PRODUCT COMPLEXITY, QUALITY OF INSTITUTIONS AND THE PRO-TRADE EFFECT OF IMMIGRANTS

Anthony Briant, Pierre-Philippe Combes, Miren Lafourcade

GREQAM

Groupement de Recherche en Economie
Quantitative d'Aix-Marseille - UMR-CNRS 6579
Ecole des Hautes Etudes en Sciences Sociales
Universités d'Aix-Marseille II et III

Document de Travail
n°2009-08

PRODUCT COMPLEXITY, QUALITY OF INSTITUTIONS AND THE PRO-TRADE EFFECT OF IMMIGRANTS

Anthony BRIANT
Pierre-Philippe COMBES
Miren LAFOURCADE

February 2009

Product Complexity, Quality of Institutions and the Pro-Trade Effect of Immigrants *

Anthony Briant† Pierre-Philippe Combes‡ Miren Lafourcade§

February 27, 2009

Abstract

The paper assesses the trade-creating impact of foreign-born residents on the international imports and exports of the French regions where they are settled. The pro-trade effect of immigrants is investigated along two intertwined dimensions: the complexity of traded goods and the quality of institutions in partner countries. The trade-enhancing impact of immigrants is, on average, more salient when they come from a country with weak institutions. However, this positive impact is especially large on the imports of simple products. When we turn to complex goods, for which the information channel conveyed by immigrants is the most valuable, immigration enhances imports regardless of the quality of institutions in the partner country. Regarding exports, immigrants substitute for weak institutions on both simple and complex goods.

JEL classification: F14, F22, R12.
Keywords: trade, immigration, quality of institutions, product complexity, gravity.

*This paper has been prepared as part of the PREDIT research contract 04-MT-5036. Financial supports from PREDIT and from CNRS (Combes) are gratefully acknowledged. Authors are also involved in the Marie Curie Research Training Network 'TOM - Transnationality of migrants' funded by the European Commission (contract: MRTN-CT-2006-035873). We are especially grateful to Alain Sauvant and Micheline Travet for providing us with trade data, as well as to Alexandre Kych and Thierry Mayer for precious help in gathering immigration and institution data. We are indebted to Laurent Gobillon, Farid Toubal, Muriel Barlet and Rodrigo Paillacar for insightful comments. We also thank the participants of the CEPR Conference on "Trade, Industrialisation and Development" (Paris, October 2006), of the first CEPR-TOM Conference (Riga, October 2007), and of the RIEF doctorial meeting (Rennes, February 2007), for useful feedbacks.

†Corresponding Author: Paris School of Economics / Paris-Jourdan Sciences Économiques (UMR 8545 CNRS-EHESS-ENPC-ENS). PSE, 48 Boulevard Jourdan, 75014 Paris, France. Email: anthony.briant@m4x.org. Fax: +33(0)143136310.

‡GREQAM-University of Aix-Marseille, PSE and CEPR. Centre de la Vieille Charité, 2 rue de la Charité, 13236 Marseille cedex 02, France. ppcombes@univmed.fr. http://www.vcharite.univ-mrs.fr/pp/combes/.

§University of Valenciennes and PSE. 48 Boulevard Jourdan, 75014 Paris, France. lafourcade@pse.ens.fr. http://www.enpc.fr/ceras/lafourcade/.

1 Introduction

Despite the widespread availability of modern communication technologies, information costs still play a crucial role in shaping world trade patterns. As surveyed by Anderson and Wincoop (2004), these costs, equivalent to an ad-valorem tax of 6%, largely account for the puzzling persistence of distance and border impediments to trade.

According to Rauch (2001), social and business transnational networks are likely to alleviate some of these information failures. Cross-border networks are prone to substitute for organized markets in matching international buyers and sellers, and this is especially true of differentiated products. In this respect, co-ethnic networks are of more particular interest, as illustrated for instance by the model of Casella and Rauch (2003). Immigrants' ties to their home country may promote trade for at least three reasons. First, immigrants have a good knowledge of the customs, language, laws as well as business practices in both the host and home countries. Accordingly, their presence helps bridging the information gap between sellers and buyers on both sides, hence promoting bilateral trade opportunities. Second, immigrant networks may provide contract enforcement through sanctions and exclusions, which substitutes for weak institutional rules and reduces trade costs. In addition to the two previous channels, immigrants bring their taste for homeland products, which should make their trade-creating impact even more salient on imports.

In this paper, we provide new evidence on the relationship between trade and immigration building on regional data for France. We investigate the pro-trade effect of foreign-born French residents on the exports and imports of French *départements* with around 100 countries in the world. The novelty consists in crossing the effect of immigration with both the quality of institutions in the home country and the complexity of traded goods.

The trade-promoting effect of immigration is now well documented (see Wagner, Head, and Ries (2002) for an extensive review). Gould (1994), Head and Ries (1998) and Girma and Yu (2002) find a significant trade-creating impact of immigrants settled in the United States, Canada, and the United Kingdom respectively. Rauch and Trindade (2002) exhibit a diaspora-network rationale ruling this pro-trade phenomenon by showing that South-Asian country pairs with a higher proportion of Chinese immigrants trade more with each other.

However, there are many reasons to suspect that, at the country level, the correlation between trade and immigration might arise from omitted common determinants (such as colonial ties, language or cultural proximity), or reverse causality if immigrants prefer to settle in countries that have good trade relationships with their home country.

Accordingly, a few recent attempts investigate the link between the spatial patterns of trade and immigrants' settlements within countries. Wagner, Head, and Ries (2002) are the first to test a causal relationship between trade and immigration at the scale of Canadian provinces. The inclusion of country fixed-effects allows to control for the common determinants of trade and immigration at the national level. At the same time, cross-sectional variability in trade and immigration at the regional level provides sufficient information to identify the pro-trade effect of immigrants. The authors confirm the positive and significant elasticity of trade with respect to immigration, at the regional level.

Further evidence is provided for the US states exports. Herander and Saavedra (2005) disentangle the impact of both the in-state and out-state stocks of immigrants. The outstanding impact of in-state immigrants pinpoints the key role of local social interactions as a major source of technological externalities. Building on the same previous data set, Dunlevy (2006) further shows that the pro-trade effect of immigrants increases with the degree of corruption and with language similarity in the partner country. Finally, Bandyopadhyay, Coughlin, and Wall (2008) explore the temporal scope of the data and regress the 1990-2000 time variation in trade on the related time variation in immigrants' settlements. This approach bears the advantage of controlling for pair-specific unobserved characteristics. The pro-trade effect of immigrants is found to exhibit a large heterogeneity driven by a few countries only. In a related strand of literature, Combes, Lafourcade, and Mayer (2005) for France and Millimet and Osang (2007) for the US show that within-country migrations affect positively the volume of inter-regional trade flows.

Our paper extends this literature in three directions. First, the relationship between trade and immigration is studied at a lower geographical scale than any previous North-American study. French *départements* are almost 30 times tinier than American states and more than 100 times smaller than Canadian provinces. A spurious correlation between trade and immigration is thus less likely to occur at this very fine geographical scale. We do find that immigration exerts a significant positive impact on trade: doubling the number of immigrants settled in a *département* boosts its exports to the home country by 7% and its imports by 4%.

Second, we address econometric questions endemic to gravity-type estimations. We first tackle the issue of specification and selection biases due to zero flows, by resorting to the Quasi-Maximum Likelihood estimator recently proposed by Head, Mayer, and Ries (2008b). We then turn to the bias arising from possibly omitted common determinants for immigration and trade or from reverse causality. To circumvent both sources of endogeneity, we include country- and region-specific fixed-effects in the regression, and we resort to an instrumental variables approach, where lagged stocks of foreign-born French residents serve as instruments. The previous orders of magnitude remain astonishingly robust to these econometric refinements.

