The University of Nottingham

# research paper series

Globalisation, Productivity and Technology

Research Paper 2008/30

## *Cultural Links, Firm Heterogeneity and the Intensive and Extensive Margins of International Trade*

**by**

Paulo Bastos and Joana Silva

The Centre acknowledges financial support from The Leverhulme Trust under Programme Grant F114/BF

GEP Leverhulme Centre for Research on Globalisation and Economic Policy

**The Authors**
Paulo Bastos is an Economist at the European Commission and a Policy Associate of GEP.

Joana Silva is a Young Professional at the World Bank and a Policy Associate of GEP.

**Acknowledgements**
We are grateful to Ana Margarida Fernandes for valuable comments. Silva thanks *Instituto Nacional de Estatística* for providing the micro data employed in this paper and *Fundação para a Ciência e a Tecnologia* for financial support. The views expressed in this paper are those of the authors and should not be attributed to the institutions they are affiliated with.

# Cultural Links, Firm Heterogeneity and the Intensive and Extensive Margins of International Trade

**by**

**Paulo Bastos and Joana Silva**

**Abstract**

It is well known that cultural links between countries increase bilateral trade. In this paper we exploit Portuguese firm-level data on exports to 199 destinations to investigate the questions: How? Do cultural links increase the number of exporters, or the shipments per exporter? What is the role of firm heterogeneity? The results reveal that cultural links, measured by common language/colonial ties and emigrant communities, are significantly associated with a lower incidence of within-firm export zeros and with larger shipments per exporter. Furthermore, they show that the former of these relationships tends to be magnified by firm size, suggesting that firm heterogeneity is key in shaping the interplay between cultural links and the extensive margin of international trade.

**JEL classification:** F1; F22

**Keywords:** Colonial ties, migration, firm heterogeneity, intensive and extensive margins

**Outline**

## Non-Technical Summary

It is by now accepted as a stylized fact that cultural links between countries increase bilateral trade. In this paper we use detailed firm-level data from Portugal to investigate the questions: How? Do cultural links increase the number of exporters, or the shipments per exporter? What is the role of firm heterogeneity?

The data we use to address these questions gather the shipments of virtually all exporting firms in 2005 to each of 199 destinations. Amongst these, eight are former colonies of Portugal and have the Portuguese as the official language. We further add information on the number of Portuguese emigrants in each importing country, which we use as a complementary measure of cultural links. For each exporting firm in our data, we define an export zero as a bilateral trade flow which could have occurred but did not. We then investigate whether, conditional on the standard gravity regressors, the presence of cultural links is systematically associated with the within-firm probability of exporting to a given market, as well as with the volume exported.

Our results show that the positive association between cultural links and bilateral exports reflects both a lower incidence of export-market zeros (the extensive margin) and higher exports per firm (the intensive margin). Furthermore, they reveal that the former of these relationships tends to be magnified by the exporting firm's size, a result that is consistent with a new generation of trade models with heterogeneous firms.

The findings of this paper may have policy relevance. They suggest that removing existing barriers to international migration could play an important role not only in fostering bilateral trade, but also in promoting the diversification of markets within exporting firms in emigrant-sending countries. The latter of these effects could be especially relevant for developing countries, which tend to display a sizeable emigration potential and firms with a low degree of export-market diversification.

# 1 Introduction

Globalization has proceeded at a rapid pace in recent decades. As a clear observable counterpart to this phenomenon, goods and production factors have become significantly more mobile across national boundaries. Yet whilst distances between nations appear to be shorter than ever, casual observation suggests that numerous international business relationships continue to rely heavily on deeply rooted links, such as colonial ties, common language, or migrant communities.

The importance of cultural links for international trade has been subject to rigorous scrutiny in recent years, with general agreement that their presence leads indeed to significantly higher bilateral exports. Our aim in this paper is to investigate the questions: How? Do cultural links increase the number of exporters, or the shipments per exporter? What is the role of firm heterogeneity? To do so, we exploit highly detailed data from Portugal gathering the shipments of virtually all exporting firms in 2005 to each of 199 destinations. Amongst these, eight are former colonies of Portugal and have the Portuguese as the official language. We further add information on the number of Portuguese emigrants in each importing country, which we use as a complementary measure of cultural links. For each exporting firm in our data, we define an export zero as a bilateral trade flow which could have occurred but did not. We then investigate whether, conditional on the standard gravity regressors, the presence of cultural links is systematically associated with the within-firm probability of exporting to a given market, as well as with the volume exported.

In addition to its coverage and detail, these data have a number of attractive features for our purposes. On the one hand, the geographic and economic situation of Portuguese-speaking countries provide a particularly good ground for disentangling the effect of colonial ties/common language from that of other determinants of export volumes. Indeed, they are relatively distant from Portugal, belong to different continents, differ markedly in terms of size, and are at distinct stages of economic development. One the other hand, Portugal has a sizeable emigrant community, which is widely, but unevenly spread across the globe.[1] Such a variation will be particularly useful for identification in the empirical analysis.

Our results indicate that cultural links reduce the incidence of within-firm export zeros and increase the shipments per exporter. Furthermore, they reveal that the former of these relationships tends to be magnified by the exporting firm's size, suggesting that firm heterogeneity is key in shaping the interplay between cultural links and the extensive margin of international trade. These findings are consistent with an asymmetric, multi-country version of the heterogeneous firms trade model pioneered by Melitz (2003), as recently presented by Baldwin and Harrigan (2007).[2]

This paper relates to several strands of existing research. Based on the so-called gravity model, there is now a substantial body of empirical work showing that common language and colonial ties increase the volume of bilateral trade (e.g. Rauch, 1999; Melitz, 2008). In a recent paper, Helpman et al. (2008) note that traditional estimates of the gravity equation do not account for the absence

[1] As reported in the Appendix, the number of Portuguese emigrants aggregate to nearly 20 percent of the population living in Portugal.

[2] See also Melitz and Ottaviano (2008), Chaney (2008) and Helpman et al. (2008).

of trade relationships between certain country pairs, and show how the observation of both zero and positive trade flows can be rationalized by a multi-country version of the heterogeneous firm trade model of Melitz (2003). Using aggregate trade data for 158 countries, they then use this model to empirically decompose the impact on trade volumes of the typical gravity regressors into the contributions of the intensive (exports per firm) and extensive margins (number of exporting firms). Their results reveal that the positive effect of common language on bilateral exports is solely driven by the extensive margin, whereas colonial ties increase both the extensive and intensive margins of trade. Importantly, they also find that traditional estimates of the gravity equation are biased, and that the bulk of the bias is not driven by sample selection, but rather by the fact that they do not account for the effect of unobserved firm heterogeneity.

