“Does Intraregional Trade Facilitate Export Diversification?”

TESIS PARA OPTAR AL GRADO DE

MAGÍSTER EN ECONOMÍA

Alumno: Esteban Verdugo Pedreros

Profesor Guía: Roberto Álvarez

Santiago, octubre 2017

Does Intraregional Trade Facilitate Export Diversification?

Esteban Verdugo Pedreros *

October 24, 2017

Abstract

Using a panel dataset that considers a large number of developing and developed economies, we find robust evidence supporting the hypothesis that higher levels of intraregional trade lead to a more diversified exporting structure. Employing a gravity based framework described in Felbermayr and Groschl (2013), we construct a time-varying instrument of intraregional exports that allow us to obtain evidence of the causal relationship between integration and diversification. We also show that intraregional trade of knowledge intensive products is highly efficient in terms of promoting diversification, mainly through the extensive margin. This finding is significant since it shed some light on the mechanism through which intraregional trade pushes a more diversified exporting structure. As Albornoz et al. (2012) maintains, firms realize their export potential from its experience in neighboring markets before expanding on to remote ones. Since knowledge intensive products are highly benefited by this learning by exporting process, intraregional flows of these type of products should promote diversification through the opening of new exporting lines. While the benefits of intraregional trade are often stressed in the literature, this paper focused on a gain on diversification that is rarely discussed.

Keywords: Intraregional Trade; Export Diversification; Gravity Equation
JEL Classification Codes: C23; F15; F43

*everdugo@fen.uchile.cl

A mis padres
y Triana

Contents

List of Figures

List of Tables

1 Introduction

In this paper, we study the relationship existing between the degree of regional integration of an economy and the qualities of its exporter sector. In particular, we address the effects that intraregional exports have on export diversification. Does a higher level of engagement in trade activities inside the region boost the emergence of new exporting products or exporter firms? As Albornoz et al. (2012) maintains, export experimentation theory proposes that firms are more likely to serve near destinations to their current active markets before going out of the region. Since nearby markets serve as testing grounds for product experimentation, a sort of learning by exporting process is developed. Expecting that this process is encouraged by regional integration, we anticipate that more intraregional trade diversifies the exporting structure.

The diversification of the exporting structure has long been one of the main topics in the policy discussions of developing economies. There are at least three main reasons behind this concern. First, higher levels of export diversification reduce macroeconomic volatility and vulnerability to external shocks (Di Giovanni and Levchenko, 2010 and Koren and Tenreyro, 2013). Second, more diversification seems to boost the local productivity (Melitz, 2003). And third, a more diversified exporting matrix leads to a faster pace of economic growth (Agosin, 2007; Lederman and Maloney, 2007, and Hesse, 2008). Although most of the literature maintains that diversification is something highly desirable, there are few empirical works analyzing its driving factors (Agosin et al., 2012 and Parteka and Tamberi, 2012). A main goal of this paper is to shed some light about the diversification determinants, but differing from the existing body of research by focusing on the role of regional integration.

To analyze properly the causal relationship existing between intraregional trade and diversification, we take into account the reversal causality existing between both variables. We follow the gravity equation proposal made by Felbermayr and Groschl (2013). This strategy operates in a panel framework, and allow us to implement a cross-country, long time panel scope. Our largest dataset includes 90 developing and developed economies over a period of

56 years. This methodology provides an exogenous time-varying instrument of high quality for intraregional exports.

As our main result, we find a strong and positive causal relationship between the share of intraregional exports of a country and the diversification of its export structure. This relationship holds through a battery of robustness checks: among others, we reduce the sample size to obtain a balanced panel, and we select alternative measures of export diversification as the dependent variable. These results allow us to provide robust evidence on the positive and causal relationship between intraregional exports and diversification. As a complementary result, we confirm a robust relevance of exchange rate policies on the export sector: an undervaluated exchange rate boosts the export diversification process. Moreover, we find evidence that more freedom to trade, better local institutions and higher levels of capital endowment boost diversification. However, these effects do not seem to be highly robust.

As a complementary measure, we study the transmission mechanism of our main result. Additional regression seems to confirms that the diversification process motivated by intraregional exports occurs mostly at the extensive margin. This means that higher intraregional exports augment the number of markets served and products exported during time. Also, differentiating products between primary and non-primary, we find evidence that diversification is motivated mainly by the intraregional export of non-primary products. Since the former are more heterogeneous and knowledge intensive than the latter, these results seem to suggest that diversification is linked to a learning by exporting process. These results are highly consistent with what export experimentation theory postulates.

The paper proceeds as follows. The second section describes the data, while section 3 outlines our empirical strategy. The results are presented in the section 4. Last section concludes.

2 Data, Definitions and Stylized Facts

2.1 Data and Definitions

We begin this section remembering that the spirit of this empirical exercise is to evaluate the role that intraregional trade has over export diversification using a cross-country, long time panel approach. For methodology purposes, in our first econometrical section we exploit a panel dataset for a large number of developing and developed economies covering the 1960-2015 period, while in our second section we use five-year non-overlapping averages of the annual data covering the same sample. The data on bilateral exports measured in current USD comes from the International Monetary Fund's (IMF) Direction of Trade Statistics (DoTS): this dataset covers all IMF members and some non-member economies with exports reported on a free on-board basis. Regarding geographical variables such as land area surface, bilateral distance, or contiguity status, those are obtained from the Centre d'Estudes Propesctives et d'Informations Internationales's (CEPII) Geographic and Bilateral Distance Database. Our preferred measure of bilateral distance uses latitudes, longitudes, and population data to estimate each main agglomeration's position: weighting bilateral distances between the biggest cities of every two-pair of countries to compute their bilateral country distance. As a robustness check, we employed simpler measures such as the great circle formula, but results were not significantly different using any of these alternative approaches[1]. Following Rose and Spiegel (2009), we compute financial remoteness as the logarithm of great circle distance to the closest major financial center (London, New York, or Tokyo). Since logarithm of zero value is not well defined, we imposed zero financial remoteness for those economies where financial hubs are located (United Kingdom, United States, and Japan, respectively). Meanwhile, details on population size and crude death rate come from the World Bank's World Development Indicators (WDI).

Data on natural disasters are taken from the Emergency Events Database (EM-DAT) produced by the Center for Research on the Epidemiology of Disasters (CRED). This set contains

[1] See Mayer and Zignago (2011) for a complete description about the different measures of bilateral distance included in Geographic and Bilateral Distance Database.

information on the occurrence and effects of over 22,000 mass disasters in the world starting on 1900. It is a particularly rich panel since it allows to identify the number of persons injured and dead organized by the type of disaster affecting. Such as Felbermayr and Groschl (2013) clarifies, technological and complex disasters could be easily linked to economic variables, therefore, we exclude these kind of disasters and focus exclusively on the ones that are highly presumable to be orthogonal to any economic factor[2]. In alternative regressions we work with broader definitions of "large" disaster, still excluding those of technological and complex nature.

Regarding our second econometrical section, the data on the dependent variable is taken from IMF's The Diversification Toolkit: Export Diversification and Quality Databases (2014). Among other variables, this dataset provides three Theil indexes on export diversification representing overall as well as extensive and intensive margins: higher values for all three indexes denote lower diversification levels[3]. In particular, the extensive margin index is calculated as:

$$diversification_t^e = \sum_{i \in G} \frac{n_{i,t}\mu_{i,t}}{n_t \mu_t} \times \ln\left(\frac{\mu_{i,t}}{\mu_t}\right) \qquad (1)$$

Where:

$$\mu_{i,t} = \frac{1}{n_{i,t}} \sum_{j \in i} \chi_{j,t} \qquad \text{and} \qquad \mu_t = \frac{1}{n_t} \sum_{i \in G} \sum_{j \in i} \chi_{j,t}$$

With $\chi_{j,t}$ denoting the export value of line j during time t, and sub-index i representing each group of traditional, new, and non-traded products[4]. While $n_{i,t}$ denotes the number of open lines in each one of these groups, n_t indicates the number of total lines open in the whole exporting basket. It can be demonstrated that diversification at the extensive margin occurs

[2] We adopt Felbermayr and Groschl (2013) definition of "large" droughts, wildfires, earthquakes, mass movements, volcanic activities, and storms. This definition guarantees that no natural disaster is caused by any local or global economic factor. For details about this definition see appendix B.

[3] See Cadot et al. (2011) for a complete description about different dimensions of export diversification and how to compute some measurements.

[4] Traditional products are those goods that were exported since the beginning of the sample, non-traded products are those goods that were not exported during the whole sample, and new goods are the remaining ones.

when the number of lines open rises. Similarly, the intensive margin index is computed as:

$$diversification_t^i \quad = \quad \sum_{i \in G} \frac{n_{i,t}\mu_{i,t}}{n_t \mu_t} \times \left(\frac{1}{n_{i,t}} \sum_{j \in i} \frac{\chi_{j,t}}{\mu_{i,t}} \times \ln \left(\frac{\chi_{j,t}}{\mu_{i,t}} \right) \right) \tag{2}$$

Clearly, this is a weighted average of group specific degrees of diversification with weights being $n_{i,t}\mu_{i,t}/n_t\mu_t$. It can be demonstrated that a reduction in concentration among active lines implies higher intensive diversification levels. In sum, the extensive margin reflects variations in the number of lines open, whereas the intensive margin reflects variations in the distribution of export values among those lines already opened. Finally, the overall Theil index ($diversification_t^o$) is calculated just as the sum of its extensive and intensive components. This whole set of indexes is computed using the 4-digit Standard International Trade Classification revision 1 (SITC rev. 1) as the measure of commodity classification.