Finally, we evaluate the heterogeneous impact of immigrants along two intertwined dimensions: the complexity of traded goods and the quality of institutions in the partner country. Indeed, Rauch and Trindade (2002) show that the trade-creating effect of Chinese networks is larger for differentiated products than for homogeneous or reference price goods. The fact that immigrants matter more for differentiated goods can be taken as a support for the information-cost-saving channel of transnational networks. Besides, Anderson and Marcouiller (2002) and Berkowitz, Moenius, and Pistor (2006) show that the quality of institutions impacts drastically on the volume of bilateral trade. Berkowitz, Moenius, and Pistor (2006) point out that the quality of institutions matters more for complex commodities, which exhibit characteristics difficult to fully specify in a contract. This is the reason why good institutions may reduce transaction costs when contracts are more incomplete. However, they do not study whether transnational networks could be a substitute for weak institutions, especially in the trade of complex products, as suggested by Rauch (2001).[1]

Building on these insights, we disentangle the pro-trade impact of immigrants across both the partner's institution quality and the complexity of traded goods. In this respect, we emphasize two main results. First, immigrants especially matter for the imports of complex goods, regardless of institution quality in the home country. Turning to the imports of simple products, immigrants matter only when the quality of institutions at home is weak. Second, the trends are less marked for exports. The pro-trade impact of immigrants on exports is positive only when they come from countries with weak institutions, regardless of the complexity of products.

The remainder of the paper proceeds as follows. Section 2 presents the augmented-gravity specification we use to evaluate the trade-creating impact of foreign-born French residents, and discusses several econometric issues. It also describes the trade and immigration data for French regions. Section 3 presents the benchmark empirical results. Section 4 disentangles the trade-creating impact of immigration across simple or complex products, and across countries with different quality of institutions. Section 5 concludes.

[1] In this respect, Dunlevy (2006) is a noticeable exception. He shows that the impact of immigrants on US state exports is more important when institutions in the home country are weak.

2 Model specification, econometrics and data

To investigate the pro-trade effect of social networks, we need a benchmark to evaluate the amount of trade expected absent any immigrants' settlements. Following Combes, Lafourcade, and Mayer (2005), we present the gravity norm we use to provide this benchmark. This section also discusses some of the econometric pitfalls traditionally encountered in gravity estimations. The following presentation draws on the exposition of Head, Mayer, and Ries (2008a).

Model specification

The rationale behind the gravity model is that the value of trade between two locations (y_{ij}) is generated by the adjusted economic sizes of both the supplying location i (S_i) and the demanding location j (M_j), and inhibited by all the sources of "trade resistance" between them (ϕ_{ij}):

$$y_{ij} = GS_iM_j\phi_{ij} \tag{1}$$

G is a factor that does not vary across regions. Head, Mayer, and Ries (2008a) refer to S_i and M_j as the monadic terms, and ϕ_{ij} as the dyadic term. The usual practice is to log-linearize this equation and to find proxies for the monadic and dyadic terms:

$$\ln y_{ij} = \ln G + \ln S_i + \ln M_j + \ln \phi_{ij} \tag{2}$$

Anderson and Wincoop (2003) provide clear-cut theoretical micro-foundations for the monadic terms: they depend on nominal economic size (for instance GDP), but also on non-linear functions of all pairwise dyadic terms, called the "Multilateral resistance Indices" (hereafter MRIs). A proper control for these monadic terms in gravity estimations is challenging.[2] The primary question we focus on is whether the spatial distribution of immigrants coming from a country j affects the trade of hosting *départements* with such country. Hence, we are not interested in the country- or *département*-specific determinants of trade. This is the reason why we adopt a fixed-effect approach *à la* Anderson and Wincoop (2003), and introduce two sets of dummies in the gravity equation. The inclusion of country fixed-effects (f_j) is meant to control for all standard

[2]Head, Mayer, and Ries (2008a) give a clear review of the state-of-art on the econometric specification of the gravity equation. Four solutions are encountered in the literature: 1/ a non-linear approach, proposed by Anderson and Wincoop (2003), where MRIs are explicitly computed, 2/ a fixed-effect approach, also proposed by Anderson and Wincoop (2003), where monadic terms are controlled for by a set of importer and exporter dummies, 3/ the *bonus vetus OLS* approach, proposed by Baier and Bergstrand (2007) and recently adapted by Behrens, Ertur, and Koch (2007) based on spatial econometrics, where first-order Taylor expansions of MRIs are introduced in the specification, and 4/ the *tetrad* approach, proposed by Head, Mayer, and Ries (2008a), where monadic terms are suppressed thanks to the computation of export ratios.

country-specific determinants of trade: membership to a common trade or currency bloc (e.g. the Euro Zone or the European Union), landlocked nature, colonial ties or common languages. The other set of dummies (f_i) controls for the *département*-specific determinants of trade, such as the density of economic activity or any natural or man-made endowments. Finally, it is worth noting that, in this two-way fixed-effect setting, only the dyadic determinants (ϕ_{ij}) of bilateral trade can be identified.

Regarding this dyadic term, we follow Combes, Lafourcade, and Mayer (2005) and assume that trade costs do not only depend on distance and contiguity. They are also inversely correlated with the number of immigrants coming from country j settled in region i. We choose ϕ_{ij} as a multiplicative function of (i) the great circle distance between i and j, (ii) a dummy indicating whether or not the *département* and the country are contiguous,[3] and finally (iii) the stock of foreign-born residents in i originating from country j, mig_{ij}:

$$\phi_{ij} = \text{dist}_{ij}^{\beta}(1 + \text{mig}_{ij})^{\alpha} \exp(\gamma \text{contig}_{ij}) \tag{3}$$

We add an error term (ϵ_{ij}) that controls for all unobservable dyadic terms uncorrelated with distance, contiguity or immigrants' stock. The baseline specification we estimate is thus the following two-way fixed-effect log-linearized equation:

$$\ln y_{ij} = f_i + f_j - \beta \ln \text{dist}_{ij} + \gamma \text{contig}_{ij} + \alpha \ln \left(1 + \text{mig}_{ij}\right) + \epsilon_{ij}. \tag{4}$$

In what follows, we estimate this specification for exports and imports separately.

Econometric issues

Three major econometric problems are usually encountered when estimating gravity models. The first problem deals with the treatment of zero flows. The log-linearized specification (4) can only be estimated on strictly positive flows. Various methodologies have been proposed to control for the selection bias arising from keeping positive flows only. Dunlevy (2006) takes the logarithm of one plus the value of the flow as a dependent variable. He also estimates a Tobit model with an arbitrary zero threshold. Herander and Saavedra (2005) use the extended Tobit estimation first proposed by Eaton and Tamura (1994), where the threshold is estimated. This technique, also used by Wagner, Head, and Ries (2002), rests on a maximum likelihood estimation of the log-linearized model.

[3]This dummy is equal to one for only a small subset of *départements* contiguous to Belgium/Luxembourg, Germany, Switzerland, Italy or Spain.

A second issue concerns the heteroskedasticity of error terms in levels. In theoretical models indeed, the gravity equation takes a multiplicative form, as in specification (1): hence, if the error term in levels is heteroskedastic, OLS estimates for the log-linearized model are biased.[4] To tackle simultaneously the zero-flow and the heteroscedastic issues, Santos Silva and Tenreyro (2006) initiated a novel approach by estimating the gravity equation in levels. They propose a easy-to-implement Quasi-Maximum Likelihood (hereafter QML) estimation for the gravity equation, under the assumption that error terms in levels are distributed according to a Poisson distribution. These authors find that the elasticity of trade flows to distance is twice as small as the one estimated from OLS. However, the Poisson specification builds on the assumption that conditional variance equals conditional mean in the data, $\mathbb{V}(y_{ij}|x_{ij}) = \mathbb{E}(y_{ij}|x_{ij})$. Head, Mayer, and Ries (2008b) provide a more robust 2-step Negative Binomial (hereafter 2NB) procedure that allows the conditional variance to be a quadratic function of the mean, $\mathbb{V}(y_{ij}|x_{ij}) = \mathbb{E}(y_{ij}|x_{ij}) + \eta^2\mathbb{E}(y_{ij}|x_{ij})^2$.[5] Hence, in what follows, we compare baseline OLS and 2NB estimates in order to test whether the pro-trade effect of immigrants is robust to these two presumably important biases: zero flows and heteroskedasticity in levels.