Two related papers make use of firm-level data to construct more direct measures of the intensive and extensive margins. Using data on French firms, Crozet and Koening (2007) decompose industry bilateral exports into the number of exporters and the average exports per firm, and then employ each of these variables as the dependent variable in a one-sided gravity model. They find that the positive effect of common language on bilateral exports is solely accounted for the extensive margin, while colonial ties increase both the number of exporters and the average exports per firm. Anderson (2007) exploits firm-level data from Sweden to investigate the link between familiarity and country-level measures of the intensive and extensive margins. Familiarity in his empirical model is captured by dummy variables for Nordic, Baltic and English-speaking nations, and the results suggest that its positive effect on bilateral trade is primarily due to the extensive margin. A key difference between this line of work and our paper is that, by using aggregate measures of the intensive and extensive margin, these papers do not explicitly account for the role of firm heterogeneity in shaping the interplay between cultural links and exports.[3]

Our paper also adds to existing research in that, besides colonial ties and common language, we examine the role of emigration links in determining the extensive and intensive export margins. As first pointed out by Gould (1994), migrant communities may stimulate bilateral through several channels, such as the knowledge of home country markets, common language, preferences, or business contacts. Using US data and an augmented gravity specification, he provides evidence that immigration links increase trade flows with emigrant-sending countries in both directions, a result that has subsequently been shown to hold for many other countries – e.g. Canada (Head and Ries, 1998) and the UK (Girma and Yu, 2002). In a related study, Rauch and Trindade (2002) find that ethnic Chinese networks, proxied by the product of ethnic Chinese population shares, increase bilateral trade more for differentiated than for homogeneous products, which they interpret as evidence that business and social networks help to match buyers and sellers. None of these papers, however, makes use of firm-level data, nor decomposes the effect of migration ties on the intensive and extensive margins.

Finally, we would like to draw attention to recent empirical work by Baldwin and Harrigan

[3]Importantly, Crozet and Koening (2007) also use firm-level data to structurally estimate the model of Chaney (2008), but restrict their focus to the effect of geographic distance on the extensive and intensive margins of trade.

(2007), who examine the ability of different trade models to predict the incidence of bilateral export zeros. Using product-level data for the US, they find that the incidence of within-product export zeros rises with distance and falls with the size of the destination market, as predicted by an asymmetric, multi-country version of the heterogeneous firm trade model of Melitz (2003). There are, however, a number of important differences between their work and our own. First, they do not consider the role of cultural links in determining export zeros, which is the main focus of this paper. Second, they focus on within-product export zeros, whereas we examine the incidence of within-firm export zeros. Finally, we investigate the role of firm heterogeneity in shaping the interplay between the customary gravity model regressors and the incidence of within-firm export zeros, which requires firm-level data.

The remainder of the paper is organized as follows. In Section 2, we describe the data employed. Section 3 presents preliminary evidence based on country-level data. In Section 4, we analyze the extensive and intensive margins at the level of the individual firm. Finally, Section 5 offers some concluding remarks.

## 2 Data

Our main data source is the *Foreign Trade Statistics* (FTS) of Portugal for 2005. This is the country's official information source on international trade statistics, gathering the shipments of virtually all exporting firms to each destination market. The FTS data are collected in two different ways. Data on trade with countries outside the EU (external trade) are collected via the customs clearance system, which covers the universe of external trade transactions. Data on the transactions with other EU member States (internal trade) are obtained via the Intrastat system, where the information providers are companies engaged in internal trade and registered in the VAT system whose value of annual shipments exceeds a legally binding threshold (85,000 Euros in 2005).

Export values in these data are "free on board", thus excluding any duties or shipping charges. The 2005 FTS dataset comprises information on 16,541 exporting firms and 220 destination markets. Despite the above mentioned constraint, the export transactions included in these data aggregate to 97 percent of the total value of merchandise exports reported in the official national accounts.

Table 1 reports some descriptive statistics on the distribution of exporters across markets. As can be seen, on average each firm exported to 3.4 countries. However, the mean hides significant firm heterogeneity. More than one-half of all exporters sell to only one destination (54.2 percent), but they tend to be relatively small exporters, accounting for only 6.8 percent of the total export value. By contrast, only 7 percent of firms export to more than 10 countries, but they account for 60.2% of the total export value.

We have complemented the FTS data with information on each importing country, namely its real GDP (measured at PPP), GDP per worker, and the distance between its most populated city and Lisbon (measured in Kms). The source of the information on distance is CEPII. The remaining variables come from the *World Development Indicators* (WDI) of the World Bank. Whenever

WDI data were reported missing, we have used instead information from the *CIA factbook*. The unavailability of information for some small importing countries forced us to restrict the data set used in the analysis to 199 destination markets.[4] Among these, eight are former colonies of Portugal and had the Portuguese as their official language in 2005: Angola, Brazil, Cape Verde, East Timor, Guinea Bissau, Macau-China, Mozambique, Sao Tome and Principe.[5] Lastly, we have added information on the number of Portuguese emigrants in each importing country. These data come from the *Global Migrant Origin Database* of the World Bank, refer to the year 2000, and are available for 193 of the 199 countries mentioned above.[6]

## 3 Country-level evidence

We begin by presenting preliminary evidence based on aggregate measures of the extensive and intensive margins of bilateral exports. We decompose the shipments of Portugal to each importing country into two different terms:

$$X_c = \sum_j F_{jc} \frac{X_c}{\sum_j F_{jc}} \tag{1}$$

where $\sum_j F_{jc}$ is the number of firms exporting to country $c$ and $X_c/\sum_j F_{jc}$ is the average exports per firm. Figure 1 plots each of these components against the share of each trading partner in total exports ($X_c/X$). A clear pattern emerges from the data: The number of firms exporting to the former Portuguese colonies is relatively high when compared with the share of those countries in Portugal's total exports. These simple descriptive statistics point, therefore, to an important role of the extensive margin in shaping the association between colonial ties/common language and exports.

