)	(4.925)	(4.935)	(4.292)
English	1.078***	1.101***		0.653***	0.622***	
	(9.481)	(9.566)		(5.523)	(5.280)	
French		0.048			-0.479***	
		(0.436)			(-2.759)	
German		0.241***			0.051	
		(3.586)			(0.411)	
Russian					1.642***	
					(6.159)	
Cumulat[a]			0.573***			0.336***
			(7.463)			(3.739)
N	1470	1470	1470	5634	5634	5634
Adj. R^2	0.965	0.966	0.964	0.923	0.924	0.923

Note: 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). Country-specific time dummies are not reported. *t*-statistics based on robust standard errors are in parentheses. ***, **, and * denote significance at 1 per cent, 5 per cent, and 10 per cent, respectively.

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***
	(26.842)	(26.607)	(26.713)	(22.885)	(26.618)	(22.482)	(25.764)
GDP	0.947***	0.961***	0.948***	0.945***	0.947***	0.958***	0.962***
	(36.054)	(35.987)	(36.228)	(35.541)	(36.046)	(35.607)	(36.023)
Distance	-0.839***	-0.850***	-0.835***	-0.783***	-0.839***	-0.791***	-0.829***
	(-11.149)	(-11.228)	(-11.066)	(-9.034)	(-11.058)	(-8.988)	(-10.992)
Contiguity	0.485***	0.490***	0.488***	0.478***	0.485***	0.488***	0.486***
	(4.749)	(4.788)	(4.775)	(4.728)	(4.732)	(4.786)	(4.753)
English		0.944***				0.928***	
		(3.491)				(3.430)	
French			-2.490			-2.536	
			(-1.019)			(-1.037)	
German				2.164***		2.073***	
				(2.738)		(2.624)	
Russian					0.306	0.916	
					(0.138)	(0.426)	
Cumul[a]							1.097***
							(4.313)
N	2,006	2,006	2,006	2,006	2,006	2,006	2,006
Adjd R^2	0.880	0.880	0.880	0.880	0.880	0.881	0.881

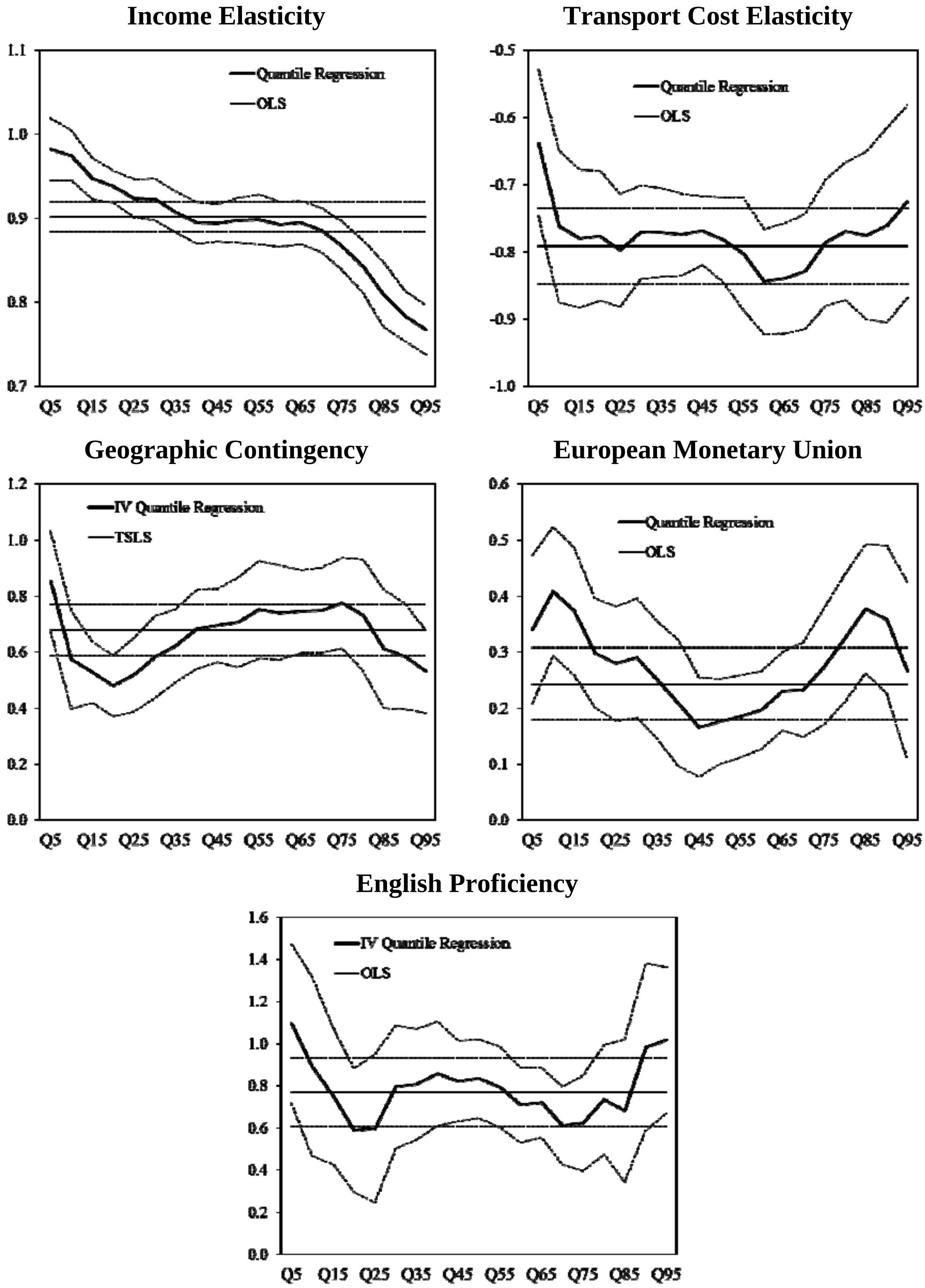
Note: a – cumulative probability that two inhabitants of the country pair can communicate in English, French or German (reflecting knowledge of two or all three languages). Country-specific time dummies are not reported. *t*-statistics based on robust standard errors are in parentheses. ***, **, and * denote significance at 1 per cent, 5 per cent, and 10 per cent, respectively.

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

Note: Robust *t*-statistics using bootstrapped (1000 replications) standard errors are in parentheses. ***, **, and * denote significance at 1 per cent, 5 per cent, and 10 per cent, respectively.

Figure 5: OLS Regression and Quantile Regression Estimates



Note: Bootstrapped standard errors are used for 95% confidence bands.