The third issue is endogeneity, which may arise from two major sources: omitted variables and reverse causality. At the national scale, one can imagine that preferential links between two countries (resulting from a common colonial history for instance) generate simultaneously trade and immigrant flows. Furthermore, the existence of a strong trade partnership may push people to migrate, creating a reverse causality between trade and immigration. Gould (1994) provides two reasons to believe that cross-section estimations actually preclude the endogeneity bias, at the national level. First, migrations are expected to be more exogenous than trade flows, because they are determined by family reunifications in the first place. As recently analyzed by Thierry (2004), this is also a plausible explanation for France. Second, in addition to family entrance motivations, immigration inflows are conveyed by wage differentials and the pre-existence of a same native/speaking community, rather than by trade opportunities. This is also what suggests the analysis conducted by Bartel (1989) or Munshi (2003) for the US, and by Jayet and Bolle-Ukrayinchuk (2007) for France.

[4]This is due to Jensen's inequality, according to which the expected value of the logarithm of a random variable is not equal to the logarithm of the expected value of this variable. Furthermore, the expected value of the logarithm of a random variable depends not only on the expected value of the variable, but also on the other moments of its distribution, especially the variance. Under heteroskedasticity in levels, this variance is a function of explanatory variables, which generates endogeneity in the log-linearized model.

[5]Gourieroux, Monfort, and Trognon (1984) show that QML estimators are consistent as long as the expected value of the dependent variable is well specified, and thus robust to an error in the specification of the true data generating process for the error term. See Cameron and Trivedi (2005) for further details.

Furthermore, these two sources of endogeneity are partially mitigated when we turn to infra-national data. In specification (4), the country- and region-fixed effects control for a large set of common observable and unobservable determinants for trade and immigration flows. Nevertheless, it could be argued that reverse causality and omitted variables are still likely to prevail at the infra-national level. To be sure that this relationship is not driven by omitted variables, Wagner, Head, and Ries (2002) control for the commonality of language, i.e. the probability that a random citizen of a given region speaks the same language as a random citizen of the trading partner. We cannot compute such a variable in the French case. We follow another route and instrument the current stock of immigrants with past stocks in 1975, 1982 and 1990. These lagged stocks are valid instruments as long as they determine the current stock of immigrants, and do not determine current trade flows beyond their effect on the current stock of immigrants. We provide further support for this view in what follows. The instrumental variables approach has been rarely implemented in the literature.[6]

Data

Trade data consists in the exports and imports of 94 French metropolitan *départements* with around 100 countries. French decentralized customs services record the value of trade flows exclusive of transit shipments, as well as the origin/destination of shipments, i.e. those where goods are actually produced/consumed. Although trade values are available since 1978, we focus exclusively on the recent period to ensure data compatibility with immigrants' stocks. Furthermore, in order to prevent noisy observations due to time-specific shock (as the euro adoption), we average trade flows over three years (1998, 1999 and 2000) for each *département*-country pairs.

Trade flows are originally available at a very disaggregated industrial level, according to the Standard Goods Classification for Transport Statistics (NST/R classification). We match this classification with the one proposed by Rauch (1999) to characterize the complexity or the degree of differentiability of products.[7]

The 1999 French population census provides us with exhaustive information on the number of foreign-born residents by *département* and country pairs. We define immigrants as residents born abroad with a foreign nationality. In the empirical part, we also use the lagged stocks of immigrants to tackle the endogeneity issue. These figures are provided by French population censuses for the years 1975, 1982 and 1990. Appendix A provides further details on exports,

[6]Combes, Lafourcade, and Mayer (2005) stands as an exception.
[7]See Appendix B for details.

imports and immigration data.

It is worth stressing that most of the variability in the data comes from the cross-country dimension of the sample. For instance, the regression of trade flows on country-specific dummies returns an adjusted-R^2 of 51% for exports, 61% for imports and 70% for immigration.

We wipe out this cross-country variation with a set of country fixed-effects. We also include *département* dummies to control for the common observable or unobservable determinants of trade and immigrants inside France.

Due to the introduction of these two sets of dummies, the pro-trade impact of immigrants is identified along the within-country and within-*département* data variability. Table 1 depicts the within-country and within-*département* correlation between exports, imports, distance and immigration.[8] As expected, distance is negatively correlated with exports and imports, the correlation being stronger for imports. By way of contrast, immigration is significantly and positively correlated with both exports and imports. Distance and immigration are also negatively correlated, as it is well known that immigration flows also share a gravity pattern. Appendix A provides further summary statistics on the data.

Table 1: *Within-country, within-département* correlations

Variables	Exports	Imports	Distance	Immigrants
Exports	1.000			
Imports	0.144	1.000		
Distance	-0.090	-0.137	1.000	
Immigrants	0.066	0.043	-0.090	1.000

Note: All correlations are significant at the 1% level.

3 The pro-trade effect of immigrants

3.1 Benchmark results

Table 2 provides the basic results drawn from estimating specification (4). In columns labeled OLS, we report the results drawn from the log-linear form (null flows are left out of the sample). We also estimate the same specification in levels (columns 2NB). We run this regression twice: first on the sample restricted to positive flows (columns (3) and (7)), and second on the whole sample (columns (4) and (8)). We run two sets of regressions for exports and imports separately.

[8] More formally, this is the correlation between the residuals of the regression of each variable on country-specific and *département*-specific dummies.

Table 2: Benchmark results

	Exports				*Imports*			
	OLS	OLS	2NB >0	2NB ≥ 0	OLS	OLS	2NB >0	2NB ≥ 0
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Distance	-0.81^{a}	-0.777^{a}	-0.963^{a}	-0.961^{a}	-1.488^{a}	-1.480^{a}	-1.612^{a}	-1.638^{a}
	(0.089)	(0.085)	(0.1)	(0.104)	(0.128)	(0.127)	(0.143)	(0.157)
Contiguity	0.452^{a}	0.273^{c}	0.123	0.099	0.445^{b}	0.342^{c}	0.029	-0.0009
	(0.167)	(0.163)	(0.163)	(0.169)	(0.198)	(0.201)	(0.205)	(0.237)
Immigrants		0.102^{a}	0.091^{a}	0.109^{a}		0.054^{b}	0.094^{a}	0.089^{b}
		(0.018)	(0.019)	(0.021)		(0.027)	(0.035)	(0.041)
Obs.	9033	9033	9033	9400	8110	8110	8110	9494
Adj. R^2	0.844	0.844			0.8	0.8		

Note: Country and *département* fixed effects are not reported here. Robust standard errors in brackets, with ^{a}, ^{b} and ^{c} denoting significance at the 1%, 5% and 10% levels respectively.

Log-linear specification

In columns (1) and (5), trade impediments are proxied by distance and contiguity only. Elasticities have expected signs. Exports, as well as imports, decrease with distance and increase with contiguity. The elasticity of imports to distance is twice larger than that of exports. Although there is not any obvious reason for such a phenomenon, it is worth recalling that, in this two-way fixed-effect setting, elasticities are estimated on the within-variability of the data. Hence, identification relies drastically on close countries for which distance differentials across regions remain high in comparison with countries located further away. For instance, Paris and Marseille are almost equally distant from the United States, but not from Germany. For more distant countries, the variability in distance is reduced. Nevertheless, the variability in trade flows remains fairly high: a small difference in distance can be associated with a large difference in trade values.

In columns (2) and (6), we add the stock of immigrants in the specification in logs. Contrary to most of the previous regional studies, we are able to assess separately the impact of immigration on exports and imports. Immigrants have a strongly significant impact. They promote exports as well as imports: doubling their number yields a 7% ($2^{0.102} \approx 1.07$) increase in the value of exports and a 4% ($2^{0.054} \approx 1.04$) increase in the value of imports. The pro-trade effect of immigration on imports is almost twice smaller than on exports. This casts doubt on the existence of a preference channel. However, we will see later that such a difference, which is barely significant here, is in any case not very robust.

The impact on exports is also almost twice smaller than what has been previously found for U.S. states exports. We argue that previous estimations could be tainted with an upward omitted variable bias that can be controlled for by using country fixed-effects. The impact of distance and contiguity is reduced when the stock of immigrants is accounted for. Contiguity is only significant at the 10% level. Its impact reduces drastically once immigrants are controlled for. Indeed, immigrants coming from neighboring countries, such as Belgium, Germany or Italy, locate according to a gravity pattern. Consequently, the share of immigrants originating from these neighboring countries is much higher in the regions behind the border than anywhere else in France.