To investigate more systematically the relationship between cultural links and exports, we estimate a one-sided gravity equation of the following form:

$$\ln X_c = \beta_1 \ln Y_c + \beta_2 \ln(Y/L)_c + \beta_3 \ln DIST_c + \beta_4 LANG_c + \beta_5 \ln EMIGR_c + \mu_c \tag{2}$$

where: $X_c$ is the value of exports to country $c$, $Y_c$ is country $c$'s real GDP, $L_c$ its labour force, $DIST_c$ is the distance between Lisbon and country $c$'s most populated city, and $\mu_c$ an exogenous disturbance.[7] Our main interest lies in the coefficient associated with $LANG_c$ and $EMIGR_c$. The former is a dummy variable which takes the value of one for each of the eight countries that have the Portuguese as the official language, while the latter is the number of Portuguese emigrants in

[4] As a result, 1 percent of the total export value, 175 firms and 38 product categories were excluded from the data set used in the econometric analysis.

[5] In July 13 2007, the Portuguese became one of the official languages of Equatorial Guinea.

[6] Emigration data are reported in the Appendix. Data are not available for Democratic Republic of the Congo, Montenegro, Serbia, Serbia and Montenegro, East Timor, and US Minor Outlying Islands. In the trade data, Serbia, Montenegro, and Serbia and Montenegro are considered as three different destinations. We treat them as so in the analysis, but none of our results is influenced by this decision.

[7] Bernard et al. (2008) adopt a similar specification to analyse US bilateral exports, but do not focus on the role of cultural links.

country $c$.[8] We then proceed by regressing each component of $X_c$ identified in (1) against the same set of explanatory variables. Since OLS is a linear operator, these regressions additively decompose the margins whereby each regressor impacts on bilateral exports (Hummels and Klenow, 2005).

Table 2 reports the econometric results. Not surprisingly, the estimates in column (1) indicate that exports are significantly higher in the presence of colonial ties/common language. The remaining variables are significant and present the expected sign. Our main interest, however, lies in the relative contributions of the extensive and intensive margins − columns (2) and (3). The econometric results suggest that the number of exporters accounts for most of the positive effect of this variable: in column (2), the coefficient attached to $LANG_k$ is significant at the 1% level and its magnitude is just slightly below the estimate presented in column (1); in column (3) the coefficient is much smaller and only weakly significant.

We now turn to the relationship between emigrant communities and exports. The results reported in column (4) suggest that, conditional on the core gravity regressors, the number of Portuguese emigrants in the destination market has a positive and statistically significant impact on exports. Specifically, they indicate that if the number of emigrants doubles, bilateral exports rise by 11.3%. Furthermore, they suggest that the extensive margin accounts for the bulk of this positive relationship. The coefficient associated with $EMIGR_c$ in column (5) is statistically significant and its magnitude is close to that in column (4); the corresponding estimate reported in column (6) is insignificant.

The same pattern of results holds when both variables are included simultaneously − columns (7) to (9). Since there are sizeable emigrant communities in the former Portuguese colonies, the simultaneous inclusion of these variables allows to examine the extent to which the country-specific dummies might be capturing the effect exerted by their presence. Albeit the magnitude of the coefficients declines, both variables remain statistically significant, as would be expected if they were to have an independent impact on bilateral exports.

## 4 Firm-level evidence

While their simplicity is attractive, the aggregate measures of the extensive and intensive export margins used in the preceding section suffer from important limitations. In fact, a key insight of the asymmetric multi-country heterogeneous firms trade model (HFT) is that the effects of trade costs and market size on the within-firm probability of exporting are systematically associated with firm attributes (Baldwin and Harrigan, 2007)[9]. Importantly, recent results by Helpman et. al (2008), based on country-level data, suggest that unobserved firm heterogeneity rather than sample selection is the main source of bias in traditional estimates. Clearly, we are unable to fully account for such heterogeneity if we measure the extensive margin as the number of exporters. Another concern about this measure stems from the existence of a minimum statistical threshold for the

[8] For 7 countries in our data, the number of Portuguese emigrants is zero. Since the log of zero is undefined, this issue was dealt with by transforming all zero values to 0.00001, then taking the log.

[9] See also Chaney (2008) and Helpman et al. (2008).

collection of data on internal trade, which implies that the number of exporters is measured with error.

On the other hand, the use of average shipments per exporter for measuring the intensive margin is also problematic. Indeed, if firms are heterogeneous, the exports per firm are also likely to vary with firm attributes, which we are unable to account for in the estimation if we use average measures of the intensive margin. To address these concerns, in this section we exploit the firm-country detail of our data to examine the relationship between cultural links and the extensive and intensive margins at the micro-level.

### 4.1 Cultural links, firm heterogeneity and the incidence of export zeros

We begin by focusing on the determinants of within-firm export zeros. For each exporter, we define an export zero as a bilateral trade flow which could have occurred but did not. More specifically, a zero occurs if there is at least one market that an exporting firm does not serve. We then proceed by estimating a specification of the form:

$$\Pr(z_{jc}=0) = F(\beta_1 \ln Y_c + \beta_2 \ln(Y/L)_c + \beta_3 \ln DIST_c + \beta_4 LANG_c + \beta_5 \ln EMIGR_c + \eta_j) \quad (3)$$

where: $z_{jc}$= 0 if firm $j$ exports to country $c$, $F$ is a probability distribution function, $\eta_j$ a firm-specific unobserved effect. The remaining variables have the meaning defined above.

Our empirical strategy for examining the drivers of export zeros is similar in spirit to that proposed by Baldwin and Harrigan (2007). Three main novelties are introduced. First, we focus on firm-level rather than product-level export zeros, which is more in line with the asymmetric HFT model. Second, we extend the gravity model to consider the role of colonial ties/common language and emigration links. Third, we explicitly explore the role of firm heterogeneity, by analyzing how the estimated coefficients vary with the size of the exporting firm.

We estimate (3) through a linear probability model with firm fixed-effects (LPM-FE). Hence, we analyze the extent to which cultural links are systematically associated to the within-firm probability of exporting to a given market, conditional on exporting to at least one market. As noted by Baldwin and Harrigan (2007), the LPM-FE estimator is more appropriate than the fixed-effects probit and logit estimators, as the latter are inconsistent when the number of effects is large, which is clearly the case here. On the other hand, the LPM-FE is also preferable to random-effects logit models, as the latter embody the (unsuitable) assumption that firm-effects are orthogonal to country characteristics, and statistical inference relies on standard errors that do not account for clustering by importing country.

