Foreign Languages and Trade:
What are you sinking about?

Jarko Fidrmuc[†]
Jan Fidrmuc[‡]

November 2008

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 Europe. We show that widespread knowledge of languages is an important determinant for foreign trade, with English playing an especially important role. Furthermore, we document non-linear effects of foreign languages on trade.

Keywords: Gravity models, foreign trade, language effects, quantile regression.
JEL Classification: C23, F15, F40, Z10.

[†] University of Munich, Department of Economics; CESifo Institute Munich; and Comenius University Bratislava, Slovakia Institute of Applied Mathematics and Statistics, e-mail: jarko.fidrmuc@lrz.uni-muenchen.de. Contact information: Department of Economics, University of Munich, Geschwister-Scholl-Platz 1, 80539 Munich, Germany
[‡] Department of Economics and Finance, and Centre for Economic Development and Institutions (CEDI), Brunel University; CEPR, London; and WDI, University of Michigan. Contact information: Department of Economics and Finance, Brunel University, Uxbridge, UB8 3PH, United Kingdom. Email: Jan.Fidrmuc@brunel.ac.uk or jan@fidrmuc.net. Phone: +44-1895-266-528, Fax: +44-1895-203-384. Web: http://www.fidrmuc.net/.

1 Introduction

Languages facilitate communication and ease transactions. Two individuals who speak the same language can communicate and trade with each other directly whereas those without a sufficient knowledge of a common language must often rely on an intermediary or hire an interpreter. The additional complexity inherent in such a mediated relationship, the potential for costly errors[1] and their increased cost may be large enough to prevent otherwise mutually beneficial transactions from occurring. Consequently, ability to speak foreign languages should have a positive economic payoff embodied in better employment opportunities and higher wages[2] -- in addition to other, non-pecuniary benefits such as ability to travel, study and live abroad, to meet new people, to read foreign books or newspapers, and the like.

In this paper, 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 both country A and country B speak the same language, they will be able to communicate with each other more readily. Consequently, trade between these two countries will be easier and cheaper. 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 dozens of former French and Belgian colonies). Such studies invariably find that sharing language translates into greater trade intensity. However, languages need not have the official status in order to foster trade:

---

[1] A well-known, while tongue-in-cheek, example is a commercial by Berlitz, a language school, in which a German coastguard receives a distress call 'We are sinking!', to which he responds 'What are you sinking about?' See http://www.youtube.com/watch?v=8vBn2_ia8zM.

[2] Most empirical studies focus on immigrants (e.g. Chiswick and Miller, 2002 and 2007) where positive returns to the ability to speak the host-country language is not surprising. Ginsburgh and Prieto-Rodriguez (2006) estimate the returns to using a foreign language at work for native Europeans and find positive returns which depend on the relative *scarcity* of the foreign langauge (for instance, English has a much lower return in Denmark than in Spain).

international commerce is increasingly conducted in English, even if neither party to the transaction is from an English speaking country.

We utilize a new and previously little used survey data set on language use in the member and candidate countries of the European Union. Importantly, the data contain detailed information not only on European's 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 estimate probabilities that two randomly chosen individuals from two different countries will be able to communicate. We investigate the effect of such *communicative probabilities* on bilateral trade flows in Europe.

While most gravity-model types of analyses considered only official languages, Mélitz (2008) went a step further by considering all (indigenous) languages spoken in a country and accounting for the fraction of the population speaking them. English, for example, is spoken in dozens of former British colonies but often only a small fraction of the population speak it, and Chinese is spoken in a number of South Asian countries even while it does not enjoy an official-language status in all of them. Nevertheless, by focusing on languages that are indigenous, Mélitz fails to take account of foreign languages: a Chinese tradesman in French-speaking Africa may be more inclined to communicate with his business partners in English than in either French or Chinese.

We find that greater density of linguistic skills indeed translates into greater trade intensity. In the earlier 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% (this probability makes no distinction between native speakers of English and those who speak it as a foreign language except that we require that the self-assessed proficiency for the latter is at least good or very good). This raises intra-EU15 trade, on average, by approximately 30%. German and French, in contrast, produce only weak and mixed results. It appears, indeed, that English is the main driver of international trade, at least in Western Europe.

We find furthermore that the effect of foreign languages is not uniform across countries. When we expand our analysis to include all 29 member and candidate countries[3], the

[3] At present, Croatia and Turkey are the only countries with the candidate status.

effect of English appears weaker or outright insignificant (nevertheless, English appears significant in a sample including only the new members and candidates for membership). This could be either due to their much shorter and more limited history of integration. Furthermore, we show that the effect of languages is in fact non-linear (on average, fewer people speak English in the new member and candidate countries). This finding is also consistent with the pattern observed for the more marginal European languages (marginal in the sense of not being spoken widely in Europe, except in their native countries): Italian, Spanish, Russian, Swedish and Hungarian. These appear with relatively large coefficients in our regressions, indicating that languages may have diminishing returns with respect to trade.

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 main body of our empirical analysis while section 5 presents some robustness checks. 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), relates bilateral trade to the aggregate supply and aggregate demand of, respectively, the exporting and importing country, to transport and transaction costs, and to specific trade factors (e.g. free trade agreements). It has proved an extremely popular tool for applied trade analysis. In particular, 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' (McCallum, 1995) and to estimate the effects of currency unions on trade (Rose, 2000). Further research applies gravity models to trade in services (Kimura and Lee, 2006) and FDI (Egger and Pfaffermayr, 2004).

Accounting for common official languages has become a standard feature of gravity models. The gravity equation is augmented to include a common-language dummy, alongside other potential determinants of bilateral trade such as common border, landlocked dummy and indicators of shared colonial heritage.[4] Most studies, however,

[4] More recent studies include these factors usually as fixed effects.

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. Several languages, for example, have the status of the official language in two or more European countries: German (Austria, Germany and Luxembourg), French (France and Belgium), Dutch (Belgium and Netherlands), Swedish (Sweden and Finland), and Greek (Greece and Cyprus). 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 effect of economic integration in the EU.

Some studies, such as Rauch and Trindade (2002), find that the presence of immigrants helps foster trade links between their country of origin and the ancestral country. To the best of our knowledge, the only paper that focuses specifically on the relationship between bilateral trade and languages is Mélitz (2008). He goes beyond focusing on official languages and instead considers all indigenous languages spoken by at least 4% of the population, in addition to official languages.[5] He finds that both categories of languages that he defines, 'open-circuit' and 'direct communication'[6] languages, increase bilateral trade. Nevertheless, as he only considers indigenous languages, he fails to measure the effect of foreign languages.

3 Data

We base our analysis on data on bilateral trade flows among 29 countries that are at present member states or candidates for membership of the European Union, which are taken from Bussière et al. (2005 and 2008). The trade flows are observed between 2001 and 2007. The data are compiled from the IMF Direction of Trade Statistics; they are expressed in US dollars. Nominal GDP data converted to US dollars are from the IMF

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

[6] 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.