Regarding our set of factor endowment used as control, we proxied human capital accumulation using the average years of schooling of the population. The source of this data can be found in Barro and Lee (2013), this database provides details on educational attainment and years of schooling for 146 countries in five-year intervals since 1950. We selected the average years of total schooling for the population aged 25 and over of every country as its human capital endowment value. Unfortunately, this variable represents just a quantity base measure of education, quality based measures have proven to capture capital accumulation in a better way, but this type of data is not suited for a cross-country, long time panel which is the spirit of this analysis. Data on capital stock per capita is computed using IMF's Investment and Capital Stock Dataset (2015). This set provides comprehensive details on the stock of public and private capital measured in billions of constant 2005 international USD. We obtain per capita figures adjusting these values by population size. Since we consider as relevant those institutions that directly shape market production and trade behavior, we focus on the freedom to trade internationally and the legal system and property rights quality of a country as proxies of its overall institutional quality. Data on both indexes come from The Fraser Institute's Economic Freedom dataset and details can be found in Gwartney et al. (2016). Meanwhile, data on real gross domestic product per capita in constant 2010 USD is taken from World Bank's WDI database.

Data on exchange rate related variables come from Darvas (2012). This dataset includes details on nominal and real effective exchange rate at monthly and annual frequency for a broad set of countries. Real exchange rate overvaluation is computed using Rodrik (2008) methodology. Defined this way, the overvaluation variable is comparable across countries and over time. If this variable has a positive value, it indicates that the real exchange rate is set such that domestic goods are relatively cheap; if negative, then local currency is undervaluated. This index is essentially a real exchange rate variable adjusted for the Balassa-Samuelson effect: it adjusts the price of tradables to non-tradables considering that as countries get richer, the relative price of non-tradables tends to rise since they get higher productivity levels. Finally, data on the terms of trade index is taken from World Bank's WDI dataset.

Some deeper analyses about potential heterogeneous effects are performed using a more disaggregated bilateral export dataset, with Robert Feenstra's World Trade Flows (WTF) as a source. This dataset is constructed by Feenstra et al. (2005) and includes bilateral trade information at commodity level, organized by the 4-digit Standard International Trade Classification revision 2 (SITC rev. 2), over the 1962-2014 period with values measured in current USD. This information is more reliable than the one provided by the United Nations Comtrade database because, among other things and when available, it is constructed with importer records which are more trustworthy than the exporter ones. Since this and other checks have not been extended to post-2000 data, some robustness checks using the WTF dataset are performed with pre-2000 data only.

Country grouping is primarily based on the World Bank's regional classification scheme. Following this international organization, non-high-income economies are spanned into six regions: East Asia and Pacific (EAP), Eastern Europe and Central Asia (ECA), Latin America and the Caribbean (LAC), Middle East and North Africa (MENA), South Asia (SAR), and Sub-Saharan Africa (SSA). We split the seventh group of high income economies into two additional regions: North America (NAM) and Western Europe (WEU). It is important to notice that every country in the sample is member of just one of these eight regions[5]. We recognized that every scheme comes with a set of limitations and biases, but this classification seems to provide a reasonable method of organizing data where economic sense prevails.

Finally, we focus our analysis on two different country samples. Although resulting in an unbalanced panel, our preferred selection maximizes sample size to include several countries from every region of the world. Altogether, this set of 90 economies accounts for more than 70% of the world exports over the 1960-2015 period and, as Figure 1 presents, just 13% of its country-pair exports are missing values. Some robustness checks are performed with an additional sample that considers a strongly balanced panel of 58 countries where just 1% of its bilateral export flows are missing values. This set includes economies from all over the regions and excludes global dominant oil-producers and almost every ex-Soviet Union member[6]. To close this sub-section, we point out that a detailed description of the variables employed in the estimation process is presented in appendix B.

2.2 Stylized Facts and Descriptive Statistics

Regarding some empirical facts derived from our dataset, we begin by analyzing the evolution of intraregional exports using regional values from our main sample[7]. A general glance of Figure 2 reveals the existence of clear differences across regions. To begin with, the high degree of heterogeneity existing in path trends stands out: while some regions present steady paces of integration, others display flat paths. The data shows that EAP and NAM start with figures close to 30% and end with values above 50%, whereas MENA and SAR remain always close to its initial value. Likewise, the figure also reveals important differences in the

[5] Although according to the raw classification Mexico belongs to LAC, we place this economy in NAM region. A notorious high level of economic integration justifies this decision: influenced by the North America Free Trade Agreement (NAFTA), today Mexico is the second destination of United States' exports and the United States is the top destination of Mexico's exports being Canada the next one.

[6] Recognizing that this alternative sample is small, we remark that its strongly balanced condition is something desirable because of the panel methodology here employed. For details about the list of countries included in both samples, see Table 1.

[7] Intraregional exports at regional level is computed as the aggregate exports that economies members of some region destine to economies members of the same region relative to the sum of its whole export activity.

degrees of integration across regions: on one end, MENA and SAR remains steady with figures permanently lower than 10%, on the other, WEU shows values consistently above 60%. Complementing, Panel (a) of Figure 3 reaffirms the presence of these differences. In this figure, EAP and NAM are placed well above the diagonal line, suggesting more integration over time, and there is also a clear long distance separating WEU from the regions of MENA and SAR. Regarding LAC, this region slowly rose its integration during the sample period: it begins with values close to 10% and ends with figures over 20%. Although NAM is its main historical trading partner, at the end of the sample NAM shares this position with EAP.

In respect to the diversification variable, Panel (b) of Figure 3 depicts the evolution of the overall Theil index using regional simple averages. Since most of them lie above the diagonal, except for WEU, this figure reveals that almost every region improved its diversification situation during the second half of the sample. This improvement appears to be more notorious in the regions of EPA and LAC. Besides that, important differences of level persist: the regions of MENA and SSA still show figures significantly higher than ECA, NAM, and WEU. Analyzing the results found for both integration and diversification, the existence of a high level of heterogeneity across regions is clear, but more important, it seems that the regions that show better diversification levels are the more integrated ones. The most extreme cases correspond to MENA and WEU: each one represents an equilibrium of high levels in one variable and low levels in the other one. Using data at a country level, Figure 4 and Figure 5 confirm these findings.

[Insert Figure 6 here.]

In order to keep motivating our research question, Figure 6 offers more insight into the existing relationship between integration and diversification by plotting intraregional export values against the overall Theil index. While in the left panel this image is presented at country level, the one on the right does it at a regional level. As expected, this figure reveals a clear negative relationship for both exercises: more integration seems to be highly tied to a better diversification status. In a similar test, Figure 7 presents country by country correlations between both variables: although nearly 75% of the countries reveal negative figures, there are economies with high positive values, suggesting some degree of heterogeneity in the outcomes. Figure 8 clarifies the results revealing that most of the economies with positive correlations are members of the more developed regions such as EAP, ECA, and WEU, while developing regions such as LAC, SAR, and SSA, host most of the countries with negative values. This result suggests that poor regions could be more beneficed by the integration process. Even though these general exercises are just raw results from which no causal relationship can be inferred, it is encouraging to see that our predictions are confirmed by such an unprocessed cut of the data.

To close this section, we present two tables summarizing the main descriptive statistics for our baseline dataset. Table 2 describes the data employed in our first econometrical section, whereas Table 3 focuses on the data used on the second section. It is important to notice that the data shows that the average country suffers 0.22 "large" natural disasters every year, a figure that has associated a notoriously high standard deviation of 30%. As we explain in detail in the next section, this degree of volatility is important since the exogenous variation that natural disasters bring to bilateral trade flows is what allows us to construct our necessary time-varying instrument for intraregional trade.

3 Empirical Strategy

3.1 Baseline Specification

In this section we discuss the identification strategy and estimation methodology employed to carry out this investigation. We begin by presenting the empirical specification used to identify the effect that intraregional trade has over diversification.

$$diversification^{i}_{\tau} \;=\; \xi_1 + \xi_2 \times diversification^{i}_{\tau-1} + \xi_3 \times \Omega^{i}_{\tau} + \xi_4 \times \Im^{i}_{\tau} + \varsigma^{i} + \varsigma_{\tau} + \varepsilon^{i}_{\tau} \quad (3)$$

In this equation, $diversification^{i}_{\tau}$ denotes a measure of export diversification for country i during period τ. It is important to remark that for every measure of diversification that we used, higher values denote lower diversification levels. Since we are primarily interested in the emergence of new exporting products and exporter firms, i.e. diversification at the extensive margin, we decide to employ the Theil index as our main diversification measure. As we explained above, one of the main advantages of this index is that it is a weighted average of inequality within groups plus inequality between those groups, so it can be easily decomposed in its intensive and extensive components. While our main regression employs the overall index, further analyses are made with both sub-components. Because of its great persistence over time, our favorite model considers the dependent variable lagged one-period as a regressor. This procedure helps to mitigate problems of omitted variable bias and allows to eliminate residual serial correlation.

Meanwhile, intraregional trade is included via the term Ω^{i}_{τ}. Since our interest is focused on exports rather than overall trade, Ω^{i}_{τ} actually denotes aggregate exports that country i destines to economies members of its own region relative to its whole export activity. This means that import flows are excluded from the analysis and that ξ_3 is our main parameter of interest[8]. As Albornoz et al. (2012) maintains, firms are more likely to serve near destinations to

[8] Since our analyses of the mechanism through which regional integration push a more diversified exporting structure are closer to the export experimentation theory, focusing only on export flows is more suitable to our empirical exercise. Moreover, the basic gravity theory explains the value of spending by one economy on the goods produced by another, that is, our primal reference theory is made about unidirectional flows only.

their current active markets before going out of their neighborhood. This conclusion suggests that nearby markets serve as testing grounds for product experimentation and that higher levels of intraregional exports could lead to a more diversified exporting matrix. Following this prediction, we anticipate a negative sign for the corresponding parameter.