The data we use in the econometric analysis cover the shipments of 16,446 exporting firms across 199 destinations. There are, therefore, more than 3 million potential firm-level bilateral trade flows. As we have seen in section 2, however, a large proportion of firms serve only a small subset of these markets. For this reason, 98.41% of total potential trade relationships are zero.[10]

[10] Since information on the number of Portuguese emigrants is only available for 193 destinations, our sample is

Table 3 reports the econometric results. The estimated coefficients give the marginal effects on the probability of a firm exporting to a given market, conditional on exporting to at least one country. Clearly, they suggest that cultural links − measured by common language/colonial ties and emigrant communities − have a positive and significant effect on the probability of exporting to a given destination − columns (1) to (3). When controlling for the effect of emigration, the probability of exporting is 8.2 percentage points higher if the destination market has the Portuguese has the official language. On the other hand, doubling the number of Portuguese emigrants in the destination market, lowers the probability of an export zero by about 0.2 of a percentage point.[11]

Is the positive effect of common language/colonial ties common to all eight former Portuguese colonies or driven by a subset of them? To investigate this question, we re-estimate (2) using country-specific dummy variables for each of the eight countries. The corresponding estimates are presented in columns (4) and (5). While the magnitude of the coefficients varies across countries, the estimate is positive and significant for all eight Portuguese former colonies.

The remaining variables are all significant and present the expected sign. Specifically, the estimates suggest that the within-firm incidence of export zeros is smaller for larger, richer and geographically closer nations. The results shown here are therefore consistent with the predictions of the asymmetric HFT model. They are also in line with the empirical findings of Baldwin and Harrigan (2007), based on product-level data for the US. Moreover, the negative effect of distance on the incidence of within-firm export zeros is compatible with the results of Crozet and Koenig (2007), based on French firm-level data.

**The role of firm heterogeneity** We now take the analysis one step further by investigating whether the relationships examined above are influenced by firm heterogeneity, as predicted by the asymmetric HFT model. To do so, we make use of information on two different measures of firm size: the number of employees; the volume of sales. Based on each of these measures, firms are included into one of three size categories: *Small* if the number of employees is lower than 50 (volume of sales is lower than 3 million euros); *Medium* if they employ between 50 and 250 workers (volume of sales higher than 3 but smaller than 40 million euros); *Large* if the number of workers is greater than 250 (volume of sales greater than 40 million euros).

Table 4 reports the econometric results on the interaction between the firm-size dummies and each of the regressors considered previously (columns 1 and 2). For each gravity regressor, we also conduct tests on the equality of the coefficients across firm-size categories. As can be seen, the econometric results reveal that the effects of the gravity variables on the within-firm probability of exporting tends to be significantly magnified by the size of the exporting firm. This suggests that firm heterogeneity plays a key role in shaping the extensive margin of bilateral trade, as predicted by the HFT model. Importantly, however, the estimates also suggest that one should be cautious interpreting the positive effect of cultural links on the extensive margin as evidence that they serve

reduced accordingly when this variable is considered in the econometric analysis.

[11] Notice that the magnitude of this estimate should be gauged against the fact that the distribution of emigrants across importing countries is very wide − see Appendix.

to reduce trade costs. Indeed, in the light of the estimates for market size, the hypothesis that cultural links lead to greater demand for *nationally-differentiated* products should not be excluded.

### 4.2 Cultural links, firm heterogeneity and the intensive margin

We now turn to the relationship between cultural links and the intensive margin at the level of the individual firm. In other words, we analyze whether cultural links also impact on the volume exported by each firm, conditional on exporting to that market. To do so, we adopt a fixed-effects estimator and focus solely on non-zero export flows. Therefore, identification comes from the within-firm variation of export volumes across the markets to which the firm exports. The estimated equation can be written as:

$$X_{jc} = \alpha \ln Y_c + \beta \ln(Y/L)_c + \ln \gamma DIST_c + \delta LANG_c + \varepsilon \ln EMIGR_c + \eta_j + \mu_{jc} \qquad (4)$$

The econometric results are reported in Table 5. We find a positive and statistically significant relationship between within-firm export volumes and the measures of cultural links. The estimated coefficients indicate that within-firm export volumes are 1.2% higher, on average, for Portuguese speaking countries − column (3). Additionally, they suggest that doubling the number of emigrants increase export volumes by 6.4%. Moreover, the estimates in columns (4) and (5) show that the positive association between colonial ties and export volumes applies to all countries, with exception for Brazil. Overall, the results reveal that, when explicitly accounting for the role of unobserved firm heterogeneity, cultural links are also positively associated with the intensive margin of international trade. If interpreted in the context of the HFT model, this result indicates that cultural links reduce variable trade costs and/or increase demand for *nationally-differentiated* products.

**The role of firm heterogeneity** Using the same strategy of the preceding section, we now examine the extent to which the effect of cultural links on within-firm export volumes is influenced by firm heterogeneity. The corresponding results are reported in Table 6. Unlike for the extensive margin, we find scant evidence of a systematic effect of firm heterogeneity on the interplay between the gravity regressors and the intensive margin. The only exception concerns the interactions of firm-size with geographic distance, for which we find robust evidence that the negative effect of this variable on the intensive margin tends to be magnified by firm size.

## 5 Concluding remarks

In this paper we have exploited detailed firm-level data from Portugal to analyze the relationship between cultural links and exports. Our results suggest that this association reflects both a lower incidence of export-market zeros (the extensive margin) and higher exports per firm (the intensive margin). Furthermore, they indicate that the former of these relationships is magnified by the exporting firm's size, a result that is consistent with a new generation of trade models with heterogeneous firms.

The findings of this paper may have policy relevance. They suggest that removing existing barriers to international migration could play an important role not only in fostering bilateral trade, but also in promoting the diversification of markets *within* exporting firms in emigrant-sending countries. The latter effect could be especially relevant for developing countries, which tend to display a sizeable emigration potential and firms with a low degree of export market diversification.

Future research might usefully extend this line of work by determining the extent to which the positive effect of cultural links on the intensive and extensive margins of international trade is due to reduced trade costs or increased demand for *nationally-differentiated* products.

## References


Andersson, M. (2007). "Entry Costs and adjustments on the extensive margin - an analysis of how familiarity breeds exports.", *Royal Institute of Technology of Stockholm, CESIS Working Paper* 81.

Baldwin, R. and Harrigan, J. (2007). "Zeros, quality and space: Trade theory and trade evidence.", NBER Working Paper 13214.