International Financial Statistics. The distance term is measured in terms of great circle distances between the capitals of country *i* and country *j*.

We augment the trade and output data with survey data on European's ability to speak various languages. This Eurobarometer survey[7] was carried out in the late 2005 in all member states and candidates countries of the European Union. The respondents, who had to be EU citizens (although not necessarily nationals of the country in which they were interviewed), were asked to list their mother's tongue (allowing for multiple entries when applicable) 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. These surveys are nationally representative (with the limitation that they do not account for linguistic skills of non-EU nationals) and therefore we can use them to estimate the share of each country's population that speaks each language.[8]

English is the language spoken by the largest number of Europeans: 33% of the 29 countries included in our analysis speak it as their native language or speak it well or very well (Figure 1). Furthermore, five EU non-English-speaking countries have majority of their population proficient in English and only two countries have proficiency rates below 10%. German is spoken by 22%, French by 17% and Russian by 4% (Figure 2 through Figure 4).[9] 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 presumably because of immigrants who do not possess good linguistic skills in the host-country language.

---

[7] 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.

[8] The data report figures for 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).

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

Figure 1: Proficiency in English (native and good/very good skills)

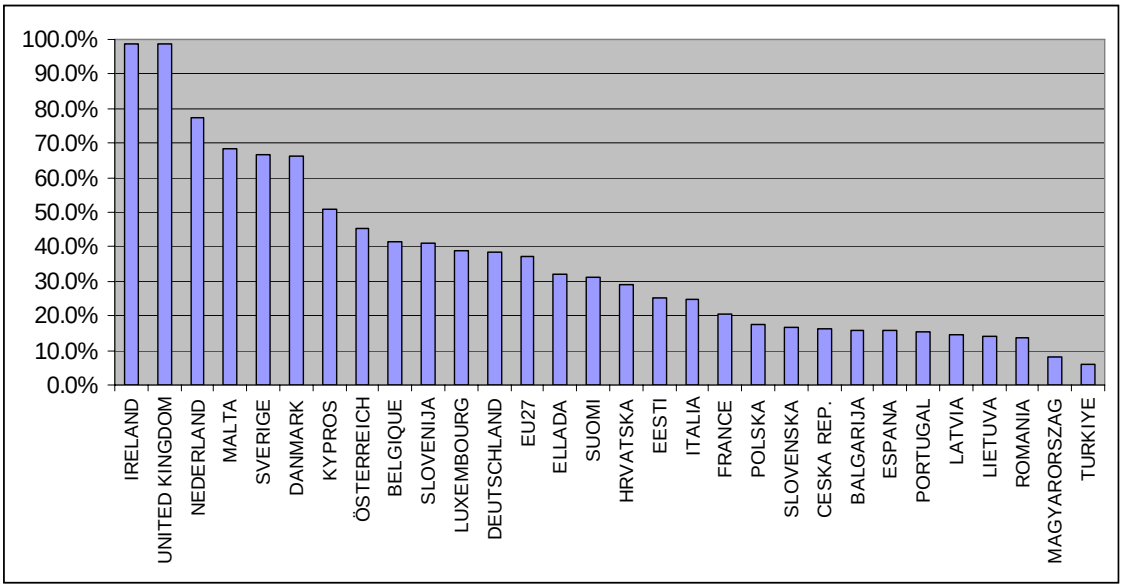


Figure 2: Proficiency in German (native and good/very good skills)

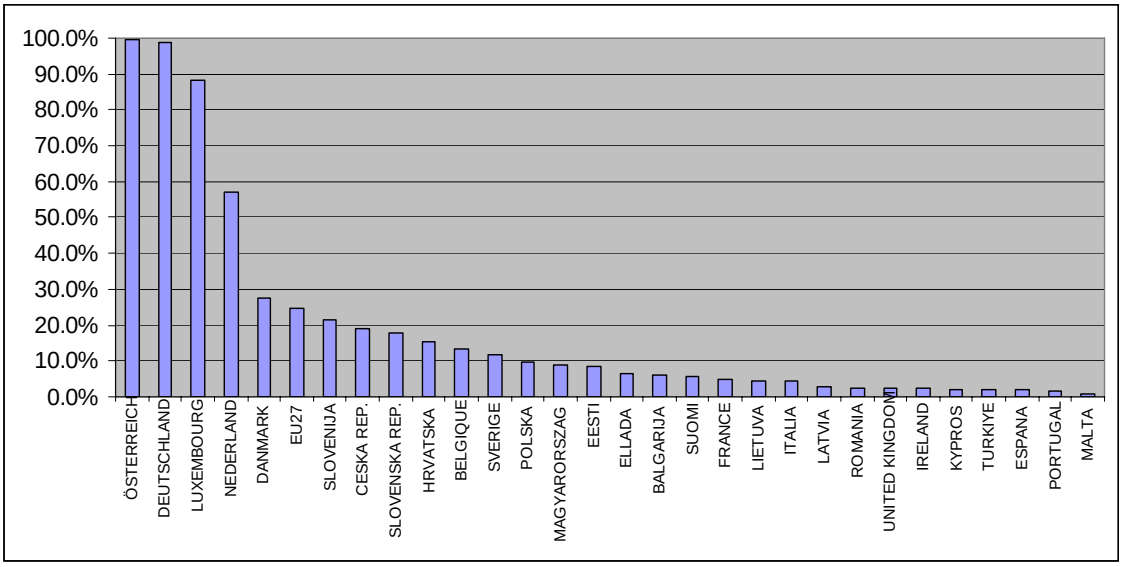


Figure 3: Proficiency in French (native and good/very good skills)