Regarding time-varying controls, those are considered by $\Im^i_\tau$ matrix. We divide this set of variables into three categories: indicators capturing factor endowment bulk, variables related to institutions or long-term policy, and exogenous good or bad luck factors. In respect to this first group, the classical Hecksher-Ohlin model sustains that economies have comparative advantages in those products for which the required factors of production are relatively abundant at a local level, thus production and export patterns are largely determined by endowment variables. Our set of factor endowment considers indicators related to the stock of human and physical capital. For these variables, recent theoretical works are available that provide microeconomic foundations about their potential effects under a heterogeneous firms framework. Regarding human capital, the common feeling is that its accumulation allows economies to reallocate their factors more easily across industries, accelerating changes on the pattern of specialization (Melitz, 2003 and Bernard et al., 2007). This conclusion of more plasticity in the producing and exporting matrix suggests that we may anticipate a negative coefficient on the estimation of the $education^i_\tau$ variable. Similarly, Acemoglu and Zilibotti (1997) proposes a model that links scarcity of capital accumulation and limited diversification opportunities in a context of market incompleteness. Owing to the scarcity of capital, only a limited number of imperfectly correlated projects are undertaken since agents avoid risky investments and opt for a small group of safe but less productive assets. As a result, countries with poor infrastructure endogenously present higher concentration levels.

Regarding variables capturing institutional or political affinities to trade, they are represented by the set $\{$ $free\ to\ trade^i_\tau$, $institutions^i_\tau$, $volatility^i_\tau$, $overvaluation^i_\tau$ $\}$. The literature sustains that institutional differences act as a major source of trade (Egger et al., 2008, 2011, and Levchenko, 2007) and that contract enforcement sometimes explains even a bigger portion of bilateral flows than the endowment bulk (Nunn, 2007). Since we are interested in those institutional elements that directly shape market production and trade behavior, we focus on two areas to proxy the institutional quality of an economy: the level of freedom to trade internationally ($free\ to\ trade^{i}_{\tau}$) and the state of the legal system and property rights ($institutions^{i}_{\tau}$). Regarding the freedom to trade, Dornbusch et al. (1977) proposes a model where tariff reductions by two trading economies rise the diversification levels for both partners. Since a tariff reduction means lower trading costs for every product in the economy, ex-ante non-traded products can now be traded because of its ex-post lower cost, expanding the exporting basket. More recently, Melitz and Ottaviano (2008) develops a monopolistically competitive model with firm heterogeneity where the variety of products increases because of liberalization via a pro-competitive mechanism. In this setting, trade liberalization implies integrated markets that exhibit larger and more productive firms as well as higher product variety, lower prices, and lower mark-ups. On the other hand, Miranda and Wagner (2015) finds evidence pointing that contract enforcement institutions in neighboring economies affect significantly the local pattern of specialization. Employing an empirical framework, it concludes that better levels of rule of law in neighboring partners encourage spillover mechanisms that allows higher degrees of diversification mainly through the development of contract intensive industries. In conclusion, we anticipate negative coefficients for both of these institutions related variables.

Regarding exchange rate volatility, its positive expected sign comes from the fact that theory usually considers volatility as a variable cost that implies uncertainty and generate persistence in firm decisions, damaging the diversification process (Melitz, 2003). In particular, Baldwin and Taglioni (2004) develops a model where exchange rate uncertainty is highly detrimental for the number of products and firms in the economy. Likewise, Hausmann and Rigobon (2003) suggest that volatility produces a vicious cycle since it discourages investment in non-resource tradables and boosts concentration in the resource based sector bringing even more volatility into the real exchange rate. On the other hand, since an appreciated real exchange rate reduces real profitability of the exporting firms, Agosin et al. (2012) proves overvaluation as an increase in the trading costs in the model proposed by Melitz (2003). These higher costs make products with marginal comparative advantages less likely to be produced and exported reducing diversification. In the same fashion and employing some empirical evidence and a simple model of real exchange rate and growth, Rodrik (2008) argues that tradables suffer more than non-tradables from institutional weakness and market failures. Thus, an increase in the relative price of tradables alleviates some of these distortions and encourages a desirable structural change: undervaluation boosts diversification through its expansion of the share of tradables in the economy. As a result, we expect a negative role of $overvaluation_{\tau}^{i}$ on the diversification index.

The literature usually associates improvements in the terms of trade with higher concentration levels. This hypothesis is consistent with the broadly covered Dutch disease phenomenon where commodity booms imply general equilibrium effects that reduce the size and heterogeneity of the tradable sector (Corden and Neary, 1982). In this context, price and income effects induce factor reallocation towards those sectors facing improved export opportunities reducing diversification. However, Agosin et al. (2012) recognizes that under the model proposed by Melitz (2003), terms of trade improvements could be interpreted as an increase in the profitability of the exporting firms. In this context, income effects imply an expansion of the exporting basket. Under this frame of ambiguous theoretical effects, Agosin et al. (2012) finds evidence that a positive shock on terms of trade implies more concentration.

To close the description of equation (3), we point that our specification properly controls for country and year fixed-effects eliminating the need for many time-invariant and common time-variant controls (ς^{i} and ς_{τ}, respectively). Furthermore, to purge the data from the influence of business-cycle and to identify a clean long-run relationship, we divided our whole sample period into ten sub-periods of non-overlapping five years each to which simple averages are applied. We implement this procedure to every variable included in our dataset and use the sub-index τ to denote these sub-periods. Among other things, this common operation also allows the reduction of outliers' influence over the results. As a robustness check, we ran additional regressions with an annual frequency dataset denoted by the sub-index t.

Finally, we point that the results obtained from a simple OLS estimation of equation (3) may be inconsistent because the correlation existing between the error term (ε_{τ}^{i}) and our trade related regressor. While the inclusion of the lagged dependent variable as regressor, the set of controls selected, and the inclusion of fixed-effects, alleviates some of the concerns related with omitted variables, our results could still be biased since both main variables present potential endogeneity problems. There are at least two reasons that explain our endogeneity concerns. First, there could be unobserved conditions that simultaneously affect the degree of regional integration and export diversification of an economy. And second, and more important, there could be a process of reverse causality between both variables. As a mirror of the export experimentation theory, diversification improvements imply a strengthening of the exporting lines of an economy, process that occurs through a higher number of open lines or a more even distribution between export values, since firms are more likely to strength near destinations to their current market before going out of its neighborhood, higher diversification levels may lead to higher intraregional exports. As we explain in the next sub-section, employing the instrumental variables setup proposed by Felbermayr and Groschl (2013) helps us mitigate this concern.

3.2 Instrument Construction

In this sub-section we discuss in detail the identification strategy here followed. As we mentioned above, our empirical specification does not consider the probable presence of reversal causality in both main variables, leaving open the possibility for the results to be biased. To tackle the potential endogeneity, we adopt the panel generalization of the gravity equation proposed by Felbermayr and Groschl (2013). Summarizing, it declares that natural disasters bring exogenous variation in bilateral trade flows allowing the implementation of a panel IV methodology to construct a time-varying instrument for intraregional trade. This procedure allows to capture export elements driven only by exogenous factors, clearing the data from any endogeneity problem. The inclusion of "large" foreign natural disasters and the exclusion of the domestic ones, ensures the validity of the instrument. We remark that our key identifying assumption is that foreign natural disasters have no effect over local diversification besides its bilateral trade influence.

The estimated specification is based on the following equation[9]:

$$\chi_t^{i,j} \quad = \quad \exp\left(\varphi_1 \times \mathcal{D}_t^j + \varphi_2 \times \left(\Theta_t^{i,j} \times \mathcal{D}_t^j\right) + \varphi_3 \times \Xi_t^{i,j} + \omega^i + \omega^j + \omega_t\right) + \upsilon_t^{i,j} \qquad (4)$$

Where $\mathcal{D}_t^j$ denotes the number of "large" natural disasters that hit the importer country j during year t. The matrix $\Theta_t^{i,j} = \{\ financial_t^j\ ,\ area^j\ ,\ population_t^j\ ,\ adjacency^{i,j}\ \}$ considers importer financial remoteness, its surface area and its population size. Details on the contiguity status between the country-pair (i, j) are captured by $adjacency^{i,j}$: it takes the value of one if the economies in question share a common land border and zero otherwise. On the other hand, the matrix $\Xi_t^{i,j} = \{\ population_t^i\ ,\ population_t^j\ ,\ distance^{i,j}\ ,\ adjacency^{i,j}\ ,\ language^{i,j}\ \}$ considers importer and exporter population mass, any measure of geographical distance between the country-pair (i, j), the adjacency status between them, and a language dummy that takes the value of one if this country-pair share a common official language. Based on Frankel and Romer (1999), we presume $\{\Theta_t^{i,j}, \Xi_t^{i,j}\}$ as a set of orthogonal controls.

We include time dummies to capture common shocks faced by the whole sample of countries during the same period (ω_t). As Anderson and van Wincoop (2003) suggests, we also control properly for multilateral resistance terms by including exporter and importer fixed-effects (ω^i and ω^j, respectively). Furthermore, we estimate applying standard errors clustered at country-pair level with $\upsilon_t^{i,j}$ being the error term. It should be noticed that the proposal of Felbermayr and Groschl (2013) diverges from the traditional and micro-founded gravity specification. Based on an export pull factor dynamic, this strategy uses exogenous shocks in the importing country (such as droughts, wildfires, earthquakes, mass movements, volcanic activities, floods, landslides, and storms) to identify variations in local flows.