Bernard, A., Jensen, J.B., Redding, S. and Schott, P. (2008). "Firms in international trade.", *Journal of Economic Perspectives*, forthcoming.

Chaney, T. (2008). "Distorted Gravity: The intensive and extensive margins of international trade.", *American Economic Review*, forthcoming.

Crozet, M. and Koenig, P. (2007). "Structural gravity equation with extensive and intensive margins.", Université of Paris X-Nanterre, *Working Paper* 2007-36.

Girma, S. and Yu, Z. (2002). "The link between immigration and trade: Evidence for the United Kingdom.", *Weltwirtschaftliches Archiv/Review of World Economics*, vol. 138, pp. 115-130.

Gould, D. (1994). "Immigrant links to the home country: Empirical implications for U.S. bilateral trade flows." *Review of Economics and Statistics*, vol, 76, pp. 302-316.

Head and Ries (1998). "Immigration and Trade Creation: Econometric Evidence from Canada." *Canadian Journal of Economics,* vol 31, pp. 47-62.

Helpman, E., Melitz, M. and Rubinstein, Y. (2008). "Estimating trade flows: Trading partners and trading volumes." *Quarterly Journal of Economics,* vol. 123, pp. 441-487.

Hummels, D. and Klenow, P. (2005). "The variety and quality of a nation's exports.", *American Economic Review*, 95, 704-723.

Melitz, M. (2003). "The impact of trade on intra-industry reallocation and aggregate industry productivity.", *Econometrica*, vol. 71, pp. 1695-1725.

Melitz, M. and Ottaviano, G. (2008). "Market Size, Trade and Productivity.", *Review of Economic Studies*, vol. 75, pp. 295-316.

Melitz, J. (2008). "Language and foreign trade.", *European Economic Review*, vol. 52, pp. 667-699.

Rauch, J. (1999). "Networks versus markets in international trade.", *Journal of International Economics*, 48, pp. 7-35.

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

**Table 1: Markets per firm**

| Number of markets covered | % of firms | % of value |
|---|---|---|
| 1 | 54.2 | 6.8 |
| 2 | 15.0 | 6.2 |
| 3 | 7.7 | 4.7 |
| 4 | 5.0 | 3.2 |
| 5 | 3.3 | 3.7 |
| 6 | 2.4 | 3.9 |
| 7 | 1.8 | 2.5 |
| 8 | 1.5 | 3.6 |
| 9 | 1.1 | 2.6 |
| 10 | 1.0 | 2.6 |
| More than 10 | 7.0 | 60.2 |
| Average number of markets per firm | 3.4 | |
| Maximum number of markets per firm | 84 | |

**Table 2: The extensive and intensive export margins**

| | $X_c$ | $\sum_j F_{jc}$ | $\frac{X_c}{\sum_j F_{jc}}$ | $X_c$ | $\sum_j F_{jc}$ | $\frac{X_c}{\sum_j F_{jc}}$ |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $Y_c$ | 0.776 | 0.517 | 0.260 | 0.705 | 0.448 | 0.257 |
| | (12.66)*** | (13.95)*** | (6.32)*** | (9.83)*** | (9.52)*** | (5.54)*** |
| $(Y/L)_c$ | 0.411 | 0.309 | 0.101 | 0.383 | 0.288 | 0.095 |
| | (2.68)*** | (3.42)*** | (1.26) | (2.65)*** | (3.36)*** | (1.20) |
| $DIST_c$ | -1.602 | -0.911 | -0.691 | -1.638 | -0.900 | -0.738 |
| | (8.28)*** | (7.79)*** | (5.58)*** | (8.91)*** | (7.66)*** | (6.03)*** |
| $LANG_c$ | 4.461 | 4.067 | 0.394 | | | |
| | (7.60)*** | (8.29)*** | (1.79)* | | | |
| $EMIGR_c$ | | | | 0.113 | 0.098 | 0.015 |
| | | | | (2.87)*** | (3.26)*** | (0.71) |
| Observations | 199 | 199 | 199 | 193 | 193 | 193 |
| R-squared | 0.70 | 0.75 | 0.38 | 0.67 | 0.66 | 0.39 |

Robust t-statistics in parentheses. *significant at 10%; **significant at 5%;***significant at 1%.

All continuous variables are in logs.

**Table 2: The extensive and intensive export margins (continued)**

| | $X_c$ | $\sum_j F_{jc}$ | $\frac{X_c}{\sum_j F_{jc}}$ |
|---|---|---|---|
| | (7) | (8) | (9) |
| $Y_c$ | 0.748 | 0.489 | 0.259 |
| | (10.95)*** | (12.86)*** | (5.42)*** |
| $(Y/L)_c$ | 0.419 | 0.322 | 0.097 |
| | (2.77)*** | (3.64)*** | (1.21) |
| $DIST_c$ | -1.678 | -0.938 | -0.740 |
| | (9.14)*** | (8.48)*** | (6.01)*** |
| $LANG_c$ | 4.047 | 3.865 | 0.181 |
| | (6.12)*** | (6.91)*** | (0.85) |
| $EMIGR_c$ | 0.067 | 0.054 | 0.013 |
| | (1.98)** | (2.39)** | (0.58) |
| Observations | 193 | 193 | 193 |
| R-squared | 0.72 | 0.77 | 0.39 |

Robust t-statistics in parentheses. *significant at 10%; **significant at 5%; ***significant at 1%. All continuous variables are in logs.