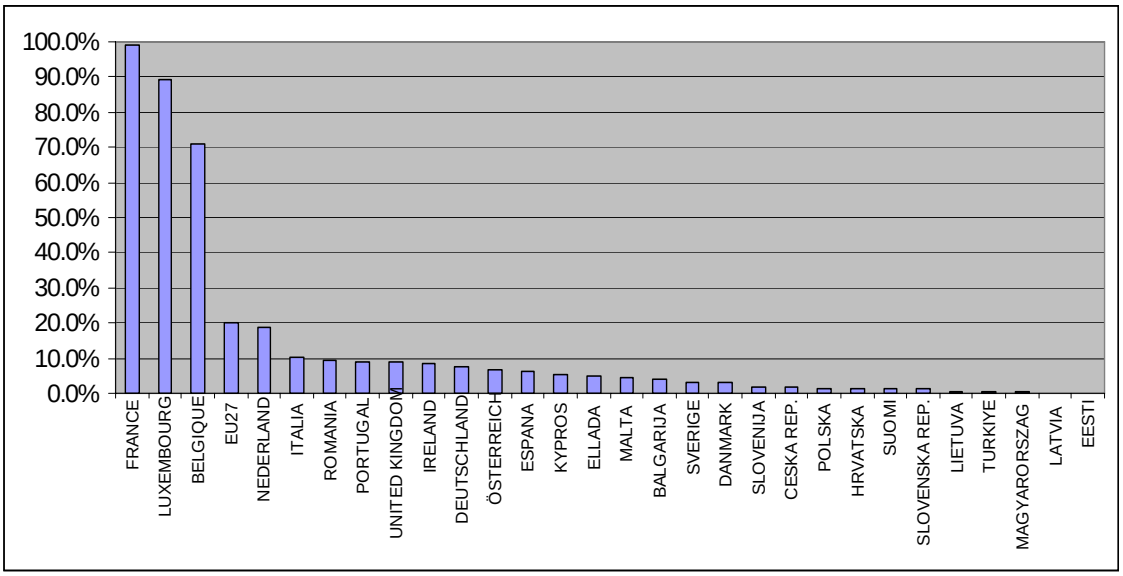
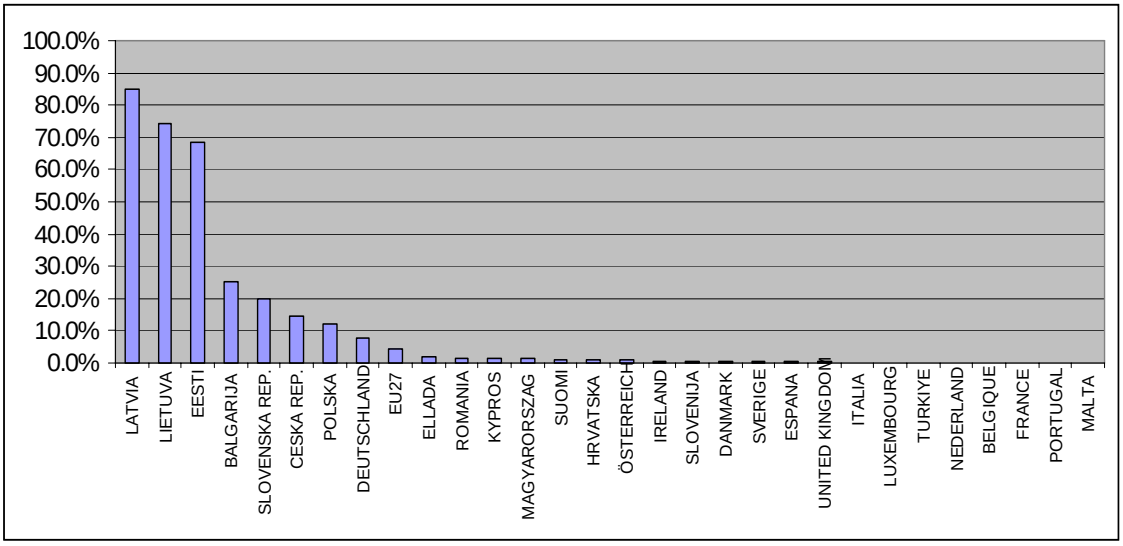


Figure 4: Proficiency in French (native and good/very good skills)



We use the average proficiency rates to estimate probabilities that two randomly chosen individuals from two different countries will be able to communicate with each other. 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 rather than merely basic. To include a language in our analysis, we start with the requirement that it should be spoken by at least 10% of the population in at least three countries. This yields four languages: English, German, French and Russian – the last being spoken

mainly in the new member countries, while also Germany is close to this threshold (8% of population). Note that this relatively strict definition leaves out Italian, spoken by 3-5% of Austrian, Belgian, French and Luxembourgish population and 7-9% of Croats and Slovenes. Similarly, Spanish, spoken widely outside of the EU and by between 2-7% of Austria, Denmark, France, Germany, Netherlands and Portugal, is not included. Lowering the threshold to 4% therefore adds these two languages and also Swedish (spoken by 8% of Danes and 20% of Finns) and Hungarian (spoken by 7% of Rumanians and 16% of Slovaks).

Again, English is most likely to serve as a 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 still 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%; in general these are all countries with Romance languages.

German and French lag far behind English, with 5 and 3% respectively (or 7 and 5% in the EU15). Nevertheless, there are some cases where the communicative probability is relatively high. There is a 16% probability that a Dutchman and a Dane will be able to use German in their communication. For all the remaining languages, the average communicative probability is essentially zero, although it is often non-negligible for specific pairs of countries.[10]

4 Gravity Models

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

$$T_{ijt} = \theta_{ijt} + \beta_1 (y_{it} + y_{jt}) + \beta_2 d_{ij} + \beta_3 b_{ij} + \beta_4 f_{ij} + \sum_{d}^{D} \delta_d L_{d,ij} + \sum_{f}^{F} \delta_f P_{f,ij} + \varepsilon_{ijt}, \qquad (1)$$

[10] 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%).

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 the countries i and j at time t, and d_{ij} is the distance variable proxying for transport costs. The income elasticity of foreign trade, β_1 is expected to be positive, while transport cost elasticity, β_2, should be negative. We also include a control variable for geographic adjacency, b, and for former federations in East Europe, f, which broke up in the last two decades. Both variables are expected to have positive effects on trade. Finally, L_{dij} and P_{fij} are indicators for 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 gravity model estimations. Firstly, we define trade volume as the average of logs of exports and imports, instead of log of average of exports and imports. This precludes 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 US dollars 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 all time-invariant and time-variable country specific factors.[11]

In addition to the core variables of gravity models, we include two sets of indicators on bilateral language relationships between the countries. First, we use standard official-language dummies, which are used commonly in gravity models. Thus, we use dummies for English (Ireland, Malta and the UK), French (France, Belgium and Luxembourg), German (Germany, Austria and Luxembourg), Swedish (Sweden and Finland), Dutch (Belgium and the Netherlands), and Greek (Greece and Cyprus). Second, we include communicative probabilities for English, French, German, Russian, Spanish, Italian, Swedish and Hungarian (constructed as explained in section 3). These indicators measure the probability that two randomly chosen inhabitants of country i and j can communicate in the specific language. Importantly, we make no distinction whether the individuals are native speakers of the language or whether one or both of

[11] Alternative specifications of gravity models with simple country dummies (Mátyás, 1997 and 1998) or as a standard OLS, which are also popular in the literature, are available upon request.

them speaks it as a foreign language other than the dummies for common official languages. Clearly, language can facilitate trade also when one or both parties to the transaction speak an acquired rather than native 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 many economic, historical and cultural characteristics. Still, language differences may pose a significant barrier to trade also within this group.