Specification in levels

We push further the evidence by testing the robustness of the results to two kinds of possible biases: specification and selection due to neglecting zero flows in the log-linear specification.

Columns (3)-(4) and (7)-(8) in table 2 report the results of the 2-step negative binomial estimation procedure (equation (4) in levels). The positive and significant impact of immigrants is confirmed. Furthermore, it is of the same order of magnitude than in the log-linear specification: doubling the number of immigrants from a country yields a 6.5% increase in both the values of exports and imports with this trade partner. Hence, the results do not change drastically when moving to a log-linear specification. Furthermore, they are not driven by the zero-flow truncation. In columns (4) and (8), where null flows are included in the sample, results remain barely the same.

Finally, we provide further robustness checks based on different estimation techniques (see table 11 in Appendix C). The orders of magnitude are virtually the same in all procedures but the Poisson QML estimation. This is probably due to the assumption that conditional mean equals conditional variance, which would not be valid in our data. Therefore, the pro-trade effect of immigration is robust to both specification and selection biases. We now turn to the endogeneity problem in the log-linear specification.

3.2 An instrumental variables approach

Despite the inclusion of fixed effects and the use of a fine geographical scale, our results could still be plagued by the endogeneity of immigrants' stocks. We use an instrumental variables approach to circumvent this issue within the log-linear model.[9] We choose the lagged stocks of immigrants for the years 1975, 1982 and 1990 as instruments.

[9]Non-linear models, as the negative binomial model, remain quite hard to instrument, as reviewed by Windmeijer (2006). Instrumenting is all the more challenging in our setting that we include numerous dummies. This is the reason why, in this section, we exclusively focus on the log-linear specification.

Relevance of instruments

In order to be relevant, instruments have to be correlated with the current stock of immigrants. Hence, we should observe some persistence in the geography of immigrants' settlements within France, by country of origin. This is a well-known established empirical fact. For instance, Jayet and Bolle-Ukrayinchuk (2007) find that, in France, past settlements strongly determine the location of new immigrants, due to the existence of social networks or to family motives. Table 3 reports the pairwise correlations between past and current stocks of immigrants. We see that these correlations are indeed fairly high, even though they decrease as time-lag raises. This is a first support for validating instruments.

Nevertheless, strict relevance depends on the partial correlation between the endogenous variable and the instruments, once the other exogenous regressors have been controlled for. Table 4 reports the OLS estimates of the traditional first step of the 2-step instrumented regression. We further report the F-test of the joint significance of excluded instruments, as well as the Bound, Jaeger, and Baker (1995) partial R^2 (BJB R^2 herafter). As shown by Baum, Schaffer, and Stillman (2003), in the case of a single endogenous explanatory variable, these tests are sufficient to assess the relevance of instruments. According to the Staiger and Stock (1997) rule of thumb,[10] our instruments are thus relevant. Nevertheless, in regression (4), the elasticity of the 1968 stock of immigrants is not significant. The weakness of instruments being often worse that the endogeneity bias itself, we choose to remain parsimonious, and leave this instrument out of the list.

Table 3: Pairwise correlations for instruments

	ln(1+Immigrants 1999)	
	Correlation	Nb. obs.
ln(1+Immigrants 1990)	0.92	8011
ln(1+Immigrants 1982)	0.92	5697
ln(1+Immigrants 1975)	0.87	4366
ln(1+Immigrants 1968)	0.79	4162

Note: All correlations are significant at the 1% level.

Supporting the validity of instruments

In what follows, we estimate two instrumented models. In the first, we use the stock of immigrants in 1990 as the only instrument. This variable is actually the most highly correlated with

[10] In the case of a single endogenous explanatory variable, a F-statistic below 10 is of concern. All our F-statistics are far greater than 10.

Table 4: Relevance of the lagged stocks of immigrants as instruments

Dependent variable:	ln(1+ Immigrants 1999)			
	(1)	(2)	(3)	(4)
Immigrants 1990	0.566^a	0.503^a	0.488^a	0.505^a
	(0.007)	(0.01)	(0.012)	(0.013)
Immigrants 1982		0.218^a	0.242^a	0.24^a
		(0.01)	(0.012)	(0.013)
Immigrants 1975			0.045^a	0.061^a
			(0.011)	(0.013)
Immigrants 1968				-0.012
				(0.011)
Distance	-0.055	0.106^b	0.155^a	0.146^a
	(0.047)	(0.041)	(0.038)	(0.036)
Contiguity	0.854^a	0.665^a	0.573^a	0.534^a
	(0.112)	(0.094)	(0.08)	(0.075)
Obs.	8011	5471	4038	3558
Adj. R^2	0.934	0.949	0.961	0.965
$F(N_1, N_2)$	6069.6	3969.4	2886.1	2285.2
N_1	1	2	3	4
N_2	7805	5306	3881	3400
BJB R^2	0.44	0.6	0.69	0.73

Note: Country and *département* fixed effects are not reported here. Standard errors in brackets, with ^{a}, ^{b} and ^{c} denoting significance at the 1%, 5% and 10% levels respectively.

the endogenous regressor, and it is non-missing for most of the observations. Consequently, the model is just-identified and the validity of the instrument, which cannot be tested, must be assumed. In the second model, we run a GMM-type instrumentation by introducing simultaneously the lagged stocks of immigrants in 1975, 1982 and 1990. Even though the number of missing observations drastically increases, the model is now over-identified. Hence, we can test for over-identification restrictions. We follow the suggestion of Baum, Schaffer, and Stillman (2003) in the presence of heteroskedasticity, and run the Hansen-J test. A rejection of the null hypothesis implies that the instruments do not fulfill the orthogonality conditions. Regarding exports, the statistic is equal to $\chi^2(2) = 0.45$ with a p-value at 0.8, whereas for imports, the value is $\chi^2(2) = 1.25$, with a p-value at 0.53. In both cases, we thus fail to reject the null hypothesis. The fail of the rejection of the null is a further proof of the validity of instruments.

Results from instrumented regressions

In the columns (1) and (5) of table 5, we estimate the log-linear specification for all the observations for which the stock of immigrants in 1990 is non-missing. This slightly reduces the sample. The pro-trade effect of immigrants is broadly the same for exports and imports, with an elasticity at 0.112. Doubling the stock of immigrants yields a trade increase of 8%. This is the new benchmark against which we assess the endogeneity bias.

In columns (2) and (6), we report the estimates drawn from the just-identified model. Instrumentation confirms the significant and positive impact of immigration on exports and imports. Even though the elasticities are slightly reduced, which means that benchmark estimates were plagued by a small upward endogeneity bias, the orders of magnitude remain fairly stable, around 0.095. To the best of our knowledge, no such a formal robustness check had been proposed in the literature.

Columns (3) and (7) provide OLS estimates for the log-linear specification, based on the country-pairs for which all past stocks of immigrants are non-missing. This reduces drastically the number of observations. However, instrumented regressions reported in columns (4) and (8) provide estimates that are not significantly different from OLS results. This confirms that, even on this small sub-sample, the positive impact of immigration on trade is not driven by a reverse causality or an omitted variable bias.

Table 5: Instrumented regressions at the *département*-level

	Export				*Imports*			
	Just-identified		*Over-identified*		*Just-identified*		*Over-identified*	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Distance	-0.704^a	-0.711^a	-0.62^a	-0.62^a	-1.533^a	-1.541^a	-1.318^a	-1.312^a
	(0.083)	(0.083)	(0.074)	(0.072)	(0.128)	(0.127)	(0.117)	(0.115)
Contiguity	0.322^b	0.357^b	0.274^c	0.281^b	0.167	0.205	0.18	0.081
	(0.161)	(0.164)	(0.142)	(0.141)	(0.196)	(0.2)	(0.192)	(0.191)
Immigrants	0.115^a	0.094^a	0.162^a	0.159^a	0.12^a	0.099^b	0.186^a	0.239^a
	(0.018)	(0.026)	(0.021)	(0.025)	(0.029)	(0.041)	(0.035)	(0.042)
Obs.	7833	7833	4022	4022	7097	7097	3880	3880
Adj. R^2	0.854	0.854	0.882	0.882	0.809	0.809	0.843	0.843

Note: Country and *département* fixed effects are not reported here. Standard errors in brackets with ^{a}, ^{b} and ^{c} denoting significance at the 1%, 5% and 10% levels respectively.