**Table 3: The drivers of within-firm export zeros**

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| $Y_c$ | 0.005 | 0.004 | 0.004 | 0.005 | 0.004 |
| | (4.81)*** | (3.51)*** | (4.41)*** | (4.91)*** | (4.67)*** |
| $(Y/L)_c$ | 0.005 | 0.004 | 0.005 | 0.005 | 0.005 |
| | (3.11)*** | (2.39)** | (3.04)*** | (3.34)*** | (3.42)*** |
| $DIST_c$ | -0.019 | -0.018 | -0.019 | -0.018 | -0.018 |
| | (3.02)*** | (2.90)*** | (3.02)*** | (2.79)*** | (2.86)*** |
| $LANG_c$ | 0.084 | | 0.082 | | |
| | (2.80)*** | | (2.38)** | | |
| $EMIGR_c$ | | 0.003 | 0.002 | | 0.002 |
| | | (3.47)*** | (2.73)*** | | (2.93)*** |
| Angola | | | | 0.288 | 0.282 |
| | | | | (159.06)*** | (107.31)*** |
| Brazil | | | | 0.028 | 0.017 |
| | | | | (6.24)*** | (2.54)** |
| Cape Verde | | | | 0.132 | 0.124 |
| | | | | (28.77)*** | (21.01)*** |
| Guinea Bissau | | | | 0.041 | 0.034 |
| | | | | (8.50)*** | (6.58)*** |
| Macau | | | | 0.012 | 0.009 |
| | | | | (2.77)*** | (2.05)** |
| Mozambique | | | | 0.066 | 0.053 |
| | | | | (19.44)*** | (12.87)*** |
| Sao Tome and Principe | | | | 0.061 | 0.053 |
| | | | | (13.44)*** | (12.27)*** |
| East Timor | | | | 0.043 | |
| | | | | (4.35)*** | |
| Firm-fixed effects | Yes | Yes | Yes | Yes | Yes |
| Observations | 3,272,754 | 3,174,078 | 3,174,078 | 3,272,754 | 3,174,078 |
| Firms | 16,446 | 16,446 | 16,446 | 16,446 | 16,446 |
| Destinations | 199 | 193 | 193 | 199 | 193 |
| R-squared | 0.04 | 0.04 | 0.05 | 0.06 | 0.07 |

Robust t-statistics in parentheses, based on standard errors clustered by importing country.

*significant at 10%; **significant at 5%;***significant at 1%. All continuous variables are in logs.

**Table 4: Export zeros and firm size**

| | Number of employees | | | Volume of sales | | |
|---|---|---|---|---|---|---|
| | (1) | Test on equal. of coeff. (Ref. group: Medium) | | (2) | Test on equal. of coeff. (Ref. group: Medium) | |
| | Coeff. | F-stat | P-value | Coeff. | F-stat | P-value |
| $Y_c*Small$ | 0.003 | 29.3 | 0.00 | 0.003 | 35.2 | 0.00 |
| | (3.58)*** | | | (3.55)*** | | |
| $Y_c*Medium$ | 0.010 | | | 0.008 | | |
| | (5.09)*** | | | (5.14)*** | | |
| $Y_c*Large$ | 0.017 | 67.6 | 0.00 | 0.013 | 64.0 | 0.00 |
| | (7.14)*** | | | (6.69)*** | | |
| $(Y/L)_c*Small$ | 0.003 | 12.9 | 0.00 | 0.003 | 14.53 | 0.00 |
| | (2.42)** | | | (2.44)** | | |
| $(Y/L)_c*Medium$ | 0.012 | | | 0.009 | | |
| | (3.46)*** | | | (3.45)*** | | |
| $(Y/L)_c*Large$ | 0.018 | 18.9 | 0.00 | 0.014 | 19.4 | 0.00 |
| | (4.12)*** | | | (4.18)*** | | |
| $DIST_c*Small$ | -0.013 | 15.1 | 0.00 | -0.013 | 13.8 | 0.00 |
| | (2.64)*** | | | (2.72)*** | | |
| $DIST_c*Medium$ | -0.040 | | | -0.032 | | |
| | (3.46)*** | | | (3.28)*** | | |
| $DIST_c*Large$ | -0.047 | 2.96 | 0.09 | -0.041 | 15.5 | 0.00 |
| | (4.67)*** | | | (4.09)*** | | |
| $LANG_c*Small$ | 0.080 | 0.02 | 0.88 | 0.074 | 14.3 | 0.00 |
| | (2.27)** | | | (2.18)** | | |
| $LANG_c*Medium$ | 0.079 | | | 0.094 | | |
| | (2.69)*** | | | (2.67)*** | | |
| $LANG_c*Large$ | 0.142 | 26.11 | 0.00 | 0.179 | 89.4 | 0.00 |
| | (3.62)*** | | | (4.19)*** | | |
| $EMIGR_c*Small$ | 0.001 | 9.1 | 0.00 | 0.001 | 9.43 | 0.00 |
| | (2.54)** | | | (2.54)** | | |
| $EMIGR_c*Medium$ | 0.004 | | | 0.003 | | |
| | (2.94)*** | | | (2.91)*** | | |
| $EMIGR_c*Large$ | 0.004 | 0.00 | 0.95 | 0.004 | 1.78 | 0.18 |
| | (2.80)*** | | | (2.87)*** | | |
| Firm fixed-effects | Yes | | | Yes | | |
| Observations | 3,174,078 | | | | | |
| Firms | 16,446 | | | | | |
| Destinations | 193 | | | | | |
| R-squared | 0.07 | | | 0.06 | | |

Robust t-statistics in parentheses, based on standard errors clustered by importing country.

*significant at 10%; **significant at 5%;*** significant at 1%. All continuous variables are in logs.

**Table 5: The drivers of within-firm export volumes**

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| $Y_c$ | 0.312 | 0.177 | 0.248 | 0.338 | 0.264 |
| | (7.01)*** | (4.17)*** | (4.61)*** | (7.92)*** | (4.94)*** |
| $(Y/L)_c$ | 0.079 | -0.036 | 0.058 | 0.102 | 0.076 |
| | (1.03) | (0.46) | (0.91) | (1.07) | (0.98) |
| $DIST_c$ | -0.933 | -0.768 | -0.861 | -0.925 | -0.847 |
| | (10.53)*** | (8.11)*** | (10.63)*** | (10.28)*** | (10.23)*** |
| $LANG_c$ | 1.425 | | 1.162 | | |
| | (3.59)*** | | (2.68)*** | | |
| $EMIGR_c$ | | 0.097 | 0.064 | | 0.069 |
| | | (3.29)*** | (2.07)** | | (2.23)** |
| Angola | | | | 2.130 | 1.925 |
| | | | | (10.47)*** | (9.49)*** |
| Brazil | | | | 0.163 | -0.095 |
| | | | | (1.06) | (0.57) |
| Cape Verde | | | | 1.659 | 1.346 |
| | | | | (6.84)*** | (4.50)*** |
| Guinea Bissau | | | | 1.338 | 0.968 |
| | | | | (4.07)*** | (2.81)*** |
| Macau | | | | 1.190 | 1.025 |
| | | | | (5.37)*** | (4.56)*** |
| Mozambique | | | | 1.668 | 1.144 |
| | | | | (6.19)*** | (3.51)*** |
| Sao Tome and Principe | | | | 2.154 | 1.649 |
| | | | | (6.46)*** | (4.20)*** |
| East Timor | | | | 2.891 | |
| | | | | (7.17)*** | |
| Firm-fixed effects | Yes | Yes | Yes | Yes | Yes |
| Observations | 56,024 | 55,784 | 55,784 | 56,024 | 55,784 |
| Firms | 16,446 | 16,414 | 16,414 | 16,446 | 16,414 |
| Destinations | 199 | 193 | 193 | 199 | 193 |
| R-squared | 0.15 | 0.14 | 0.15 | 0.16 | 0.16 |

Robust t-statistics in parentheses, based on standard errors clustered by importing country.
*significant at 10%; **significant at 5%;*** significant at 1%. All continuous variables are in logs.