Table 1 compares the results obtained with the various alternative ways of controlling for bilateral language relations between countries. Column (1) is the baseline result, with official-language dummies only. English, French and German raise bilateral trade between their countries using these languages: they increase trade between 1.3 (French) and 2.6 times (English). Swedish, in contrast, has little effect and Dutch even appears to lower trade slightly. This may be due to the fact that although Swedish and Dutch are official languages of Finland and Belgium, respectively, each of these two countries has one other official language (Finnish and French). In column (2), we add the communicative probability for English. As a result of accounting for the probability of communicating in English (which relates to English-speaking countries just as well as the other countries), the effect of having English as an official language falls by approximately one third (from factor of 2.7 to 1.9). English communicative probability has a positive impact on trade and is strongly significant: the communicative probability for the UK and Ireland is 0.97 which translates into 2.9-fold increase in trade. Overall, trade between UK and Ireland is more than 5 times higher than what can be ascribed only to economic factors and geography. The proficiency in English is an important conduit of trade between other countries as well. For example, the trade between the Netherlands and Sweden, is increased by three quarters and Dutch trade with the UK is more than doubled. With English communicative probability 22% in the EU15 on average, ability to communicate in English increases trade by approximately one fifth.

In the rest of the table, we add communicative probabilities in further languages: French and German in column (3), Italian and Spanish in column (4) and Swedish in column (5). Adding communicative probabilities for these two languages seems to strengthen the impact of German as an official language while the corresponding impact of French

disappears. Communicative probability in French appears to raise trade but its effect becomes insignificant in the last column. German appears even to have a negative impact in column (3), is not significant in column (4) and has a positive impact in column (5). Italian, Spanish and Swedish all appear with very large and strongly significant coefficient. Note, however, that these languages are only spoken in a handful of countries each so that the vast majority of observations is clustered very close to zero. This makes the interpretation of these results difficult. Importantly, adding further languages affects the regression estimates for English little.

Table 2 presents similar results for the new member states and candidate countries. Because French, Italian and Spanish are marginal in this group of countries, we are not including them. Instead, column (3) features Russian and column (4) Hungarian. No official-language dummies are included because there are no two or more countries with the same official language. English communicative probability again has a strong impact on trade. In fact, the effect appears much larger in this group of countries than in the EU15. However, one must bear in mind the generally lower levels of English proficiency in the new members and candidates (the average communicative probability is 11%). Nevertheless, the effect is sizeable: on average, the ability to communicate in English raises trade by 60% in these countries.

Finally, Table 3 merges the two groups of countries. We now add one more common official language, Greek, along with, gradually, the communicative probabilities in all of the above-listed languages. English only appears significant in column (2), where we do not account for any other communicative probabilities, and is insignificant in the rest of the table. Moreover, its effect appears lower than in either of the preceding two sets of regressions. French communicative probability is not significant at all and German even appears to lower trade. The remaining languages, while again being spoken only in relatively few countries, all appear with positive and strongly significant coefficient estimates.

The mixed and generally disappointing results in Table 3 can be due to two factors. First, while the EU15 countries share a legacy of long and gradual economic integration, the EU29 is much more heterogenous. Second, the impact of language proficiency on trade can be non-linear. In particular, communicative probability can

have diminishing returns so that trade is increased more for low to moderate levels than for relatively high levels.

To explore the possibility of non-linear relationship between communicative probability and trade, we add the square of the communicative probability into our regressions. Table 4 presents again first the results for the EU15 countries. Focusing on the impact of English communicative probability, all regressions suggest that it has a hump-shaped effect on trade flows (see Figure 6 in the Appendix). The effect peaks when the communicative probability is approximately 70%. Note, however, that although this seems to suggest that English-speaking countries could do better by lowering their English proficiency, they also receive the positive impact of having English as their official language (captured by the common-language dummy) – and this effect rises when we control for the English communicative probability. Table 5 presents similar results for the new members and candidates and the regression results again suggest a hump-shaped effect of English communicative probability – although the coefficient estimates are again different from those estimated for the EU15. Finally, the hump-shaped effect is confirmed also in the regressions for all countries in Table 6. It appears, therefore, that the returns to English proficiency indeed are diminishing. The estimated impact, however, varies substantially across the three sets of countries (see also Figure 6 through Figure 8 in the Appendix). This again suggests that the EU15 and the new members and candidates are very heterogenous.

We can use our estimates to demonstrate the potential effects of improvement in English proficiency. Considering the linear specification, an improvement in English proficiency in all EU countries increased by 10 percentage points (keeping UK and Ireland constant) would increase the intra-EU15 trade by 15% on average. This increase would not be shared uniformly by all countries: while Portuguese trade would go up by some 9%, Dutch trade could increase by as much as 24% (UK and Ireland would be close behind with 21% trade increases). An even greater increase, one that would bring all countries to the level of English proficiency attained by the Netherlands (again, assuming that the UK and Ireland's proficiency levels would remain unchanged), would bring about an average increase in EU15 trade by 70%.

Table 1: Trade effects of Foreign Languages, EU15

Variable	**(1)**		**(2)**		**(3)**		**(4)**		**(5)**	
	teu_td0		teu_td1		teu_td2		teu_td3		teu_td4	
Intercept	14.841	***	15.175	***	15.415	***	15.318	***	14.678	***
	(50.231)		(49.699)		(45.150)		(46.108)		(46.115)	
GDP	1.004	***	0.897	***	0.885	***	0.880	***	0.895	***
	(53.232)		(47.047)		(44.808)		(44.863)		(47.537)	
Distance	-0.772	***	-0.748	***	-0.761	***	-0.750	***	-0.668	***
	(-26.965)		(-26.831)		(-25.893)		(-26.295)		(-24.708)	
Contiguity	0.499	***	0.471	***	0.491	***	0.364	***	0.157	***
	(13.558)		(13.310)		(13.696)		(8.001)		(5.371)	
Official languages										
English	0.908	***	0.543	***	0.570	***	0.662	***	0.775	***
	(11.572)		(6.536)		(6.646)		(8.701)		(12.480)	
German	0.556	***	0.581	***	0.853	***	0.841	***	0.667	***
	(12.581)		(13.379)		(10.409)		(10.875)		(11.433)	
French	0.150	**	0.186	**	0.101		0.295		0.788	***
	(1.877)		(2.328)		(0.382)		(1.109)		(3.093)	
Swedish	0.158		0.279	***	0.235	**	0.323	***	-2.974	***
	(1.932)		(3.300)		(2.728)		(3.459)		(-20.104)	
Dutch	-0.344	***	-0.263	***	-0.340	***	-0.180	**	0.150	***
	(-5.978)		(-4.529)		(-5.028)		(-2.347)		(2.549)	
Proficiency:										
English			1.152	***	1.074	***	0.944	***	1.022	***
			(9.261)		(8.352)		(7.399)		(8.437)	
French					0.080		0.065		-0.321	
					(0.226)		(0.186)		(-0.940)	
German					-0.408	***	-0.274	***	0.102	
					(-3.948)		(-2.641)		(1.391)	
Italian							8.724	***	11.687	***
							(8.074)		(12.705)	
Spanish							8.938	***	12.071	***
							(8.363)		(12.987)	
Swedish									19.793	***
									(22.672)	
N	1470		1470		1470		1470		1470	***
Adjusted R^2	0.972		0.974		0.974		0.975		0.980	