Since trade data are always prone to the presence of heteroskedasticity, our gravity equation is evaluated using the methodology of poisson pseudo-maximum likelihood (PPML) as the estimator[10]. As Santos Silva and Tenreyro (2006) declares, this non-linear method is ro-

[9] Arguments that supports the inclusion of this set of variables as regressors and its interaction with $\mathcal{D}_t^j$, can be found in Felbermayr and Groschl (2013).

[10] The conditional variance of trade data goes to zero as we approach to zero conditional expected value. In other words, bilateral flows in small economies show lower volatility than in larger economies allowing the presence of heteroskedasticity.

bust to different patterns of heteroskedasticity and also provides a natural way to deal with zero values on the trading variable ($\chi_t^{i,j}$)[11]. While some empirical works estimate log-linear models using ordinary least squares (OLS), thus dropping trade data with zero value since logarithm of zero is not well defined, others estimate adding a common minimum positive value to every data on the dependent variable. Santos Silva and Tenreyro (2006) concludes that, under the presence of heteroskedasticity, both procedures are not negligible and lead to inconsistent estimators[12]. This work suggests that the PPML estimator should be used as a substitute for the estimation of any standard log-linear model.

We clarify that the results from this estimation represent only a preliminary step necessary for the construction of our instrument. We employ the resulting set of estimated coefficients of the gravity equation to predict exogenous export flows. After constructing our exogenous intraregional exports variable, we proceed to estimate equation (3) using the methodology of two-stage least squares (2SLS). Thus, the goal of this section is to obtain the highest correlation between export flows and our set of exogenous controls over time to use this variation to predict the exogenous component of trade[13].

$$\hat{\Omega}_t^i = \sum_{j \in R(i)} \hat{\chi}_t^{i,j} / \sum_{k \in W} \hat{\chi}_t^{i,k} \tag{5}$$

Using the fitted data from the estimation of the equation (4) ($\hat{\chi}_t^{i,j}$), we construct an exogenous instrument $\hat{\Omega}_t^i$ employing the procedure illustrated by the equation (5). There, $R(i)$ denotes the set of countries member of the same region that country i belongs to and W denotes the whole sample set. In sum, $\hat{\Omega}_t^i$ proxies the intraregional export share Ω_t^i based on gravity equation predicted bilateral flows[14]. Again, this constructed variable is free of any reverse causality regarding the export diversification index. While the validity of this

[11] Another peculiarity of bilateral trade data is that it is also prone to the presence of a large mass of zero observations. This value represents country-pairs without bilateral trade between them. As Figure 1 presents, more than 24% of the trade values of our main sample are zeros.

[12] Santos Silva and Tenreyro (2006) also proves that the popular non-linear least squares (NLS) is also very inefficient under the presence of heteroskedasticity. By construction, heteroskedasticity implies that this procedure gives more weight to noisier observations, so results depend heavily on a small, noisy share of the sample.

[13] In that sense, Felbermayr and Groschl (2013) call its proposal of modified gravity equation as a "data reduction device".

[14] Endogenous intraregional trade variable is constructed using $\Omega_t^i = \sum_{j \in R(i)} \chi_t^{i,j} / \sum_{k \in W} \chi_t^{i,k}$

procedure requires exogeneity of every regressor in (4), the quality of the instrument is defined just by its correlation with Ω_t^i. The final instrument $\dot{\Omega}_t^i$ is obtained via first stage procedure of the 2SLS estimation. As equation (6) shows, first stage procedure consists in regressing Ω_t^i against $\hat{\Omega}_t^i$ and the whole set of control variables of equation (3) (Π_τ^i), fitted values of this estimation correspond to our final instrument $\dot{\Omega}_t^i$.

$$\Omega_\tau^i = \alpha_1 + \alpha_2 \times \hat{\Omega}_\tau^i + A \times \Pi_\tau^i + \varsigma^i + \varsigma_\tau + \epsilon_\tau^{1,i} \tag{6}$$

With:

$$\Pi_\tau^i = \begin{bmatrix} diversification_{\tau-1}^i \\ \Im_\tau^i \end{bmatrix}$$

In addition to the procedures described above, we take additional measures to limit further concerns about the exclusion restriction and endogeneity problems. Some robustness checks are performed including variables such as exporter logarithm of gross domestic product per capita or the number of preferential trade agreement in which the exporter economy is a member. These checks controls more carefully for possible income effects. Furthermore, some regressions are estimated with the whole set of independent variables lagged one period alleviating threats of reversal causality in other variables beside intraregional exports. Finally, the number of natural disasters that hit the exporter country is considered as a regressor in the second stage estimation. This strategy shows that the impact of intraregional exports on diversification is not assignable to a direct impact of disasters, and also mitigates alarms that importer and exporter natural disasters are correlated in countries member of the same region. More details about these and others robustness checks are presented in the next section.

4 Results

We begin this section by showing broad insights into the causal relationship between intraregional trade and export diversification. As a benchmark, Table 4 presents the results of a simple OLS estimation of equation (3) using actual intraregional exports (Ω^i_τ) as a regressor. Once more, it is important to remark the following: since we presumed the presence of endogeneity, this estimation is recognized as a naive approach that could lead to biased results. We build from the simplest specification in column (1) to our preferred model in column (6). The overall Theil index is the dependent variable in each one of these regressions, and for methodology purposes explained in section 3, averages of non-overlapping five years are applied to the whole set of variables.

[Insert Table 4 here.]

Column (1) estimates the relationship between both main variables controlling only by country and time fixed-effects. As we expected, this coefficient is negative and highly significant: it reveals that a rise of one percentage point in intraregional exports significantly reduces concentration by 0.013 units in the Theil index scale. Likewise, column (2) demonstrates that this relationship holds at the inclusion of the lagged dependent variable: although a slight reduction in its absolute value, the sign and significance of the estimated coefficient are not altered. This last regression also reveals the great persistence of export diversification over time. Finally, columns (3) to (6) demonstrate that the inclusion of the set of control variables does not alter either the signs or significance of both already mentioned coefficients.

In general terms, this simple empirical exercise shows that intraregional exports seem to be a highly significant determinant of diversification: for every model, this variable is significant at 1% level. Results also suggest that diversification seems to be highly persistent over time: with a significant coefficient, an increase of one unit on the lagged dependent variable implies an increase on the interval of 0.633-0.676 on its actual value. Regarding the set of control variables, results are consistent with theoretical predictions and what previous literature finds:

every coefficient has the expected sign and most of them are highly significant. We stress again that these results do not consider the presence of endogeneity, leaving open the possibility for the estimators to be biased. To identify more carefully a clear causal relationship, in the following sub-sections we examine whether above results hold in the context of IV procedure and under a battery of robustness checks.

4.1 The Gravity Equation

Table 5 presents the results from estimating our baseline model of the panel type gravity equation. As Felbermayr and Groschl (2013) and Blanchard and Olney (2017) remark, the goal of this econometrical step is to achieve the best possible fit, from completely exogenous regressors, to construct our instrument. Here, the number of "large" natural disaster that hit the importer country j, i.e. $\mathcal{D}_t^j$ variable on the equation (4), is denoted by $natural_t^j$. Since bilateral exports is the dependent variable, we selected the PPML non-linear technique as the estimator. As we mentioned in section 3, this method is robust to different patterns of heteroskedasticity and also provides a natural way to deal with zero values on the dependent variable. Robust standard errors are clustered at country-pair level, and importer, exporter, and time fixed-effects are included but not reported.

[Insert Table 5 here.]

Table 5 shows that, as a direct effect, disasters in the importer economy j imply a statistically significant reduction in its demand of products coming from economy i ($\chi_t^{i,j}$). Results reveal that the presence of an extra foreign natural disaster implies a reduction of 1.376% on the local exports destined to this affected trading partner. However, the final effect is conditioned by the importer size on surface area and population. Similar to what Felbermayr and Groschl (2013) finds, if independent variables take its mean value, foreign natural disasters still reduce local exports.

As we explained above, the traditional gravity model affirms that trade between any pair of countries is proportional to the product of its economic size and it is inversely related to the physical distance existing between them. Here, we introduce population mass as a broad proxy of GDP size. Regarding this variable, results show that the coefficient of the logarithm of population is positive and highly significant for both importer and exporter economies. Since these coefficients imply an elasticity lower than the unit, results are consistent with the hypothesis that smaller economies tend to be more open to international trade. Similarly, this distance variable shows a large and significant negative coefficient. Consistent with the traditional gravity theory, recent micro-founded theoretical works consider that distance operates as a variable cost on trade, reducing bilateral flows (Melitz, 2003; Baier and Bergstrand, 2009). Regarding our second geographical proximity variable, results goes in the same direction: sharing a common land border significantly rises bilateral flows.

In sum, Table 5 is coherent with gravity equation predictions and typical empirical results: bilateral flows are positively related to the product of the size of both trading partners and to the geographical proximity existing between them. But the goal here is to find the best possible fit for the construction of our instrument. In this sense, results affirm that 86.7% of the bilateral flows variation is explained by the set of independent variables considered. Using fitted data from this estimation, we proceed to construct our exogenous instrument computing equation (5). Since the gravity equation predicts bilateral trade flows, those values can be aggregated across all trading partners to obtain the share of exports that each country directs to each region. Regarding the share of the region from which the exporter country is a member, it will proxy the exogenous component of the intraregional exports variable. As we mentioned above, the validity of this procedure requires the exogeneity of every regressor, and the quality of the instrument is defined just by its correlation with the endogenous variable.