To sum up, immigrants do have a positive and significant impact on both exports and imports. A doubling of the stock of immigrants increases the value of exports by 7 to 12%, depending on the sample and the estimation procedure. The impact on imports, between 7 and 18%, is slightly more variable but of the same order of magnitude. We further find that these results are robust to specification and selection biases and that endogeneity introduces only a slight upward bias in OLS estimates.

4 Product complexity, quality of institutions and immigration

In this last section, we study the pro-trade effect of immigration along two intertwined dimensions: the degree of complexity (or differentiation) of traded products, and the quality of institutions in partner countries.

The complexity of traded goods

Rauch (1999) is the first to argue that trade impediments would depend on the degree of differentiability of traded products. He distinguishes differentiated goods from those sold on an organized market or possessing a reference price. In a gravity-type model of international trade, he provides convincing evidence that proximity, common language and colonial ties matter more for the former than for the latter. Using the same classification, Rauch and Trindade (2002) even argue that the trade-creating impact of immigration, the Chinese diaspora in their study, is much more salient for differentiated than for homogeneous goods. Hence, transnational networks would bridge the information gap between international sellers and buyers in a more salient way for trade in differentiated goods.

We investigate a similar conjecture for French *départements* and their international trade partners. We first match the NST/R industrial classification with the 4-digit SITC classification of Rauch.[11] We consider two types of goods only: simple and complex goods. Simple goods are either those exchanged on an organized market or those possessing a reference price. Complex goods are all the other ones, classified by Rauch as differentiated goods.[12]

We estimate now:

$$\ln y_{kij} = f_{ki} + f_{kj} - \beta \ln \text{dist}_{ij} + \gamma \text{contig}_{ij} + \alpha_k \ln \left(1 + \text{mig}_{ij}\right) + \epsilon_{kij}, \tag{5}$$

where k indices the type of goods, with $k \in (simple, complex)$. Exports and imports, as well as country and *département* dummies, are now commodity-specific. Whereas we assume that the distance and contiguity effects do not vary across goods,[13] the elasticity of trade with respect to the stock of immigrants is also commodity-specific. Contrary to Rauch and Trindade (2002), we run two separate regressions for exports and imports.

Table 6 reports the OLS estimates for specification (5) in log (columns OLS) and the 2-step

[11] See Appendix B for further details.

[12] Berkowitz, Moenius, and Pistor (2006) follow the same dichotomy. Results are not drastically changed if we consider three categories separately.

[13] Allowing these elasticities to be commodity-specific does not change the estimates of the impact of immigrants. However, it reduces the precision of the distance and contiguity estimates but, as noted above, this remains difficult to interpret.

Table 6: Product type and immigration

	Exports				*Imports*			
	OLS		2NB$\geq$ 0		OLS		2NB$\geq$ 0	
	Simple	*Complex*	*Simple*	*Complex*	*Simple*	*Complex*	*Simple*	*Complex*
Distance	-0.775^a (0.072)		-0.951^a (0.086)		-1.492^a (0.099)		-1.603^a (0.124)	
Contiguity	0.371^a (0.143)		0.19 (0.134)		0.425^a (0.155)		0.082 (0.181)	
Immigrants	0.141^a (0.025)	0.074^a (0.018)	0.123^a (0.025)	0.095^a (0.022)	0.029 (0.035)	0.075^a (0.027)	0.05 (0.044)	0.113^a (0.043)
Obs.	17711		18800		15396		18988	
Adj. R^2	0.809				0.766			

Note: Country and *département* fixed effects are not reported here. Robust standard errors in brackets, with ^{a}, ^{b} and ^{c} denoting significance at the 1%, 5% and 10% levels respectively.

negative binomial QML estimates for specification (5) in levels (column 2NB$\geq$ 0). A first striking feature is that the trade-creating effect of immigration is now different for exports and imports. Recall that, when the type of goods was not taken into account, the pro-trade effect of immigrants was of the same order of magnitude for exports and imports. By way of contrast here, immigration boosts the imports of complex commodities (with an elasticity at 0.113), whereas it has no significant impact on the imports of simple products.[14] This is consistent with the idea that social networks, by providing market information and supplying matching or referral services, would matter more for the imports of complex products. Regarding exports, migrants have a significant impact on both simple and complex goods. The effect would be even slightly stronger for simple goods, even if the difference is not significant.

Such average elasticities could hide another source of heterogeneity, depending on the partner country characteristics, as recently suggested by Bandyopadhyay, Coughlin, and Wall (2008). In the following, we disentangle further the pro-trade impact of immigration according to the rules of law in partner countries, on aggregate flows first and then, by goods type.

The quality of the trade partner's institutions

Some recent papers study the impact of institution quality on the volume of bilateral trade. In a matching model of international trade, Turrini and van Ypersele (2006) provide new evidence on the deterrent impact of legal asymmetries on bilateral trade between OECD countries, as well as between French regions. Besides, Anderson and Marcouiller (2002) establish that good institutions would reduce predation at the border. They find that a 10% rise in a country's index

[14]In the remaining, we comment the results associated with estimations in levels only, differences with estimates in logs being most of time insignificant.

of transparency and impartiality yields a 5% increase in its import volumes, other things equal.[15]

Berkowitz, Moenius, and Pistor (2006) add that the quality of the exporter's institutions matters even more. They argue that, if some common contracts (as letters of credit, counter-trade agreements and pre-payment) exist to offset the exporter's risk of not getting paid, such devices are scarcer to offset the importer's risk of late delivery and product defects. Therefore, formal institutions, such as courts and arbitration tribunals for seeking compensation, are of primary interest for importers. Most of the time, the courts or arbitration tribunals in the export country are indeed the last fallback for resolving disputes, the reason why the quality of institutions is more important in the export country.

Rauch (2001) puts forward the idea that transnational networks could be a substitute for weak institutions or weak mechanisms of arbitration. But, as far as we know, this effect has only been empirically studied by Dunlevy (2006), who restricts the focus to U.S state exports. We further investigate the conjecture of transnational network as a substitute for weak institutions on both the international exports and imports of French *départements*. According to Anderson and Marcouiller (2002), the impact of immigration should be greater for exports, as immigrants mitigate any predation behavior at the border of the importing country. According to Berkowitz, Moenius, and Pistor (2006), this should be the reverse as immigrants substitute for weak arbitration tribunals in the exporting country.

Crossing the effects of migrants and institutions may allow us to identity which one of the two previous views is the most salient. We use the *rule of law* index (hereafter RL) provided by Kaufmann, Kraay, and Mastruzzi (2007) as a measure of the quality of institutions. This index measures "the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, the police and the courts, as well as the likelihood of crime and violence". This variable is thus very close to the reality we want to describe.[16]

We proceed with the following estimation:

$$\ln y_{ij} = f_i + f_j - \beta \ln \text{dist}_{ij} + \gamma \text{contig}_{ij} + \alpha \ln \Big(1 + \text{mig}_{ij}\Big) + \rho RL_j * \ln \Big(1 + \text{mig}_{ij}\Big) + \epsilon_{ij}, \quad (6)$$

where the (log of the) stock of immigrants is crossed with the RL index in country j (RL_j). In line with Rauch (2001), we conjecture that immigrants from partner countries with weak institutions

[15] See also de Groot, Linders, Rietveld, and Subramanian (2004) and Ranjan and Lee (2007).

[16] Kaufmann, Kraay, and Mastruzzi (2007) provide six different measures of the quality of institutions. Due to the strong correlation between these measures, we restrict the focus to the rule-of-law index. However, results are unchanged when another index is chosen. The index is decreasing in the quality of institutions and stands between -2.5 and 2.5. We proceed to a simple normalization so that our sample mean would be zero and standard deviation would be one.

have a larger impact on trade flows, in which case we expect a negative sign for ρ.