Note: Country-specific time dummies are not reported. *t*-statistics 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, NMS and Associated Countries (including Turkey)

Variable	**(1)** teu_td0		**(2)** teu_td1		**(3)** teu_td2		**(4)** teu_td4	
Intercept	19.838	***	19.372	***	17.119	***	17.145	***
	(11.294)		(11.050)		(8.450)		(8.282)	
GDP	0.571	***	0.573	***	0.566	***	0.566	***
	(2.429)		(2.446)		(2.405)		(2.405)	
Distance	-1.039	***	-1.024	***	-0.817	***	-0.820	***
	(-5.999)		(-6.148)		(-4.128)		(-4.016)	
Former Federation	2.278	***	2.292	***	1.478	***	1.471	***
	(11.131)		(11.428)		(10.418)		(9.628)	
Contiguity	0.543	***	0.531	***	0.650	***	0.654	***
	(4.652)		(4.835)		(5.473)		(5.833)	
Proficiency:								
English			5.074	***	5.182	***	5.188	***
			(3.371)		(3.440)		(3.454)	
German					13.381	*	13.239	*
					(1.738)		(1.667)	
Russian					3.748	***	3.745	***
					(8.954)		(8.873)	
Hungarian							-0.309	
							(-0.335)	
N	1254		1254		1254		1254	
Adjusted R^2	0.847		0.850		0.858		0.858	

Note: See Table 1.

Table 3: Trade effects of Foreign Languages, All Countries

Variable	**(1)**		**(2)**		**(3)**		**(4)**		**(5)**	
	tall_td0		tall_td1		tall_td2		tall_td3		tall_td4	
Intercept	19.177	***	18.808	***	18.853	***	18.694	***	18.669	***
	(41.560)		(38.491)		(37.729)		(36.468)		(36.371)	
GDP	0.865	***	0.875	***	0.879	***	0.884	***	0.884	***
	(38.867)		(38.891)		(38.981)		(38.762)		(39.032)	
Distance	-1.055	***	-1.042	***	-1.045	***	-1.034	***	-1.028	***
	(-24.613)		(-23.829)		(-22.916)		(-22.310)		(-22.078)	
Former Federation	2.472	***	2.465	***	1.948	***	1.978	***	2.034	***
	(30.017)		(30.238)		(25.097)		(24.822)		(25.272)	
Contiguity	0.318	***	0.324	***	0.338	***	0.326	***	0.267	***
	(7.013)		(7.168)		(7.519)		(7.340)		(6.047)	
Official languages										
English	0.916	***	0.669	***	0.699	***	0.709	***	0.746	***
	(7.924)		(5.175)		(5.397)		(5.479)		(5.792)	
German	0.599	***	0.601	***	0.931	***	0.911	***	0.854	***
	(10.006)		(10.037)		(8.606)		(8.398)		(8.072)	
French	0.048		0.069		0.076		0.088		0.150	
	(0.432)		(0.620)		(0.228)		(0.268)		(0.456)	
Swedish	0.150		0.174	***	0.146	***	0.168	***	-2.176	***
	(2.747)		(3.159)		(2.602)		(2.884)		(-12.073)	
Dutch	-0.617	***	-0.618	***	-0.655	***	-0.624	***	-0.554	***
	(-10.120)		(-10.158)		(-10.008)		(-9.184)		(-8.002)	
Greek	2.272	***	2.294	***	2.282	***	2.297	***	2.327	***
	(14.428)		(14.507)		(14.252)		(14.445)		(15.134)	
Proficiency:										
English			0.763	**	0.658	***	0.688	***	0.597	***
			(5.111)		(4.362)		(4.513)		(3.935)	
French					-0.064		-0.028		-0.030	
					(-0.141)		(-0.062)		(-0.067)	
German					-0.465	***	-0.424	***	-0.318	**
					(-3.204)		(-2.895)		(-2.189)	
Russian					1.675	***	1.627	***	1.623	***
					(8.687)		(8.365)		(8.342)	
Italian							1.532	***	1.606	***
							(5.287)		(5.442)	
Spanish							3.582	***	4.362	***
							(3.536)		(4.221)	
Swedish									12.824	***
									(12.658)	
Hungarian									3.679	***
									(4.492)	
N	5634		5634		5634		5634		5634	
Adjusted R^2	0.930		0.930		0.930		0.930		0.931	

Note: See Table 1.

Table 4: Trade effects of Foreign Languages, Non-Linear Specification, EU15

Variable	**(1)** tall_td0		**(2)** tall_td1		**(3)** tall_td2		**(4)** tall_td3		**(5)** tall_td4	
Intercept	14.841	***	14.084	***	14.569	***	14.183	***	13.567	***
	(50.231)		(42.399)		(40.016)		(40.485)		(41.288)	
GDP	1.004	***	0.955	***	0.921	***	0.917	***	0.932	***
	(53.232)		(47.613)		(44.312)		(43.684)		(46.944)	
Distance	-0.772	***	-0.726	***	-0.748	***	-0.702	***	-0.605	***
	(-26.965)		(-26.881)		(-26.781)		(-27.262)		(25.550)	
Contiguity	0.499	***	0.429	***	0.451	***	0.424	***	0.265	***
	(13.558)		(12.615)		(14.712)		(9.891)		(9.231)	
Official languages										
English	0.908	***	1.369	***	1.672	***	1.749	***	1.601	***
	(11.572)		(12.209)		(13.622)		(14.783)		(14.940)	
German	0.556	***	0.661	***	0.030		0.015		0.325	***
	(12.581)		(15.015)		(0.210)		(0.116)		(2.969)	
French	0.150	*	0.292	***	0.400		0.514	**	1.003	***
	(1.877)		(3.650)		(1.621)		(2.061)		(4.208)	
Swedish	0.158	**	0.362	***	0.256	***	0.279	***	17.057	***
	(1.932)		(4.428)		(3.370)		(3.474)		(19.458)	
Dutch	-0.344	***	-0.283	***	-0.404	***	-0.286	***	0.030	
	(-5.978)		(-5.053)		(-6.444)		(-4.118)		(0.545)	
Proficiency:										
English			5.157	***	6.005	***	6.008	***	5.178	***
			(10.526)		(11.581)		(10.953)		(10.061)	
French					1.119	***	1.220	***	0.040	
					(2.439)		(2.733)		(0.103)	
German					-2.633	***	-2.499	***	-1.108	**
					(-8.132)		(-7.938)		(4.700)	
Italian							46.564	***	33.852	***
							(9.759)		(6.930)	
Spanish							10.856	***	11.446	***
							(3.344)		(3.671)	
Swedish									80.606	***
									(25.049)	
Proficiency (Quadratic):										
English			-3.580	***	-4.481	***	-4.580	***	-3.690	***
			(-8.600)		(-9.879)		(-9.731)		(8.292)	
French					-1.552	***	-1.712	***	-0.872	**
					(-3.178)		(-3.491)		(1.876)	
German					3.230	***	3.172	***	1.571	***
					(7.235)		(7.453)		(4.878)	
Italian							-748.687	***	-461.089	***
							(-7.864)		(4.906)	
Spanish							-75.874	**	-52.094	
							(-2.038)		(1.431)	
Swedish									-857.98	***
									(21.793)	
N	1470		1470		1470		1470		1470	
Adjusted R^2	0.972		0.975		0.977		0.978		0.983	

Note: See Table 1.