[Insert Table 6 here.]

Table 6 presents some insights into the quality of the instrument derived from the first stage of 2SLS estimation ($\dot{\Omega}^i_\tau$). The constructed share coefficient is always large, positive, and highly significant. Even after including the whole set of fixed-effects and second stage controls, it remains significant at 1% level with a value close to the unit. Equally important, our instrument seems to be highly strong. Running some F-tests on excluded instruments proves that the null hypothesis of weak instrument is rejected not only with the p-value information (Cragg and Donald, 1993), also the figures associated to the statistic are well above the rule of thumb critical values proposed by Stock and Yogo (2005)[15]. Although the instrument attributes need to be evaluated in each estimation on their own, first stage general results show that $\dot{\Omega}^i_\tau$ seems to be an exogenous instrument of high quality. To identify more carefully the real causal relationship between intraregional trade and diversification, in the next sub-section we proceed to examine second stage results of this 2SLS estimation.

4.2 Second Stage Regression

To rule out any endogeneity concern on the trade related variable, this sub-section employs the exogenous instrument constructed on the 2SLS first stage regression as a regressor ($\dot{\Omega}^i_\tau$). Table 7 reports 2SLS second stage results where typical IV standard errors are adjusted for the fact that the instrument depends on the parameters of the gravity equation (Frankel and Romer, 1999). The leftmost column of this table presents the results of our simplest specification. There, instrumentation strategy seems to be highly efficient: although the partial R^2 statistic is slightly satisfactory since it reaches a value fairly above 6%, the result on the F-test statistic is well above the critical value (Stock and Yogo, 2005)[16]. This first model concludes that a rise of one percentage point in intraregional exports significantly reduces concentration by 0.049 units in the Theil index scale. It is noteworthy to mention that the IV estimate is larger in absolute value than the OLS result[17]. Frankel and Romer

[15] In this context, the critical value for a tolerance level of 10% is just above 11, if the tolerance decays to 5%, this value rises to nearly 20.

[16] It is important to notice that the p-value associated to the Durbin test strongly justifies the IV procedure here employed. Since the null hypothesis is that the variable under consideration can be treated as an exogenous regressor, the highly statistic obtained suggests the rejection of the exogeneity for intraregional exports variable.

[17] This is consistent with the findings of empirical works employing 2SLS strategy (Frankel and Romer, 1999; Feyrer, 2009, and Blanchard and Olney, 2017, among others). As Frankel and Romer (1999) clarifies, this difference means that diversification's partial association with the portion of intraregional exports not correlated with the instrument is lower than its partial association with the portion that is actually related. This occurs because the OLS estimate is defined by the relationship between diversification and intraregional exports, while the IV estimate is defined by the relationship between diversification and the portion of intraregional trade correlated with the instrument.

(1999) suggests that this occurs because the OLS coefficient is biased downwards by omitted variables problem and mainly by measurement error in the trade related factor. Since the instrument shows less dispersion than the endogenous variable, Blanchard and Olney (2017) adds that the instrument identifies a structural source of variation in the pattern of exports to which individuals respond more; larger fluctuations in the noisy, actual export data have less effects over diversification. In this sense, the strategy of 2SLS helps to alleviate this alarm. Although with a lower coefficient in absolute value, these general results hold at the inclusion of the whole bulk of control variables. Columns (2) to (6) demonstrate that $\dot{\Omega}^{i}_{\tau}$ is an exogenous instrument of high quality and intraregional exports are a highly significant determinant of diversification[18]. Regarding our favorite model projected in the rightmost column, it concludes that a rise of one percentage point in intraregional exports significantly enhances diversification by 0.015 units in the Theil index scale. In other words, an increase of one standard deviation in intraregional exports implies an increment in diversification higher than 30% of its standard deviation.

[Insert Table 7 here.]

Regarding control variables, every one of them has the predicted sign and most of them are statically significant. Although human capital seems to contribute positively to diversify exports, its coefficient is not always significant across specifications. Between capital endowment variables, the stock of physical capital seems to be the most robust determinant. In our favorite model, a rise on the logarithm of one billion of constant 2005 international USD per capita on infrastructure, a value close to 1.5 of its standard deviation, reduces concentration by 0.231 units in the Theil index scale. This figure represents nearly a 20% of its standard deviation. Regarding the set of institutions related variables, results suggest that higher levels of political affinity to trade internationally and better institutional quality seem to favor diversification but not always with a statistically significant impact. Our preferred specification concludes that a rise of one standard deviation in the index of freedom to trade implies a significant increment on diversification on nearly 5% of its standard deviation. More

[18] Except for column (5) where the confidence level must go to 10%, values on the F-statistic show that the null hypothesis of weak instrument is rejected at 5% level for every model according to the rule of thumb.

interestingly, we find evidence that a positive shock on the terms of trade significantly damage the diversification process. Similar to what Agosin et al. (2012) concludes, our results suggest that price and income effects seem to induce factor reallocation towards those sectors facing improved export opportunities, increasing concentration. Table 7 also suggest that high volatility on the nominal exchange rate and the presence of an overvaluated local currency contribute to concentrate exports. Results of the rightmost model conclude that a rise of one percentage point on real exchange rate undervaluation reduces concentration by 0.313 units in the Theil index scale. In other words, a decay of one standard deviation in the overvaluation variable augments diversification by 10% of its standard deviation. Finally, every model confirms the great persistence of the dependent variable over time showing highly significant coefficients that range between 0.597 and 0.643. This means that an increase of one unit on the lagged dependent variable implies an increase on the interval of 0.597-0.643 on its actual value.

In sum, these results enhance the importance of active policy as an effective mean in boosting diversification. Among others, higher levels of physical capital investment, lower tariffs, fewer controls on the movements of physical and human capital, and the presence of an undervaluated local currency, encourage the process of diversification. Nevertheless, policy exogenous terms of trade also play a major role but as a concentration booster. But more important, Table 7 results provide broad support for the general prediction of the export experimentation theory: deeper levels of regional integration increase export diversification.

[Insert Table 8 here.]

By focusing only on the overall index, our baseline results could mask heterogeneous effects that integration has on the different margins of diversification. We already concluded that intraregional trade seems to boost diversification, but does it occur on its extensive or intensive margin? As we mentioned above, the extensive margin reflects variations in the number of lines opened, whereas the intensive margin reflects variations in the distribution of export values among open lines. Table 8 explores this question by using the Theil index at extensive and intensive margin as the dependent variable. While columns (1) to (6) present the results obtained on the extensive margin analysis, the intensive margin outcomes are exposed in columns (7) to (12). These results show that the null hypothesis of weak instrument is consistently rejected for both diversification levels: every model presents comfortingly high figures on the F-statistic and on the instrument partial R^2. Equally important, they also suggest that regional integration is a highly significant determinant of extensive and intensive diversification, in other words, higher levels of intraregional trade not only help to rise the number of markets served or products exported, they also help to get a more even distribution of trade values across the open lines. While column (6) concludes that a rise of one standard deviation in intraregional exports reduces extensive margin concentration on nearly 20% of its standard deviation, column (12) affirms that this reduction is nearly 25% on the case of the intensive margin.

Regarding control variables, Table 8 shows that the intensive margin outcomes are pretty close to our baseline conclusions while those of the extensive margin differ substantially. Focusing only on the extensive margin, we can see four main differences comparing its results with our baseline scenario. First, the coefficients related to freedom to trade and to an overvaluated currency maintain its sign but lose all statistical significance. This reveals that neither more freedom to trade, or a more undervaluated currency, rise the number of markets served or products exported in a significant way. Second, although not significant across models, the terms of trade coefficient presents a persistent negative sign, suggesting that improvements in this variable open new exporting lines. Third, exchange rate volatility seems to be significantly adverse to the diversification process. Finally, perhaps the most surprising result, infrastructure shows a slightly significant positive coefficient suggesting that higher level of capital stock per capita curbs the opening of new export lines and even encourages the closure of other lines already opened.

In sum, Table 8 suggest that more freedom to trade internationally and the presence of an undervaluated local currency encourage the process of diversification through its intensive margin. On the other hand, terms of trade and nominal volatility of the exchange rate act as concentration boosters through the intensive and extensive margin respectively. The most notorious result is obtained from the infrastructure variable: while it boosts diversification through the intensive margin, through the extensive line it boosts concentration. This means that more infrastructure development implies fewer open lines but a more even distribution of trade values between them. In the next sub-section, we proceed to run different robustness checks to account for possible sensitivity of these results.

4.3 Sensitivity Analysis

4.3.1 Sample Sensitivity

In our robustness checks, we address six main concerns. First, our outcomes could present some sample sensitivity. Although resulting in an unbalanced panel, our results are obtained maximizing sample size to include several countries form every region of the world. In Table 9 we report second stage results of the 2SLS estimation of the main equation (3), restricting the sample only to those countries that present bilateral trade data every year over the 1960-2015 period. The results of this strongly balanced panel are reassuringly consistent with those found previously: the instrument is strong and the intraregional exports share is a highly significant determinant of diversification. Perhaps not surprisingly, its coefficient rises in absolute value. However, reducing the sample has a small impact on the outcomes of institutions related variables: although both maintain their negative sign, they lose statistical significance.