**Table 6: Export volumes and firm size**

| | Number of employees | | | Volume of sales | | |
|---|---|---|---|---|---|---|
| | (1) | Test on equal. of coeff. (Ref. group: Medium) | | (2) | Test on equal. of coeff. (Ref. group: Medium) | |
| | Coeff. | F-stat | P-value | Coeff. | F-stat | P-value |
| $Y_c*Small$ | 0.223 | 0.48 | 0.49 | 0.222 | 1.02 | 0.31 |
| | (4.83)*** | | | (5.47)*** | | |
| $Y_c*Medium$ | 0.249 | | | 0.254 | | |
| | (3.79)*** | | | (4.03)*** | | |
| $Y_c*Large$ | 0.311 | 2.17 | 0.14 | 0.284 | 0.34 | 0.56 |
| | (4.27)*** | | | (3.66)*** | | |
| $(Y/L)_c*Small$ | 0.040 | 0.07 | 0.79 | 0.054 | 0.03 | 0.86 |
| | (0.79) | | | (1.17) | | |
| $(Y/L)_c*Medium$ | 0.049 | | | 0.047 | | |
| | (0.71) | | | (0.63) | | |
| $(Y/L)_c*Large$ | 0.084 | 0.58 | 0.44 | 0.063 | 0.08 | 0.77 |
| | (0.97) | | | (0.99) | | |
| $DIST_c*Small$ | -0.726 | 9.14 | 0.00 | -0.658 | 8.26 | 0.00 |
| | (11.83)*** | | | (12.20)*** | | |
| $DIST_c*Medium$ | -0.902 | | | -0.918 | | |
| | (9.05)*** | | | (8.71)*** | | |
| $DIST_c*Large$ | -1.148 | 8.28 | 0.00 | -1.196 | 6.53 | 0.01 |
| | (7.25)*** | | | (6.24)*** | | |
| $LANG_c*Small$ | 1.134 | 0.16 | 0.69 | 1.030 | 0.07 | 0.79 |
| | (3.28)*** | | | (3.52)*** | | |
| $LANG_c*Medium$ | 1.061 | | | 1.088 | | |
| | (2.40)** | | | (2.37)** | | |
| $LANG_c*Large$ | 1.317 | 0.53 | 0.46 | 1.691 | 3.48 | 0.06 |
| | (1.79)* | | | (2.35)** | | |
| $EMIGR_c*Small$ | 0.042 | 2.4 | 0.12 | 0.041 | 3.73 | 0.06 |
| | (1.74)* | | | (1.57) | | |
| $EMIGR_c*Medium$ | 0.075 | | | 0.074 | | |
| | (1.98)** | | | (2.13)** | | |
| $EMIGR_c*Large$ | 0.090 | 0.63 | 0.42 | 0.082 | 0.16 | 0.69 |
| | (2.16)** | | | (2.04)** | | |
| Firm fixed-effects | Yes | | | Yes | | |
| Observations | 55,784 | | | | | |
| Firms | 16,414 | | | | | |
| Destinations | 193 | | | | | |
| R-squared | 0.16 | | | 0.16 | | |

Robust t-statistics in parentheses, based on standard errors clustered by importing country.

*significant at 10%; **significant at 5%;*** significant at 1%. All continuous variables are in logs.

**Table A.1: Emigrants data**

| | | | | | |
|---|---|---|---|---|---|
| France | 619,847 | Philippines | 1,753 | Macao, China | 445 |
| Germany | 234,840 | Bermuda | 1,750 | Chile | 413 |
| United States | 212,318 | Cape Verde | 1,656 | Indonesia | 404 |
| Brazil | 170,210 | Angola | 1,555 | Morocco | 399 |
| Canada | 155,984 | Austria | 1,473 | Tanzania | 382 |
| Switzerland | 104,159 | Ghana | 1,427 | Kyrgyz Republic | 373 |
| Spain | 56,359 | Nigeria | 1,338 | Moldova | 373 |
| Mozambique | 55,520 | Cote d'Ivoire | 1,262 | Suriname | 371 |
| Venezuela, RB | 54,414 | Malaysia | 1,251 | Japan | 368 |
| Luxembourg | 41,722 | Guinea | 1,220 | Ethiopia | 322 |
| United Kingdom | 37,910 | Hong Kong, China | 1,170 | Croatia | 322 |
| Belgium | 21,371 | Netherlands Antilles | 1,005 | Cameroon | 315 |
| Pakistan | 21,302 | Kenya | 887 | San Marino | 314 |
| Zimbabwe | 19,729 | Sao Tome and Principe | 814 | Greece | 302 |
| Australia | 15,441 | Congo, Rep. | 776 | Yemen, Rep. | 298 |
| Kuwait | 10,411 | Norway | 769 | Costa Rica | 297 |
| Netherlands | 10,218 | Cuba | 768 | Mexico | 270 |
| Andorra | 8,873 | Guinea-Bissau | 766 | Turkmenistan | 235 |
| South Africa | 8,037 | Ecuador | 759 | Turkey | 225 |
| Russian Federation | 6,451 | Lebanon | 747 | Singapore | 213 |
| Italy | 5,901 | Algeria | 713 | Bahrain | 212 |
| Argentina | 5,840 | Denmark | 686 | Syrian Arab Republic | 196 |
| Uzbekistan | 5,059 | Uruguay | 680 | India | 187 |
| Jordan | 4,806 | China | 652 | Iraq | 178 |
| Israel | 3,986 | Ireland | 601 | Estonia | 176 |
| Ukraine | 3,656 | Colombia | 589 | Togo | 170 |
| Nepal | 2,876 | Zambia | 571 | Aruba | 164 |
| Sweden | 2,533 | Namibia | 566 | Faeroe Islands | 157 |
| Malawi | 2,446 | Romania | 508 | Burundi | 155 |
| Libya | 1,945 | New Caledonia | 506 | Belarus | 149 |
| Burkina Faso | 1,846 | Taiwan | 503 | New Zealand | 148 |
| United Arab Emirates | 1,841 | Tajikistan | 450 | Thailand | 143 |