Table 5: Trade effects of Foreign Languages, Non-Linear Specification, NMS and Associated Countries (including Turkey)

Variable	**(1)** teu_td0		**(2)** teu_td1		**(3)** teu_td2		**(4)** teu_td4	
Intercept	20.193	***	19.176	***	17.181	***	17.682	***
	(14.09)		(10.30)		(8.29)		(7.92)	
GDP	0.642	***	0.701	***	0.765	***	0.763	***
	(8.51)		(7.53)		(8.31)		(8.24)	
Distance	-1.039	***	-0.994	***	-0.809	***	-0.873	***
	(5.99)		(6.11)		(4.18)		(4.02)	
Former Federation	2.278	***	2.330	***	1.367	***	1.399	***
	(11.11)		(11.79)		(12.69)		(12.31)	
Contiguity	0.542	***	0.542	***	0.643	***	0.538	***
	(4.64)		(4.98)		(5.38)		(3.86)	
Proficiency:								
English			-0.861		3.002	**	2.928	**
			(0.19)		(0.67)		(0.66)	
German					6.571	**	8.82	**
					(0.38)		(0.51)	
Russian					1.632	***	1.185	***
					(1.80)		(1.30)	
Hungarian							19.328	***
							(3.56)	
Proficiency (Quadratic):								
English			13.504		5.293		5.317	
			(1.54)		(0.60)		(0.60)	
German					143.128		42.076	
					(0.46)		(0.13)	
Russian					3.833	**	4.245	*
					(2.78)		(3.18)	
Hungarian							-128.053	***
							(3.49)	
N	1254		1254		1254		1254	
Adjusted R^2	0.847		0.850		0.857		0.858	

Note: See Table 1.

Table 6: Trade effects of Foreign Languages, Non-Linear Specification, All Countries

Variable	(1) tall_td0		(2) tall_td1		(3) tall_td2		(4) tall_td3		(5) tall_td4	
Intercept	19.534	***	19.25	***	19.283	***	19.146	***	19.166	***
	(34.83)		(31.59)		(30.41)		(28.74)		(28.72)	
GDP	0.858	***	0.867	***	0.864	***	0.871	***	0.868	***
	(36.76)		(35.82)		(33.95)		(33.69)		(33.71)	
Distance	-1.098	***	-1.090	***	-1.089	***	-1.085	***	-1.07	***
	(18.61)		(18.03)		(16.51)		(15.73)		(15.37)	
Former Federation	2.366	***	2.359	***	1.951	***	1.98	***	2.077	***
	(23.43)		(23.52)		(22.44)		(22.28)		(23.27)	
Contiguity	0.282	***	0.286	***	0.296	***	0.26	***	0.152	***
	(4.36)		(4.42)		(4.58)		(3.69)		(2.11)	
Official languages										
English	0.842	***	0.713	***	0.726	***	0.742	***	0.716	***
	(4.70)		(2.97)		(2.95)		(3.01)		(2.90)	
German	0.647	***	0.648	***	1.23	**	1.255	***	1.300	***
	(6.90)		(6.89)		(3.81)		(3.88)		(4.22)	
French	0.129		0.143		0.198		0.267		0.352	
	(0.70)		(0.77)		(0.34)		(0.46)		(0.61)	
Swedish	0.051		0.063		0.049	*	0.083		17.446	***
	(0.65)		(0.77)		(0.56)		(0.90)		(6.96)	
Dutch	-0.676	***	-0.685	***	-0.706	***	-0.651	***	-0.521	***
	(7.55)		(7.23)		(6.79)		(5.90)		(4.56)	
Greek	2.006	***	2.020	***	2.011	***	2.032	***	2.086	***
	(10.44)		(10.49)		(10.31)		(10.70)		(11.93)	
Proficiency:										
English			0.700		0.645		0.656		0.026	
			(1.12)		(0.99)		(1.00)		(0.04)	
French					-0.245		-0.335		-0.268	
					(0.25)		(0.33)		(0.27)	
German					0.216		0.345		0.799	
					(0.30)		(0.46)		(1.08)	
Russian					1.407	*	1.227		1.196	
					(1.70)		(1.44)		(1.39)	
Italian							4.799	**	6.664	***
							(2.10)		(2.91)	
Spanish							-1.337		-1.144	
							(0.28)		(0.24)	
Swedish									77.224	***
									(8.94)	
Hungarian									37.745	***
									(9.24)	
Proficiency (Quadratic):										
English			-0.229		-0.249		-0.248		0.326	
			(0.36)		(0.37)		(0.37)		(0.48)	
French					0.155		0.246		0.208	
					(0.15)		(0.23)		(0.20)	
German					-0.943	*	-1.062		-1.478	

			(0.93)	(1.03)	(1.47)	
Russian			-0.096	0.168	0.216	
			(0.08)	(0.13)	(0.17)	
Italian				-11.026	* 6.664	**
				(1.65)	(2.91)	
Spanish				76.959	96.071	
				(1.23)	(1.53)	
Swedish					-865.709	***
					(7.78)	
Hungarian					-228.74	***
					(9.37)	
N	2411	2411	2411	2411	2411	
Adjusted R^2	0.933	0.933	0.933	0.933	0.935	

Note: See Table 1.

5 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 relatively 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 regression. The median regression is frequently used in regression analysis which may be biased by outliers. While the least squares regression estimates 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 regression is more robust than with ordinary regression. While this concept is now frequently used in economics, especially in labor and family economics (see literature survey by Koenkeer and Hallock, 2001), it has found little application in trade analysis so far (see Wagner, 2006).

For simplicity, we use a parsimonious version of our gravity model specified only with linear communicative probability in English as well as a dummy for English official language. We thus estimate the following linear model for the τ^{th} conditional quantile, Q, of bilateral trade volume, T,

$$Q_{\tau}\left(T_{ijt}\right) = \alpha_{\tau} + \theta_{\pi} + \beta_{\tau 1}\left(y_{it} + y_{jt}\right) + \beta_{\tau 2} d_{ij} + \beta_{\tau 3} b_{ij} + \delta_{\tau} DL_{eng,ij} + \delta_{\tau} FL_{eng,ij} + \varepsilon_{ijt} . \qquad (2)$$

Table 7 reports the results for the 10^{th}, 25^{th}, 75^{th} and 90^{th} percentiles in addition to the median regression. The standard errors are simulated in a bootstrap procedure with 1000 repetitions. We can see that the effects of all gravity variables differ significantly between the individual quartiles. The income elasticity declines as bilateral trade increases. In turn, the transport (distance) elasticity increases slightly in absolute terms with trade volume, while the effect of contiguity tends rather to decrease with trade

volume. The test of equal coefficients for the first to third quartiles (see the last column) clearly rejects the null at the standard significance levels for all explanatory variables with the exception of distance.