4.3.2 Alternative Measure of Diversification

Second, we replicate our previous estimations but using an alternative measure of diversification: the Herfindahl-Hirschman index. This indicator is computed as:

$$diversification_t^{hh} = \frac{\sum_{i \in G} \left(\chi_t^i / \sum_{i \in G} \chi_t^i \right)^2 - 1/n_t}{1 - 1/n_t} \quad (7)$$

Where χ_t^i denotes the export value of product i during year t, n_t represents the number of total products exported, and G denotes the whole basket of exportable products. Clearly, this is an intensive margin index since it measures the dispersion of trade value across the whole set of exported products, and similar to Theil index, higher values denote higher concentration levels. Here, a country with a perfectly diversified basket shows a value close to zero, whereas a country that exports only one product shows a value close to the unit.

The outcomes of these estimations are summarized in Table 10. Not surprisingly, these results are essentially equivalent to those obtained using the intensive margin Theil index as the dependent variable. Again, intraregional exports are a highly significant determinant of diversification and the null hypothesis of a weak instrument is consistently rejected. Column (6) concludes that a rise of one standard deviation in intraregional exports reduces concentration on nearly 25% of its standard deviation. This value is highly similar to the one found under the Theil index at intensive margin analysis. In opposition to our main results, human capital accumulation seems to be the most robust determinant between capital endowment variables and shocks on the net barter terms of trade do not have significant impact on diversification.

4.3.3 Alternative Instruments

In order to check whether estimates are sensitive to the source of exogenous variation selected, we proceed to construct two alternative instruments. The first alternative strategy simply disaggregates the cumulative index of "large" natural disasters affecting the importer country ($natural_t^j$) into the different types of "large" disasters considered in its construction: droughts, wildfires, earthquakes, mass movements, volcanic activities, and storms. The definition of "large" and the presumption that the selected disasters are orthogonal to any economic factor are maintained. Based on Blanchard and Olney (2017), our second alternative instrument substitutes the importer natural disasters variable ($natural_t^j$) for the crude death rate in the importing country ($death_t^j$). Our key identifying assumption is that neither foreign disasters or foreign death rate have effects over local diversification besides its bilateral trade influence.

Table 11 report the results of estimating the gravity equation using both alternative strategies[19]. Broadly speaking, these results are highly consistent with the ones obtained using our main instrument: bilateral flows are positively related to the product of the mass of both trading partners and to their geographical proximity. In both cases, over 86% of the bilateral flows variation is explained by the set of independent variables. More important, Table 12 presents the results of the 2SLS second stage estimation using adjusted IV standard errors in both models. These findings reveal that employing these alternative instruments does not produce important differences either on the value or significance of the coefficients. Deeper grades of regional integration, higher levels of capital investment, lower tariffs, and the presence of an undervaluated currency, still encourage the diversification process. Parallel, positive shocks on terms of trade still play a detrimental role as a concentration booster. Running some F-tests on excluded instruments proves that the null hypothesis of weak instrument is always rejected with the p-value information[20]. Lastly, our favorite specification concludes that a rise of one percentage point in intraregional exports significantly reduces concentration by 0.014 or 0.011 units in the Theil index scale. Both coefficients prove to be significantly different from the one found using our main instrumental strategy.

4.3.4 Endogeneity Concerns

To limit further concerns regarding endogeneity problems, we performed some regressions with the whole set of independent variables lagged one period. This procedure alleviates threats of reversal causality in other variables beside intraregional exports, and also considers more properly the time that any economic variable needs to affect diversification. Table 13 summarizes these results. Once again, every model presents comfortingly high figures on the F-statistic and the instrument partial R^2: $\dot{\Omega}^i_\tau$ is a high quality instrument and intraregional exports significantly boost diversification. Similar to the results of the sample sensitivity robustness check, the freedom to trade coefficient maintains its sign but loses some statistical

[19] While the specification considering disaggregated natural disasters as source of variation is identified as the Felbermayr-Gröschl II model, the strategy focusing on the death rate is denoted as the Blanchard-Olney model.

[20] Regarding the rule of thumb critical values, the null hypothesis of weak instrument is not always rejected in the Blanchard-Olney model. Perhaps, this is not surprising since mortality rate could be easily linked to economic variables highly correlated with trade flows, leaving open the possibility that crude death rate could violate the exclusion restriction.

significance, and in opposition to our main results, (lagged) exchange rate volatility seems to be significantly damaging to the diversification process, while (lagged) shocks on the terms of trade lose statistical significance.

4.3.5 Alternative Frequency

As a fifth robustness check, Table 14 presents the results obtained using the dataset at annual frequency. As we warned in section 3, the outcomes from this exercise are prone to be biased by the presence of outliers and business-cycle influence. The results are highly similar to our baseline scenario, only infrastructure variable presenting major changes: the stock of capital per capita loses significance, showing no discernible impact on diversification. However, and not so surprisingly, diversification presents an elevated persistence over time: with a highly significant coefficient, an increase of one unit on the lagged dependent variable implies an increase on the interval of 0.832-0.843 on its actual value. Finally, our favorite model concludes that a growth of one percentage point in intraregional exports significantly reduces concentration by 0.006 units in the Theil index scale. This coefficient does not prove to be significantly different from the one found using our five years frequency dataset.

4.3.6 Diversification Cone Path

Finally, an stylized fact evidenced by some empirical literature affirms the existence of a non-monotone path on the diversification process: exports diversify and then re-concentrate with economic development (Imbs and Wacziarg, 2003; Kingler and Lederman, 2006, and Cadot et al., 2011)[21]. In particular, Cadot et al. (2011) sustains that as countries travel across this hump-shaped path, they fail to close some export lines that no longer belong to their comparative advantages, inflating diversification measures. When comparative advantages catch up, re-concentration occurs. To control for possible income effects, Table 15 presents the results obtained from the inclusion of the income per capita variable in its simple and quadratic form (gdp^i_τ and $(gdp^i_\tau)^2$, respectively). Since these checks probably suffer from re-

[21] Although a clear positive relationship between diversification and development is well sustained by the literature, the process of re-specialization is sometimes rejected (De Benedictis et al., 2009, and Mau, 2016).

versal causality problems, we are careful to not draw causal inference from them. The results validate the use of the quadratic form to approximate the relationship between concentration and income: as we improve income level, there appears to be a convex figure initially decreasing and then increasing in concentration. Regarding other regressors, the inclusion of the income variable reduces the absolute value of every coefficient without producing major changes either on its sign or significance. Again, intraregional export hold as a significant determinant of diversification and $\dot{\Omega}^i_\tau$ proves to be a strong instrument.

4.4 Heterogeneous Effects

In order to illustrate the mechanism through which intraregional exports push a more diversified exporting structure, we analyze the potential heterogeneous effects that certain types of products have on different diversification margins. As we mentioned above, Albornoz et al. (2012) maintains that firms are more likely to serve near destinations to their current active markets before going out of its neighborhood: once the firm realizes its export potential from its experience in neighbor's markets, it moves on to the distant ones. Expecting that this learning by exporting process may be encouraged by regional integration, we anticipate that higher levels of intraregional exports lead to a more diversified exporting matrix. Although the results shown in the previous section confirm our prediction, new questions arise regarding the mechanism through which this process is materialized. Are intraregional exports of certain types of products more efficient in boosting diversification? Do these products operate through a particular margin of diversification? Since we presume that knowledge intensive products are highly benefited by the learning by exporting process, we hypothesize that its intraregional flows are highly efficient in promoting diversification and that this promotion occurs mainly through the opening of new exporting lines.

[Insert Table 16 here.]

To address this issue in empirical grounds, we employ two common product categorization to construct proxies for intraregional trade of knowledge intensive products. Our first taxonomy is based on Rauch (1996), where an organization that defines homogeneous and non-homogeneous goods based on the possession of a well-known reference price is suggested. According to Rauch's criteria, homogeneous products are those traded on pretty standardized exchanges whereas non-homogeneous products are the remaining ones[22]. Table 16 presents the results of using intraregional exports of non-homogeneous products ($\Omega_{\tau}^{nh,i}$) as a regressor: while columns (1) to (6) present the outcomes obtained on the extensive margin, the intensive margin outcomes are exposed in columns (7) to (12). Before analyzing these results, it's important to make two relevant points about these estimations. First, since we presume heterogeneous goods as more knowledge intensive than homogeneous goods, we consider intraregional exports of non-homogeneous products as a broad proxy of intraregional exports of knowledge intensive products. Second, it is important to notice that these estimations require the repetition of the whole IV strategy including the re-estimation of the gravity equation (4) but using bilateral flows of non-homogeneous products only. That said, the results presented on Table 16 reveal that intraregional trade of non-homogeneous products is highly efficient in promoting diversification mainly through the extensive margin. Although column (6) concludes that a rise of one standard deviation in intraregional exports of non-homogeneous products reduces extensive margin concentration on nearly 20% of its standard deviation, and column (12) affirms that this reduction is nearly 25% on the case of the intensive margin, extensive margin results are highly consistent across models while intensive margin outcomes are not. Intraregional exports of non-homogeneous products significantly reduce extensive margin concentration in every model, whereas in the intensive margin this does not occur. Meanwhile, the null hypothesis of weak instrument is consistently rejected for both diversification levels, and the outcomes of the control variables are practically equal to those found in the previous section.

[Insert Table 17 here.]

Regarding our second product categorization, we employ the taxonomy proposed by the

[22] Details about the 4-digit SITC rev. 2 codes that classifies homogeneous and non-homogeneous products according to Rauch's criteria can be found in http://econweb.ucsd.edu/~jrauch/rauch_classification.html. Here we present the results using the more conservative classification, employing the alternative classification produce results that are essentially equal to the ones shown here.