**Table A1: Emigrants data (continued)**

| | | | | | |
|---|---|---|---|---|---|
| Finland | 141 | Lesotho | 45 | St. Vincent and the Grenadines | 12 |
| Panama | 137 | Somalia | 43 | Bosnia and Herzegovina | 11 |
| Kazakhstan | 127 | Cambodia | 42 | Jamaica | 11 |
| Seychelles | 123 | Paraguay | 42 | Korea, Rep. | 10 |
| Niger | 122 | Rwanda | 40 | British Virgin Islands | 8 |
| Iceland | 116 | Czech Republic | 39 | Bahamas, The | 7 |
| Madagascar | 115 | Grenada | 39 | Dominica | 7 |
| Swaziland | 113 | Bangladesh | 32 | Anguilla | 6 |
| Georgia | 102 | Gabon | 31 | Albania | 6 |
| Gambia, The | 99 | Benin | 29 | Maldives | 6 |
| Uganda | 99 | Bolivia | 28 | Solomon Islands | 6 |
| Senegal | 98 | Hungary | 28 | Northern Mariana Islands | 5 |
| Bulgaria | 96 | Eritrea | 27 | Botswana | 4 |
| Oman | 92 | Tonga | 22 | Latvia | 4 |
| Mali | 91 | St. Kitts and Nevis | 21 | Macedonia, FYR | 4 |
| Haiti | 86 | Korea, Dem. Rep. | 21 | Slovak Republic | 4 |
| Armenia | 83 | El Salvador | 20 | Samoa | 4 |
| Trinidad and Tobago | 83 | Tunisia | 20 | Antigua and Barbuda | 3 |
| Gibraltar | 81 | Comoros | 19 | Lithuania | 3 |
| Dominican Republic | 75 | Cyprus | 19 | Mauritania | 3 |
| Afghanistan | 71 | Mauritius | 18 | Cayman Islands | 2 |
| Poland | 70 | Brunei Darussalam | 17 | Equatorial Guinea | 1 |
| Iran, Islamic Rep. | 67 | Guyana | 17 | St. Lucia | 1 |
| Liberia | 67 | Sri Lanka | 16 | Monserat | 1 |
| Malta | 63 | Marshall Islands | 16 | St Hellen | 1 |
| French Polynesia | 63 | Barbados | 15 | Azerbaijan | 0 |
| Peru | 57 | Cook Islands | 15 | Central African Republic | 0 |
| Egypt, Arab Rep. | 55 | Fiji | 15 | Federal States of Micronesia | 0 |
| Djibouti | 53 | Guatemala | 15 | Qatar | 0 |
| Chad | 52 | Sierra Leone | 14 | Sudan | 0 |
| Saudi Arabia | 51 | Honduras | 12 | Slovenia | 0 |
| Papua New Guinea | 48 | Nicaragua | 12 | Tuvalu | 0 |
| Vietnam | 48 | | | | |

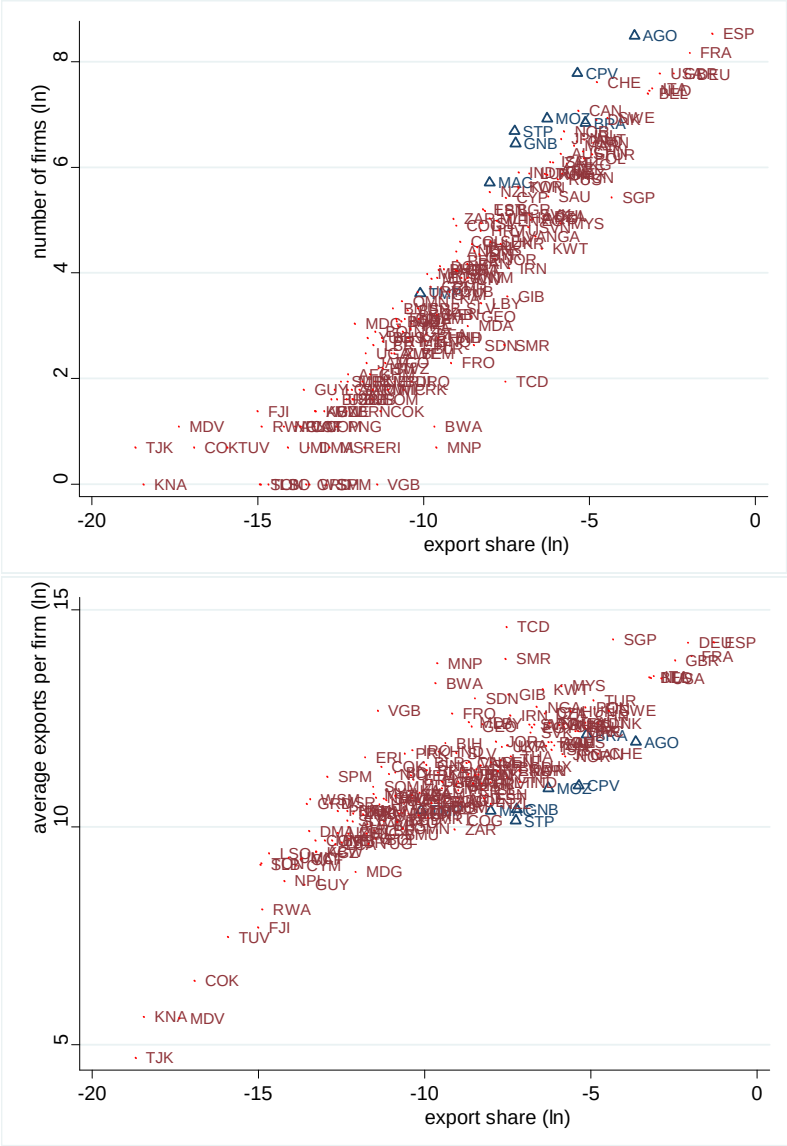


Figure 1: Decomposing bilateral exports (Portuguese-speaking countries identified by a triangle)