One could argue that the quality of institutions is endogenous to trade openness, and thus to the volume of trade. If this assertion is certainly right in general, we can forcefully argue that France remains a marginal trading partner for a large majority of countries in the sample. Hence, bilateral flows with France do not determine the quality of its trade partners' institutions. Moreover, the largest trade partners of France are high-income countries, where the quality of institutions is already high.

Table 7: Immigration and the quality of the partner's institutions

	Exports		*Imports*	
	OLS	2NB ≥ 0	OLS	2NB ≥ 0
Distance	-0.839^a	-1.014^a	-1.510^a	-1.678^a
	(0.086)	(0.108)	(0.127)	(0.16)
Contiguity	0.449^a	0.265	0.451^b	0.18
	(0.172)	(0.176)	(0.206)	(0.235)
Immigrants	0.085^a	0.096^a	0.047^c	0.078^c
	(0.018)	(0.02)	(0.027)	(0.04)
RL*Immigrants	-0.067^a	-0.053^a	-0.042^a	-0.058^a
	(0.009)	(0.013)	(0.014)	(0.02)
Obs.	9033	9400	8110	9494
Adj. R^2	0.845		0.8	

Note: Country and *département* fixed effects are not reported here. Robust standard errors in brackets, with ^{a}, ^{b} and ^{c} denoting significance at the 1%, 5% and 10% levels respectively.

Table 7 reports the estimates of specification (6). Note first that the direct trade-impact of institution quality is captured by the country-specific dummy and thus, it cannot be separately identified. Due to the normalization of the rule-of-law index to a zero mean, the average impact of immigrants is taken into account via the *Immigrants* variable. It is almost the same as in section 3. The interacted term *RL*Immigrants* accounts for an heterogeneity in the immigrant effects that depends on institution quality in partner countries. Our results support the conclusion of Dunlevy (2006). The coefficient is negative for exports: immigrants matter more when the quality of institutions is weak in the home country. We compute that the elasticity of exports to immigration ranges between 0.16, for the last country of the first decile of institution quality (Congo) to an insignificant 0.01 for the first country of the last decile (Netherlands).

In addition to Dunlevy (2006), we also provide the related estimates for imports. The impact of immigration also presents a high heterogeneity. The elasticity ranges from 0.15 for the first decile of institution quality to a zero effect for the last decile. Finally, the above-mentioned mechanisms by which weak institutions could impact on trade flows are not exclusive. However, immigrants mitigate the trade-reducing impact of weak institutions in both directions.

Complex products, quality of institutions and immigration

According to our previous discussion, the pro-trade effect of immigrants depends on both the type of goods and the quality of institutions. Hence, it makes sense to study the triple interaction. In the following, we evaluate the cross effect of institutions and immigrants for simple and complex goods separately. Results are reported in table 8.

Table 8: Product type, quality of institutions and immigration

	Exports				*Imports*			
	OLS		2NB≥ 0		OLS		2NB≥ 0	
	Simple	*Complex*	*Simple*	*Complex*	*Simple*	*Complex*	*Simple*	*Complex*
Distance	-0.856^a (0.072)		-1.008^a (0.089)		-1.527^a (0.098)		-1.654^a (0.126)	
Contiguity	0.601^a (0.151)		0.389^a (0.143)		0.554^a (0.16)		0.299 (0.183)	
Immigration	0.118^a (0.025)	0.058^a (0.018)	0.107^a (0.026)	0.084^a (0.022)	0.023 (0.035)	0.07^a (0.027)	0.038 (0.044)	0.106^b (0.042)
RL*Immigration	-0.111^a (0.013)	-0.065^a (0.01)	-0.075^a (0.015)	-0.05^a (0.012)	-0.08^a (0.019)	-0.023^c (0.013)	-0.116^a (0.02)	-0.024 (0.021)
Obs.	17711		18800		15396		18988	
Adj. R^2	0.806				0.766			

Note: Country and *département* fixed effects are not reported here. Robust standard errors in brackets, with ^{a}, ^{b} and ^{c} denoting significance at the 1%, 5% and 10% levels respectively.

For exports, neither the direct impact of immigration nor its crossed effect with institutions significantly differ between simple and complex goods. In both cases, immigrants enhance trade, even more that the quality of institutions is low, which matches aforementioned intuitions. The direct effect is slightly stronger and more heterogenous across rules-of-law for simple goods.

Significant differences between commodities are observed on imports. Regarding the imports of complex goods, the role of immigrants does not depend on the quality of institutions. Since for complex goods immigrants are a real conduit for information, they matter regardless of institution quality. For simple goods conversely, immigrants do not matter on average, because trading such goods does not require further information enhancement: hence, the direct effect is not significant. This result holds unless the quality of institutions is low. In that case, immigrants, who substitute for institutions, play an important role, as shown by the negative significant effect of the interacted variable.

5 Conclusion

The positive impact of immigration on trade is a well-established result. We add to the literature by assessing the crossed effect of immigration, goods complexity and institution quality. Even though numerous theoretical models underline this possible interaction, evidence remains very scarce.

When we do not disentangle the pro-trade effect of immigrants across goods and institutions, we find that the trade-creating impact of immigrants is slightly smaller than that found in the previous literature. This might be due to our careful estimation strategy, in which we consider variables in levels, country fixed-effects and instrumentation. However, these average effects hide a large heterogeneity across products and across trade partners.

The trade-enhancing impact of immigrants is more salient when they come from a country with weak institutions. Doubling the stock of immigrants from countries with the weakest institutions increases exports and imports by 10 to 12%. Conversely, the impact of immigrants is barely significant for countries with best institutions.

Furthermore, immigrants substitute for weak institutions for the exports of both simple and complex goods. Regarding the imports of complex commodities, i.e. those for which the information conveyed by immigrants is the most valuable, the pro-trade effect of immigrants overrides institution quality in the partner country. Conversely, even though immigrants do not enhance the imports of simple goods on average, they play an important role in interaction with the quality of institutions.

References

ANDERSON, J., AND E. V. WINCOOP (2003): "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, 93(1), 170–192.

——— (2004): "Trade Costs," *Journal of Economic Literature*, 42(3), 691–751.

ANDERSON, J. E., AND D. MARCOUILLER (2002): "Insecurity And The Pattern Of Trade: An Empirical Investigation," *The Review of Economics and Statistics*, 84(2), 342–352.

BAIER, S. L., AND J. H. BERGSTRAND (2007): "Bonus Vetus OLS: A Simple Approach for Addressing the Border Puzzle and other Gravity-Equation Issues," University of Notre Dame manuscript.

BANDYOPADHYAY, S., C. C. COUGHLIN, AND H. J. WALL (2008): "Ethnic Networks and US Exports," *Review of International Economics*, 16(1), 199–213.

BARTEL, A. P. (1989): "Where Do the New U.S. Immigrants Live?," *Journal of Labor Economics*, 7(4), 371–91.

BAUM, C. F., M. E. SCHAFFER, AND S. STILLMAN (2003): "Instrumental variables and GMM: Estimation and testing," *Stata Journal*, 3(1), 1–31.

BEHRENS, K., C. ERTUR, AND W. KOCH (2007): "Dual Gravity: Using Spatial Econometrics to Control for Multilateral Resistance," CORE Discussion Paper 2007/59.

BERKOWITZ, D., J. MOENIUS, AND K. PISTOR (2006): "Trade, Law, and Product Complexity," *The Review of Economics and Statistics*, 88(2), 363–373.

BÉNASSY-QUÉRÉ, A., M. COUPET, AND T. MAYER (2007): "Institutional Determinants of Foreign Direct Investment," *The World Economy*, 30(5), 764–782.

BOUND, J., D. A. JAEGER, AND R. M. BAKER (1995): "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak," *Journal of the American Statistical Association*, 90, 443–450.

CAMERON, C., AND P. TRIVEDI (2005): *Microeconometrics: Methods and Applications*. Cambridge Uinversity Press.

CASELLA, A., AND J. RAUCH (2003): "Overcoming Informational Barriers to International Resource Allocation: Prices and Group Ties," *Economic Journal*, 113(484), 21–42.

COMBES, P.-P., M. LAFOURCADE, AND T. MAYER (2005): "The Trade-Creating Effects of Business and Social Networks: Evidence from France," *Journal of International Economics*, 66(1), 1–29.