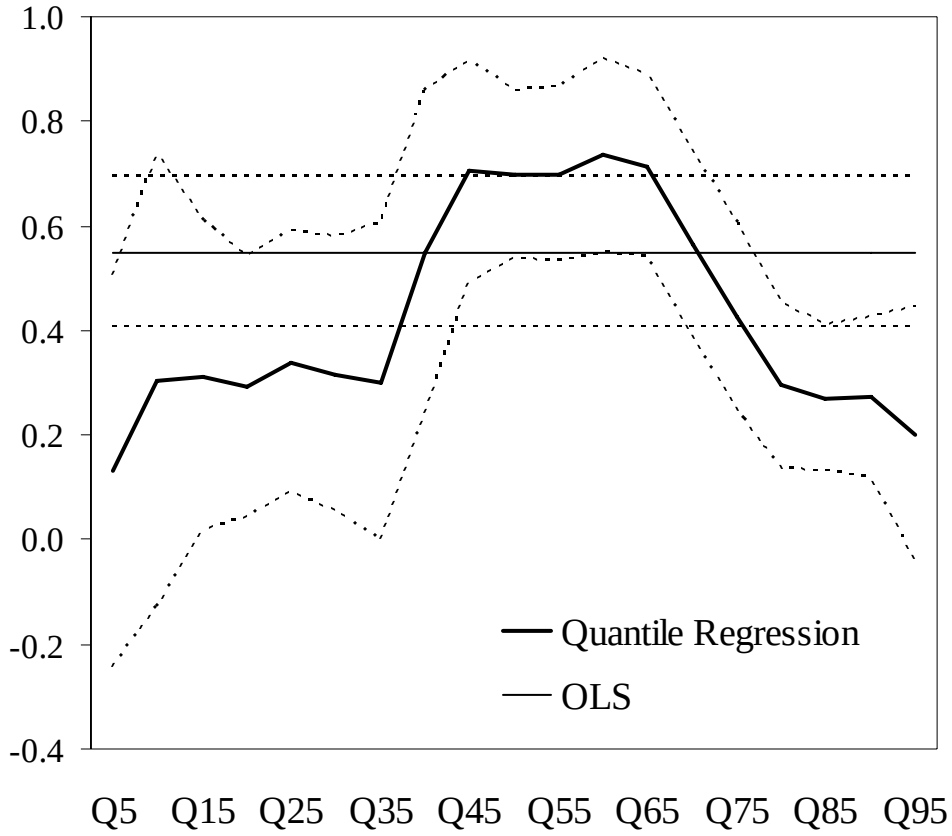
The effects of proficiency in English show an interesting non-monotonic behavior. We find that the effect is the highest in the median regression. This confirms that our previous findings are not due to outliers. There is also slight asymmetry in the coefficients showing that trade gains are higher for countries with higher trade intensity (compare the 25th and 75th percentile). The estimated coefficients are significant with the exception of the 10th-percentile. In turn, the coefficient for 10th percentile are larger that that for 90th percentile, although the former is insignificant. More detailed analysis in Figure 5 conducted for each fifth percentile confirms this pattern.

Table 7: Trade effects of Proficiency in English, Quantile Regression, EU Trade

	OLS	**Q10**	**Q25**	**Q50**	**Q75**	**Q90**	**Test[a]**
Income	0.895 ***	0.962 ***	0.931 ***	0.874 ***	0.836 ***	0.795 ***	26.15
	(103.738)	(71.456)	(108.087)	(76.435)	(61.359)	(89.185)	[0.00]
Distance	-0.694 ***	-0.464 ***	-0.695 ***	-0.709 ***	-0.787 ***	-0.852 ***	0.94
	(-25.615)	(-9.416)	(-15.490)	(-31.303)	(-11.848)	(-23.392)	[0.39]
Contiguity	0.643 ***	0.673 ***	0.483 ***	0.687 ***	0.591 ***	0.319 ***	7.06
	(16.079)	(10.082)	(7.760)	(14.501)	(6.632)	(6.246)	[0.00]
Official lang.	0.488 ***	1.088 ***	0.890 ***	0.433 **	0.426 ***	0.400 ***	5.10
(English)	(5.256)	(4.933)	(6.713)	(2.423)	(3.306)	(3.291)	[0.01]
Proficiency	0.549 ***	0.304	0.340 ***	0.697 ***	0.426 ***	0.272 ***	9.46
(English)	(7.495)	(1.374)	(2.737)	(8.308)	(4.502)	(3.583)	[0.00]
Intercept	-21.313 ***	-27.083 ***	-23.557 ***	-20.109 ***	-17.209 ***	-14.193 ***	22.42
	(0.533)	(-29.485)	(-41.532)	(-30.307)	(-15.797)	(-30.914)	[0.00]
N	1800	1800	1800	1800	1800	1800	1800
Pseudo R^2	0.918	0.738	0.735	0.722	0.716	0.714	ND

Note: Time dummies are not reported. *t*-statistics (in parentheses) are computed using bootstrap standard errors with 1000 replications. ***, **, and * denote significance at 1 per cent, 5 per cent, and 10 per cent, respectively. a – Test of equal coefficients for the first to third quartiles. p-values in brackets.

Figure 5: OLS and Quantile Regression Estimates for Proficiency in English



Note: For quantile regression estimates, the 95% confidence bands are computes on the base of bootstrap standard errors with 1000 replications. Heteroscedasticity robust standard errors are used for the OLS estimates.

6 Conclusions

Our analysis finds strong effects of languages on bilateral trade. Besides confirming that countries that share the same official language tends to trade significantly more with each other, we also shed light on the effect of foreign languages (i.e. languages that people do not speak because they are native speakers but instead they have to learn them). Our results suggest that English plays a particularly important role, both because it is the most widely spoken foreign language and because, unlike the other languages, its effect appears robust to alternative regression specifications. Our findings also suggest that the effect of English and other languages on trade flows may be non-linear, displaying diminishing returns.

Nevertheless, the gains from foreign languages are not uniform across countries: our analysis suggests that the effect is different in the EU15 compared to the new member states and candidate countries. This heterogeneity is likely due to the different history of integration and different economic, political and linguistic legacies in the two sets of countries. Further research will show to what extent we can find evidence of convergence or divergence in the effect of languages.