World Trade Organization (WTO) for primary and non-primary goods[23]. Similar to the previous exercise, we presume intraregional exports of non-primary products as a broad proxy of intraregional exports of knowledge intensive products. Table 17 presents the results of including intraregional exports of non-primary products ($\Omega_{\tau}^{np,i}$) as a regressor, it reveals that this variable seems to be efficient promoting diversification only through the extensive margin: our favorite model concludes that a rise of one standard deviation in intraregional exports of non-primary products significantly reduces extensive margin concentration on more than 10% of its standard deviation, whereas this reduction is slightly above 8% in the case of the intensive margin and has no statistical significance. Similar to the previous exercise, the null hypothesis of weak instrument is always rejected for both diversification levels, and the results of the control variables are not sensitive.

[23] This categorization defines as primary those products classified in sections 0, 1, 2, 4 excepting divisions 27 and 28, 5, 6, 7, and 8 excepting divisions 68 and 69, of the SITC rev. 2 codification. Clearly, those products classified with the remaining codes are defined as non-primary products

5 Conclusion

The main goal of this paper is to determine whether export experimentation theory is motivated by regional integration. As Albornoz et al. (2012) maintains, firms are more inclined to serve destinations closer to their current active markets before expanding on to the distant ones. Thus, nearby markets serve as testing grounds for product experimentation since firms realize its export potential from the experience in neighbor's markets before moving forward to remote ones. Expecting that this learning by exporting process may be encouraged by regional integration, we hypothesized that higher intraregional exports lead to a more diversified exporting structure.

To test this hypothesis properly, we take into account the reversal causality existing between both main variables. In particular, we employ the panel IV setup proposed by Felbermayr and Groschl (2013) to alleviate the endogeneity concern regarding intraregional exports. Broadly speaking, this proposal declares that natural disasters provide an exogenous variation in bilateral trade flows, allowing the construction of an exogenous time-varying instrument for intraregional trade. We find evidence supporting the idea that a higher degree of intraregional exports diversifies the exporting matrix at its extensive and intensive margins: more integration not only boosts the number of markets served or products exported, it also helps to obtain a more even distribution of trade values across the open lines. This result holds through different specifications and is robust to a battery of sensitivity analyses. Regarding the control variables, higher levels of physical capital investment, lower tariffs, fewer controls on the movements of physical and human capital and the presence of an undervaluated local currency, seems to encourage the diversification process. On the other hand, improvements in the terms of trade play a major role as a concentration booster.

We also show that intraregional exports of knowledge intensive products are highly efficient in terms of promoting diversification, and that this promotion occurs mainly through the extensive margin. This finding of heterogeneous effects is useful in understanding the mechanism through which intraregional exports push a more diversified exporting structure, since it is highly consistent with export experimentation theory. If we presume that knowledge intensive products are deeply benefited by the learning by exporting process, this theory hypothesizes that intraregional trade of these type of products promotes diversification through the opening of new exporting lines.

Finally, our findings bear significant policy considerations. While the benefits of intraregional trade are often accentuated in the literature, we shed some light about a gain on diversification that rarely appears, and we examine what types of exporting products are more beneficial for this process. In this sense, our results highlight the role of an active policy: if diversification is something desirable, an agenda pushing more regional integration is needed.

References

Daron Acemoglu and Fabrizio Zilibotti. Was prometheus unbound by chance? risk, diversification, and growth. *The Jounal of Political Economy*, 105:709–751, 1997.

Manuel Agosin. Export diversification and growth in emerging economies. *Universidad de Chile, Departamento de Economía Working Paper*, 233, 2007.

Manuel R. Agosin, Roberto Alvarez, and Claudio Bravo-Ortega. Determinants of export diversification arounf the world: 1962-2000. *The World Economy*, 35 (3):243–395, 2012.

Facundo Albornoz, Hector F. Calvo Pardo, Gregory Corcos, and Emanuel Ornelas. Sequential exporting. *Journal of International Economics*, 88:17–31, 2012.

James E. Anderson and Eric van Wincoop. Gravity with gravitas: A solution to the border puzzle. *The American Economic Review*, 93 (1):170–192, 2003.

Scott L. Baier and Jeffrey H. Bergstrand. Bonus vetus ols: A simple method for approximating international trade-cost effects using the gravity equation. *Journal of International Economics*, 77 (1):77–85, 2009.

Rchard Baldwin and Daria Taglioni. Positive oca criteria: Microfundations for the rose effect. *Mimeo*, 2004.

Robert J. Barro and Jong Wha Lee. A new data set of educational attainment in the world: 1950-2010. *Journal of Development Economics*, 104:184–198, 2013.

Andrew B. Bernard, Stephen J. Redding, and Peter K. Schott. Comparative advantages and heterogeneous firms. *The Review of Economic Studies*, 74:31–66, 2007.

Emily J. Blanchard and William W. Olney. Globalization and human capital investment: Export composition drives educational attainment. *Journal of Internarional Economics*, 106:165–183, 2017.

Olivier Cadot, Celine Carrer, and Vanessa Strauss-Kahn. Export diversification: What is behind the hump? *The Review of Economics and Statistics*, 93:590–605, 2011.

W. Max Corden and J. Peter Neary. Booming sector and de-industrialisation in a small open economy. *The Economic Journal*, 92 (386):825–848, 1982.

John G. Cragg and Stephen G. Donald. Testing identifiability and specification in instrumental variable models. *Econometric Theory*, 9:222–240, 1993.

Zsolt Darvas. Real effective exchange rates for 178 countries: A new database. *Bruegel Working Paper*, 2012/06, 2012.

Luca De Benedictis, Marco Gallegati, and Massimo Tamberi. Overarll trade specialization and economic development: Countries diversify. *The Review of World Economics*, 145 (1): 37–55, 2009.

Julian Di Giovanni and Andrei A. Levchenko. The risk content of exports: A portfolio view of international trade. *NBER Working Paper*, 16005, 2010.

Rudiger Dornbusch, Stanley Fischer, and Paul A. Samuelson. Comparative advantage, trade, and payments in a ricardian model with a continnum of goods. *The American Economic Review*, 67:823–839, 1977.

Hartmut Egger, Peter Egger, and David Greenaway. The trade structure effects of endogenous regional trade agreements. *Journal of International Economics*, 74 (2):278–298, 2008.

Peter Egger, Mario Larch, Kevin E. Staub, and Rainer Winkelmann. The trade effects of endogenous preferential trade agreements. *The American Economic Journal: Economic Policy*, 3 (3):113–143, 2011.

Robert C. Feenstra, Robert E. Lipsey, Haiyan Deng, Alyson C. Ma, and Hengyong Mo. World trade flows: 1962-2000. *NBER Working Paper*, 11040, 2005.

Gabriel Felbermayr and Jasmin Groschl. Natural disaster and the effect of trade on income: A new panel iv approach. *European Economic Review*, 58:18–30, 2013.

James Feyrer. Trade and income: Exploiting time series in geography. *NBER Working Paper*, 14910, 2009.

Jeffrey A. Frankel and David Romer. Does trade cause growth? *The American Economic Review*, 89(3):379–399, 1999.

James Gwartney, Robert Lawson, and Joshua Hall. Economic freedom of the world: 2016 annual report. 2016.

Ricardo Hausmann and Roberto Rigobon. An alternative interpretation of the resource curse: Theory and policy implications. *NBER Working Paper Series*, Working Paper 9424, 2003.

Heiko Hesse. Trade structure and growth. *World Bank Comission on Growth and Developmetn Working Paper*, 21, 2008.

Jean Imbs and Romain Wacziarg. Stages of diversification. *The American Economic Review*, 93:63–86, 2003.

Bailey Kingler and Daniel Lederman. Discovery, innovation, and immitation inside the global technology frontier. *World Bank Policy Research Working Paper*, 3872, 2006.

Miklós Koren and Silvana Tenreyro. Technological diversification. *The American Economic Review*, 103 (1), 2013.

Daniel Lederman and William F. Maloney. Trade structure and growth. *Natural resources: Neither curse nor destiny, D. Lederman, W.F. Maloney (eds.)*, Stanford University, 2007.

Andrei A. Levchenko. Institutional quality and international trade. *The Review of Economic Studies*, 74 (3):791–819, 2007.

Karsten Mau. Export diversification and income differences reconsidered: The extensive product margin in theory and application. *The Review of World Economics*, 152 (2):351–381, 2016.

Thierry Mayer and Soledad Zignago. Notes on ceppi's distance measures: The geodist database. *CEPII Working Paper*, 24, 2011.

Marc J. Melitz. The impact of trade on intra-industry reallocation and aggregate industry productivity. *Econometrica*, 71 (6):1695–1725, 2003.

Marc J. Melitz and Gianmarco I. P. Ottaviano. Market size, trade, and productivy. *Review of Economic Studies*, 75:295–316, 2008.

Rodrigo Miranda and Rodrigo Wagner. Bad neighbors: Bordering institutions matter for comparative (dis)advantage. *Mimeo*, 2015.

Nathan Nunn. Relationship-specificity incomplete contracts and the pattern of trade. *The Quarterly Journal of Economics*, 122 (2):569–600, 2007.

Aleksandro Parteka and Massimo Tamberi. Determinants of export diversification: An empirical investigation. *Universita Politecnica delle Marche, Dipartimento di Economia*, 327, 2012.

James E. Rauch. Networks versus markets in international trade. *NBER Working Paper*, 5617, 1996.

Dani Rodrik. The real exchange rate and economic growth. *Brooking Papers on Economic Activity*, 39 (2):365–439, 2008.