DE GROOT, H. L. F., G.-J. LINDERS, P. RIETVELD, AND U. SUBRAMANIAN (2004): "The Institutional Determinants of Bilateral Trade Patterns," *Kyklos*, 57(1), 103–123.

DUNLEVY, J. (2006): "The Influence of Corruption and Language on the Protrade Effect of Immigrants: Evidence from the American States," *The Review of Economics and Statistics*, 88(1), 182–186.

EATON, J., AND A. TAMURA (1994): "Bilateralism and Regionalism in Japanese and U.S. Trade and Direct Foreign Investment Patterns," *Journal of the Japanese and International Economies*, 8, 478–510.

FEENSTRA, R. C. (1996): "U. S. imports, 1972–1994: Data and Concordance," Working Paper 5515, National Bureau of Economic Research.

GIRMA, S., AND Z. YU (2002): "The Link between Immigration and Trade: Evidence from the United Kingdom," *Weltwirtschaftliches Archiv*, 138(1), 115–130.

GOULD, D. (1994): "Immigrant Links to the Home Country: Empirical Implications for U.S. Bilateral Trade Flows," *The Review of Economics and Statistics*, 76(2), 302–316.

GOURIEROUX, C., A. MONFORT, AND A. TROGNON (1984): "Pseudo Maximum Likelihood Methods: Applications to Poisson Models," *Econometrica*, 52(3), 701–20.

HEAD, K., T. MAYER, AND J. RIES (2008a): "The Erosion of Colonial Trade Linkages After Independence," Discussion Paper 6951, CEPR.

——— (2008b): "How Remote is the Offshoring Threat?," *European Economic Review*, forthcoming.

HEAD, K., AND J. RIES (1998): "Immigration and Trade Creation: Econometric Evidence from Canada," *Canadian Journal of Economics*, 31(1), 47–62.

HELPMAN, E., M. MELITZ, AND Y. RUBINSTEIN (2008): "Estimating Trade Flows: Trading Partners and Trading Volumes," *The Quarterly Journal of Economics*, 123(2), 441–487.

HERANDER, M., AND L. SAAVEDRA (2005): "Exports and the Structure of Immigrant-Based Networks: the Role of Geographic Proximity," *The Review of Economics and Statistics*, 87(2), 323–335.

JAYET, H., AND N. BOLLE-UKRAYINCHUK (2007): "The location of immigrants in France," mimeo.

KAUFMANN, D., A. KRAAY, AND M. MASTRUZZI (2007): "Governance Matters VI: Aggregate and Individual Governance Indicators 1996-2006," Discussion paper, The World Bank.

MILLIMET, D., AND T. OSANG (2007): "Does State Borders Matter for U.S. Intranational Trade? The Role of History and Internal Migration," *Canadian Journal of Economics*, 40(1), 93–126.

MUNSHI, K. (2003): "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market," *The Quarterly Journal of Economics*, 118(2), 549–599.

RANJAN, P., AND J. Y. LEE (2007): "Contract Enforcement And International Trade," *Economics and Politics*, 19(2), 191–218.

RAUCH, J. (1999): "Networks versus markets in international trade," *Journal of International Economics*, 48(1), 7–35.

——— (2001): "Business and Social Networks in International Trade," *Journal of Economic Literature*, 39(4), 1177–1203.

RAUCH, R., AND V. TRINDADE (2002): "Ethnic Chinese Networks In International Trade," *The Review of Economics and Statistics*, 84(1), 116–130.

SANTOS SILVA, J., AND S. TENREYRO (2006): "The Log of Gravity," *The Review of Economics and Statistics*, 88(4), 641–658.

STAIGER, D., AND J. H. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65(3), 557–586.

THIERRY, X. (2004): "Evolution de l'immigration en France et éléments de comparaison avec le Royaume-Uni," *Population*, 5, 725–764.

TURRINI, A., AND T. VAN YPERSELE (2006): "Legal Costs as Barriers to Trade," Discussion Paper 5751, CEPR.

WAGNER, D., K. HEAD, AND J. RIES (2002): "Immigration and the Trade of Provinces," *Scottish Journal of Political Economy*, 49(5), 507–25.

WINDMEIJER, F. (2006): "GMM for panel count data models," Bristol Economics Discussion Papers 06/591, Department of Economics, University of Bristol, UK.

A Data on trade and immigration

Trade flows

Trade flows come from the SITRAM data set provided by the French Ministry of Transport. It reports the value of imports and exports of 94 French metropolitan *départements* with around 200 trading partners all around the world. French *départements* are administrative units of much smaller and more regular size than US States or Canadian Provinces. The mean area of French *départements* is 5,733 km^2, with a coefficient of variation at 0.34 (when Corsica and overseas French regions are excluded), whereas the related figures are 162,176 km^2 (with a standard deviation at 0.77) for US states (when Alaska and Washington DC are included), and 606,293 km^2 (with a standard deviation at 0.82) for Canadian provinces (when Nunavut, North-West and Yukon territories are excluded). However, the instrumentation strategy requires that countries remain comparable across time. And the decade 1990-2000 has seen a large deal of modifications in the drawing of countries with, for instance, the disaggregation of the former Soviet Union and of Ex-Yugoslavia. Hence, we recover those entities as they were before the separation:

- Four former single countries have been divided during the 1990's. In order to match the data set in 1999 with our explanatory variables, we thus aggregate Armenia, Azerbaidjan, Belarus, Estonia, Georgy, Kazakhstan, Kirghistan, Lettonia, Lituania, Moldova, Ouzbekistan, Russia, Tadjikistan, Turkmenistan and Ukrainia in a single *former Soviet Union*. Czech Republic and Slovakia are aggregated in *former Czecholovakia*, Bosnia, Croatia, Serbia, Montenegro, Slovenia and Macedonia in *former Socialist Republic of Yugoslavia*, Erythrea and Ethiopia in *former Ethiopia*.

- We also aggregate three countries that have been reunified during the 1990's: Germany (former DDR and former GDR), Yemen (former South and North Yemen), and the Emirates.

We further consider as a single country: 1/Belgium and Luxembourg, 2/Italy, San Marin and Vatican, 3/Denmark and Feroe Islands, 4/Switzerland and Lichtenstein. After this manipulation, 161 countries remain in the data set, with at least one positive flow towards or from a French *département*.

As noted in the main text, the value of trade flows is generally exclusive of transit shipments. Petroleum products are however a noticeable exception. Hence, we leave them out of the sample. We also neglect postal, pipers and other too specific shipments.

The distributions of exports and imports across countries are right-skewed, with a set of few countries accounting for the largest amount of trade flows: nine countries only account for more than 70% of the value of exports and of imports (Germany, Belgium/Luxembourg, Spain, Italy, the Netherlands, United-Kingdom, United-States, Switzerland and Japan). It is also worth noting that half of the sample (80 countries) accounts for 98% (99%) of the value of exports (imports). Furthermore, import and export countries are very similar: the Spearman rank correlation between importers and exporters stands at 0.86.

Immigration

The 1999 French population census, from the French National Statistical Institute (INSEE), provides us with exhaustive information on the number of foreign-born residents by *département*. For each foreign-born resident, we know the country of birth, the nationality at birth, and the nationality at the time of the census. We are then able to distinguish between 1/French citizens born abroad, 2/foreign citizens born in France, 3/foreign citizens born abroad but having acquired the French nationality, and finally 4/foreign citizens born abroad with a foreign nationality at the time of the census.

As the place of birth is more important in the construction of a social network than the current nationality, we consider the narrower concept of *immigrant*. The French Statistical Institute disentangles a *foreigner*, i.e. a person whose current nationality is not French, from an *immigrant*, i.e. a person born abroad with a foreign nationality, regardless of his/her nationality at the time of the census. Hence, if an immigrant acquires the French nationality, he/she cannot be considered a foreigner anymore, but remains an immigrant. Note that for a few countries, it is necessary to sort apart French citizens born abroad from foreign-born French citizens. The Algerian case is very enlightening in this respect. Eighteen French *départements* count more than 10,000 French citizens born in Algeria, who are not immigrants (Algeria was a settlement colony of France until 1962). The settlement pattern of French citizens born in Algeria and Algerian-born citizens is not completely similar, with a correlation at 0.64 only.