In the past decade or two, trade has become a powerful argument in favor of deepening European integration, including introducing the common currency, the euro. Our findings suggest that gains of similar magnitude could be realized by improving linguistic skills, especially in English. 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 English does not require abandoning national languages. Substantial gains are available at relatively little cost: encouraging the learning of English could well, metaphorically, allow countries to pick up 100$ bills lying on the sidewalk.

Last but not least, our results illustrate the predominance of English as, effectively, the lingua franca in Europe. While individuals may derive private benefits from learning marginal languages, countries only benefit inasmuch as the same language is learned by other individuals in other countries. English, at present, is the only language spoken by enough people to have an economically significant effect on trade flows.

References

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

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

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.

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, Pfaffermayr M (2004) Distance, trade and FDI: A Hausman-Taylor SUR approach. Journal of Applied Econometrics 19: 227-246.

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

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.

Kimura F, Lee H-H (2006) The gravity equation in international trade in services. Review of World Economics 142: 92-121.

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

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

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.

Mélitz, J. (2008), "Language and Foreign Trade," *European Economic Review* 52 (4), 667-699.

Rauch, J.E., Trindade, V. (2002), "Ethnic Chinese Networks in International Trade," *Review of Economics and Statistics* 84 (1), 116–130.

Wagner, Joachim (2006) Export Intensity and Plant Characteristics: What Can We Learn from Quantile Regression?, Review of World Economics (Weltwirtschaftliches Archiv), 142(1), 195-203.

APPENDIX

Figure 6: Predicted Effect of Communicative Probability, EU15 Countries

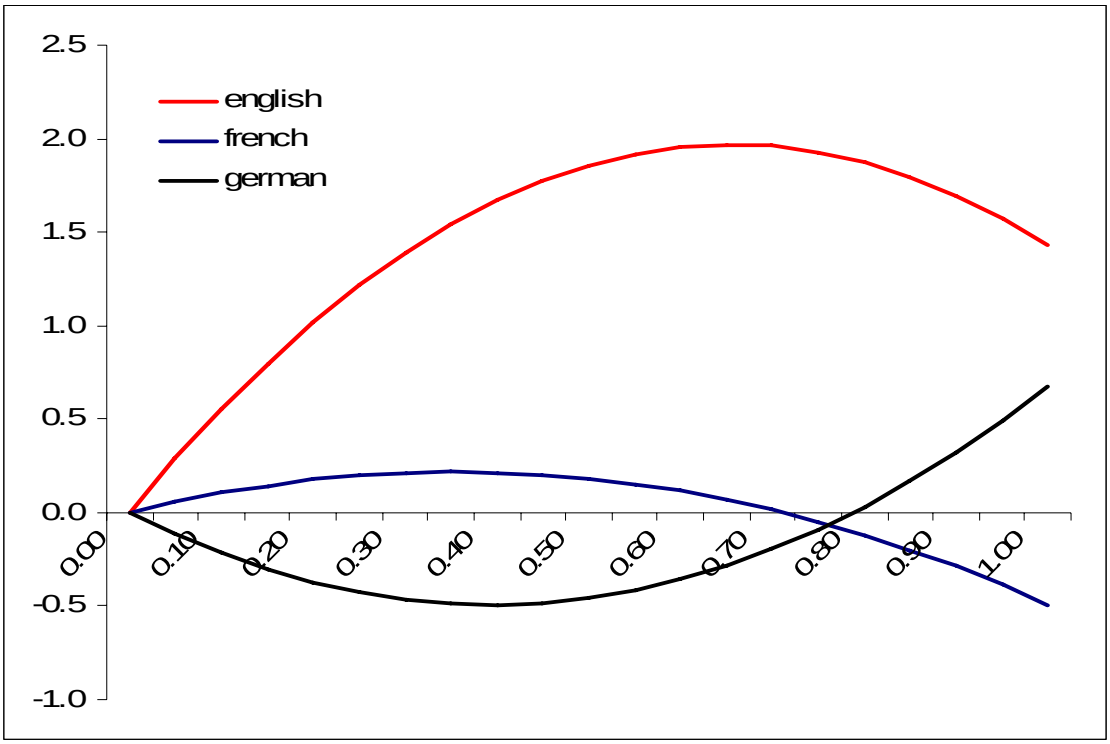


Figure 7: Predicted Effect of Communicative Probability, NMS/AC Countries

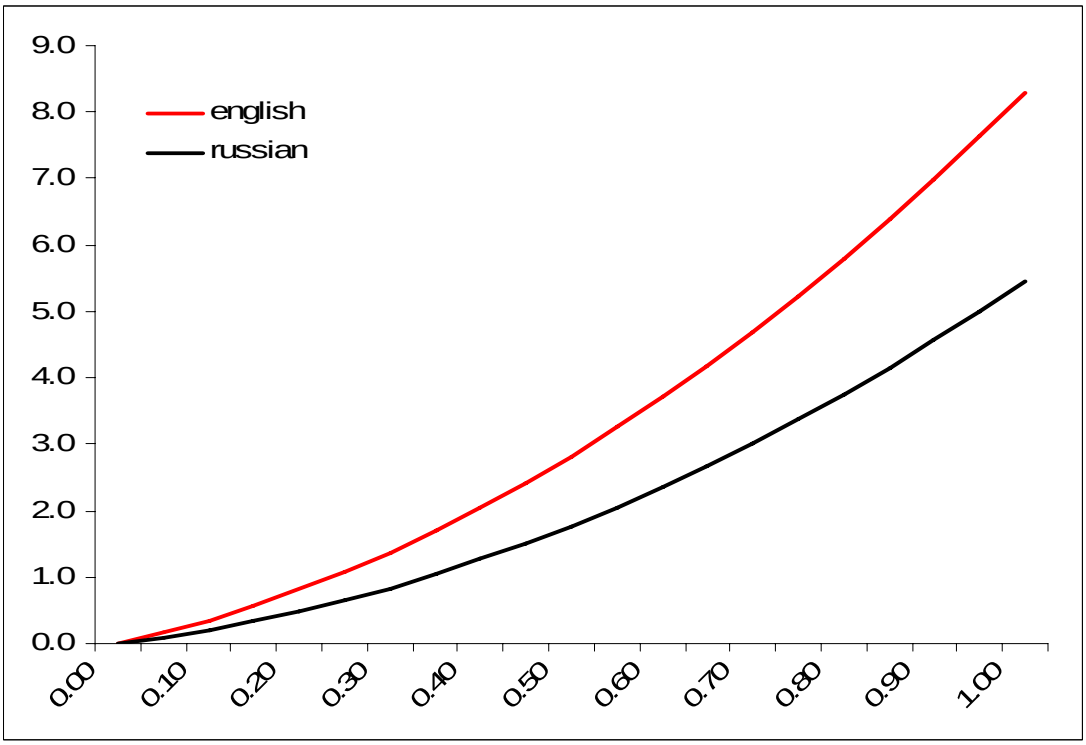


Figure 8: Predicted Effect of Communicative Probability, All Countries