The distribution of immigration across countries is also highly right-skewed. Eight countries account for more than 70% of immigrants to France (Algeria, Morocco, Portugal, Italy, Spain, Tunisia, Germany and Turkey). Most of these countries do not stand in the top-9 French trading partners. The geography of trade and immigration is thus quite different. The correlation between immigration and exports (imports) stands at 0.65 (0.56). This correlation is only 0.22 (0.20) when we restrict the sample to countries belonging to the upper-median part of the distribution.

To prevent the results from being driven by noisy observations and the skewness of our three variables of interest, we restrict the sample of exports, imports and immigration stocks to the upper-median distribution countries. This leads us to consider a sample of 100 countries for exports and a sample of 101 countries for imports.

Description of the instruments

The French population censuses of 1968, 1975, 1982 and 1990 provide us with a further reliable information on the number of immigrants by *département* and by country of origin, used as instruments to tackle the endogeneity issue. It is worth noting that, for earlier censuses (1968 and 1975), information is not exhaustive as it is extracted from a representative sample (1/4 of the whole French population). Moreover, for these years, we only know the nationality of the residents (and not the country of birth) for a limited number of countries. Hence, the number of observations reduces drastically when we use these variables as instruments. The 1982 and 1990 censuses provide the nationality of the respondent, as well as his/her country of birth. We are then able to recover an instrument variable closer to the endogenous explanatory variable.

Summary statistics

Table 9 depicts further summary statistics on the distributions of exports, imports, distance and immigration over the *département*-country pairs. In the panel of exports, there are 9033 pairs (among 9400 possibilities) of strictly positive flows, against 8110 (among 9494 possibilities) for imports, with a slightly greater pair-average value (31,980 thousands of euros against 30,443 for exports). The frequency of null flows is then quite limited here, in comparison to Helpman, Melitz, and Rubinstein (2008) for instance (half of the sample).

B Matching the NST/R and Rauch's classifications

The NST/R classification consists in a 3-tier nomenclature: 10 *chapters*, 52 *groups*, and 176 *positions*. We match each of these *positions* with the nomenclature built by Rauch (1999), who classifies the 1089 goods of the 4-digit SITC (rev. 2) system into three broad categories: the goods sold on an organized market, the reference price goods or neither of the two. Rauch (1999) provides

Table 9: Summary statistics

	Mean	Std. Dev.	Min	P25	Median	P75	Max
Strictly positive exports (9033/9400)							
Exports	30443.2	134961.7	0.2	311.4	2122.5	12621.7	3500597.5
Distance	5321.9	3758.0	110.6	1956.8	4608.3	8358.1	19839.1
Immigrants	470.6	2224.0	0.0	7.0	29.0	140.0	56540.0
All exports (9400)							
Exports	29254.6	132431.9	0.0	234.1	1848.3	11694.1	3500597.5
Distance	5338.8	3712.9	110.6	2021.2	4638.2	8325.4	19839.1
Immigrants	452.8	2181.9	0.0	6.0	27.0	131.0	56540.0
Strictly positive imports (8110/9494)							
Imports	31079.7	151225.4	0.1	54.9	890.0	9076.8	4451061.5
Distance	5626.0	3933.6	110.6	1912.3	4983.7	8908.9	19839.1
Immigrants	519.2	2341.7	0.0	7.0	34.0	170.0	56540.0
All imports (9494)							
Imports	26549.0	140197.3	0.0	7.2	392.2	6335.9	4451061.5
Distance	5577.7	3704.2	110.6	2238.5	4954.2	8615.1	19839.1
Immigrants	448.1	2171.6	0.0	5.0	26.0	128.0	56540.0

Note: Exports and imports are in thousands of euros, immigrants in number of foreign-born French residents. Distance is the average number of kilometers between capital cities, weighted by their population size.

a conservative and a liberal classification. In the main text, we use the conservative one, but we check that the results are not sensitive to the alternative classification. We cannot define a one-to-one mapping between the categories of Rauch, and the NSTR classification. Therefore, we measure how each *position* distributes across these three broad categories.

To this aim, we use a correspondence between the 6-digit Harmonized Standard (HS6) and the NST/R classifications on one side, and between the HS6 and the classification of Rauch (1999) on the other side. The distribution of each *position* across the three Rauch's categories is computed as the ratio of the number of HS6 items belonging to each category over the number of HS6 items composing a given *position*.

To compute a correspondence table between the NST/R and HS6 classifications, we first use the correspondence table between the 8-digit Combined Nomenclature (CN8) and the NST/R classifications provided by the European Statistical Institute (EUROSTAT). [17] We then use another correspondence table provided by EUROSTAT for the year 1988 to match each CN8 item with only one item of the HS6 classification.[18]

In order to compute a correspondence between the HS6 and the classification of Rauch (1999), we use a correspondence table between the 4-digit SITC (rev. 2) and the 10-digit Harmonized Standard (HS10) classifications provided by Feenstra (1996).[19]

Table 10 provides the distribution of each NST/R *chapter* across the three broad categories defined by Rauch. As expected, differentiated goods mainly appear in chapter 9 (Machinery, transport equipment, manufactured articles), and homogeneous goods in chapters 0 and 4.

[17] Available at: http://ec.europa.eu/eurostat/ramon/other_documents/index.cfm?TargetUrl=DSP_OTHER_DOC_DTL
[18] Available at: http://ec.europa.eu/eurostat/ramon/relations/index.cfm?TargetUrl=LST_REL
[19] Available at: http://cid.econ.ucdavis.edu/data/usixd/imports/conimp89_01.txt.

Table 10: Distribution of the 9 NST/R chapters across Rauch's categories (in %)

Chapters	Label	n	r	w
0	Agricultural products and live animals	19.69	25.87	54.44
1	Foodstuffs	19.26	67.6	13.13
2	Solid mineral fuels	13.77	86.23	0
4	Ores and metal waste	0	60.54	39.46
5	Metal products	29.91	63.56	6.53
6	Crude and manufactured minerals	66.6	33.4	0
7	Fertilizers	3.82	96.18	0
8	Chemicals	59.42	40	0.58
9	Machinery, transport equipment and manufactured articles	96.5	3.17	0.34

Note: n = Differentiated Goods, r = Reference Price Goods, w = Goods sold on an organized market. Chapter 4 (petroleum products) is left out of the analysis.

C Robustness Checks

The first column of table 11 reports OLS estimates equivalent to those presented in table 2. The second column, $OLS(y + 0.1)$ gives the related estimates for the log-linearized model, where the dependent variable has been replaced by the logarithm of 0.1 plus the flow (in thousands of euros). This methodology has been used by Dunlevy (2006), Bénassy-Quéré, Coupet, and Mayer (2007) among others. The third column ($ET - Tobit$) gives the gravity estimates building on a modified Tobit estimator, as suggested by Eaton and Tamura (1994). This method has been used by Herander and Saavedra (2005).

The three following columns report QML estimates. The first column ($2NB$) depicts the results of a 2-step Negative Binomial procedure similar to that of table 2. The second column ($GPML$) presents another QML estimator, where we assume that the error term follows a Gamma distribution. The third column ($PPML$) depicts the Poisson QML estimates used by Santos Silva and Tenreyro (2006).

Table 11: Results from different specifications

	In *Log*			In *Levels*		
	OLS	OLS$(y + 0.1)$	ET-TOBIT	2NB	GPML	PPML
Exports (> 0)	0.102^a	0.101^a	0.082^a	0.092^a	0.091^a	0.24^a
	0.018	0.018	0.014	0.019	0.019	0.035
Exports (≥ 0)	–	0.135^a	0.077^a	0.109^a	0.113^a	0.241^a
		0.021	0.013	0.021	0.021	0.035
Imports(> 0)	0.054^b	0.055^b	0.068^a	0.094^a	0.095^a	0.208^a
	0.027	0.026	0.024	0.035	0.035	0.035
Imports(≥ 0)	–	0.032	0.057^a	0.089^b	0.120^a	0.208^a
		0.027	0.021	0.041	0.047	0.035

Note: Standard errors in brackets, with ^{a}, ^{b} and ^{c} denoting significance at the 1%, 5% and 10% levels respectively.