Andrew K. Rose and Mark M. Spiegel. International financial remoteness and macroeconomic volatility. *Journal of Development Economics*, 89:250–257, 2009.

Joao M. C. Santos Silva and Silvana Tenreyro. The log of gravity. *The Review of Economics and Statistics*, 88 (4):641–658, 2006.

James H. Stock and Motohiro Yogo. Testing for weak instruments in linear iv regression. *Andrews DWK Identification and Inference for Econometric Models. New York: Cambridge University Press*, pages 80–108, 2005.

Appendix A

Figure 1: Data distribution of the bilateral exports in the main sample (1960-2015).

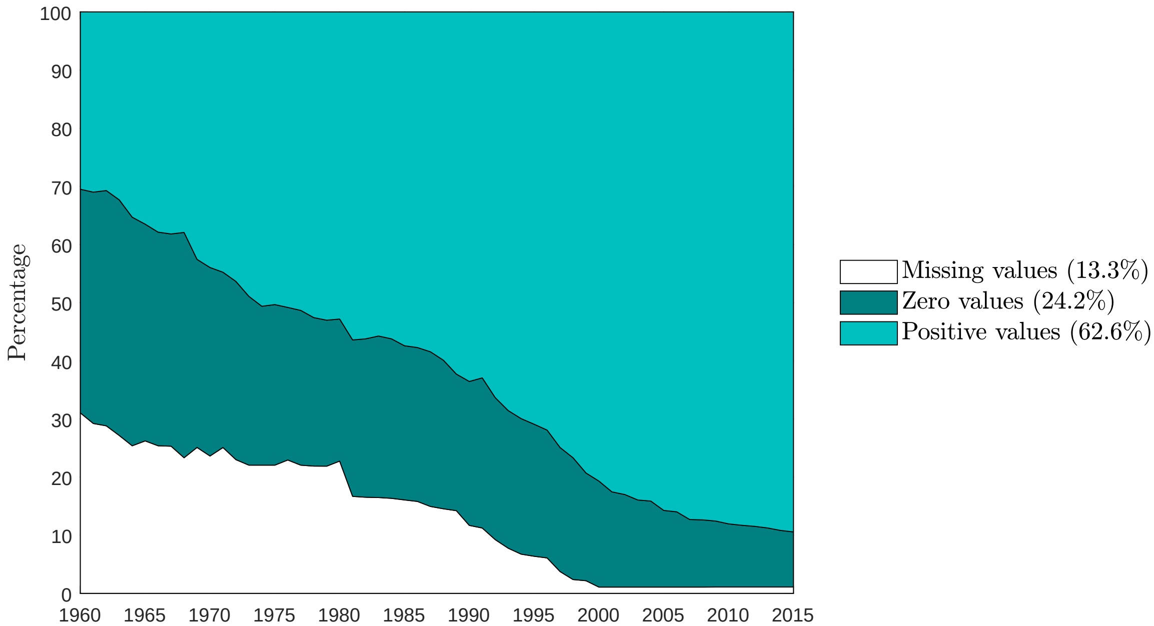


Figure 2: Intraregional exports by region over time (1960-2015).

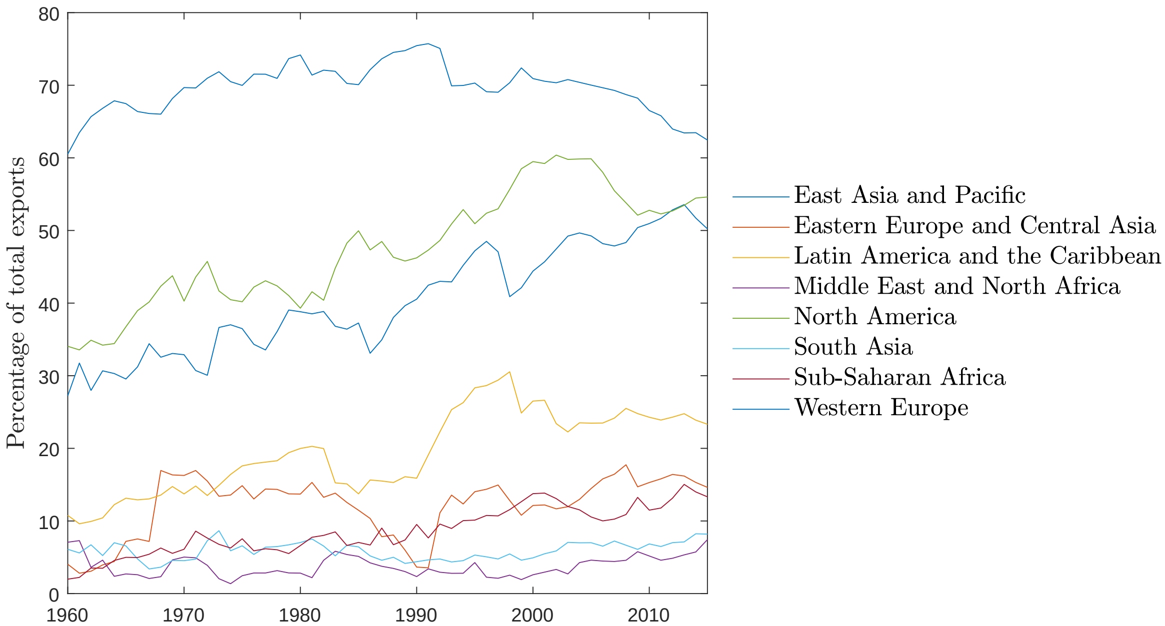
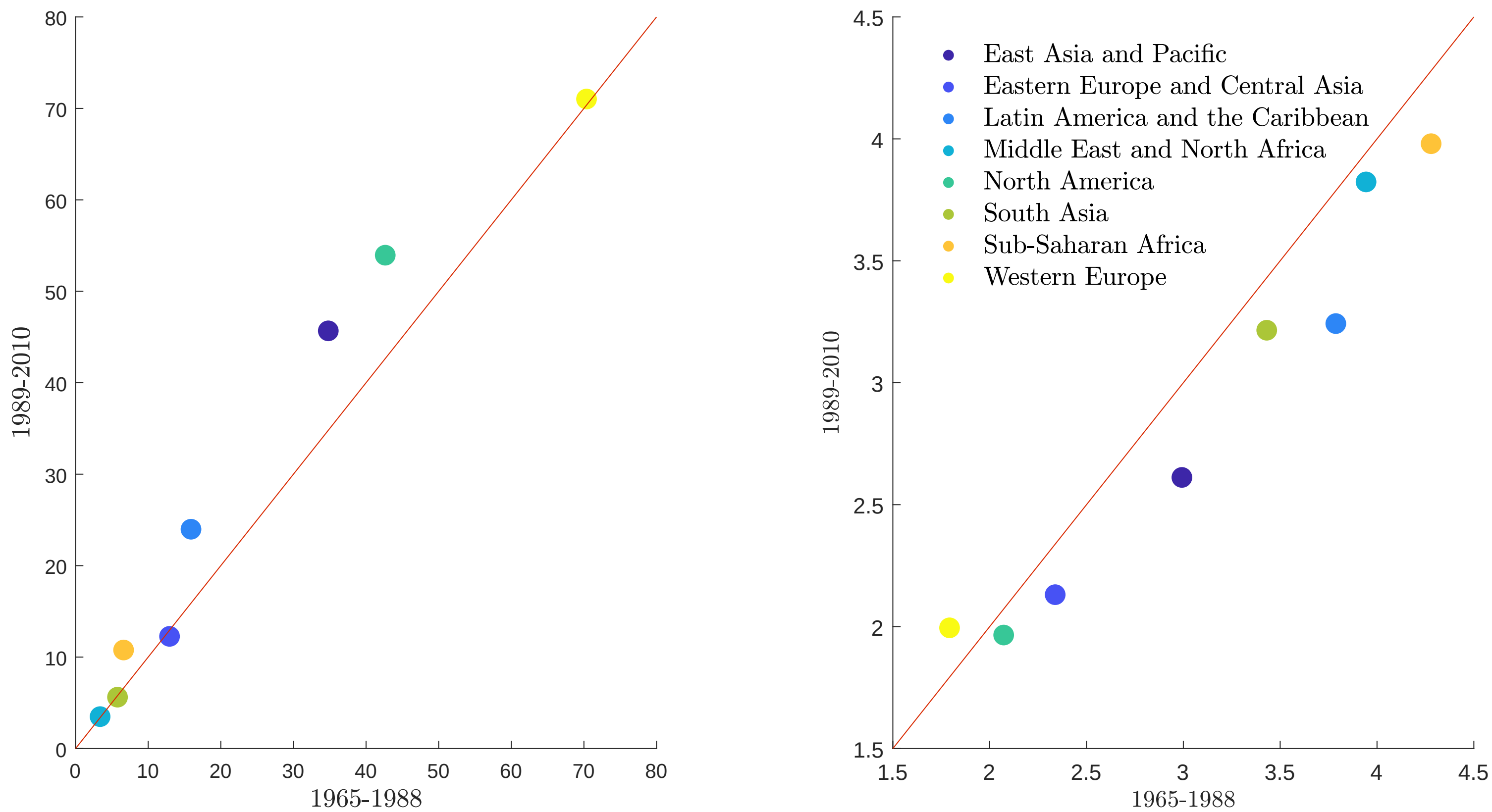


Figure 3: Intraregional trade and export diversification over time at regional level.



(a) Intraregional trade

(b) Export diversification

Figure 4: Intraregional trade over time and ordered by region.

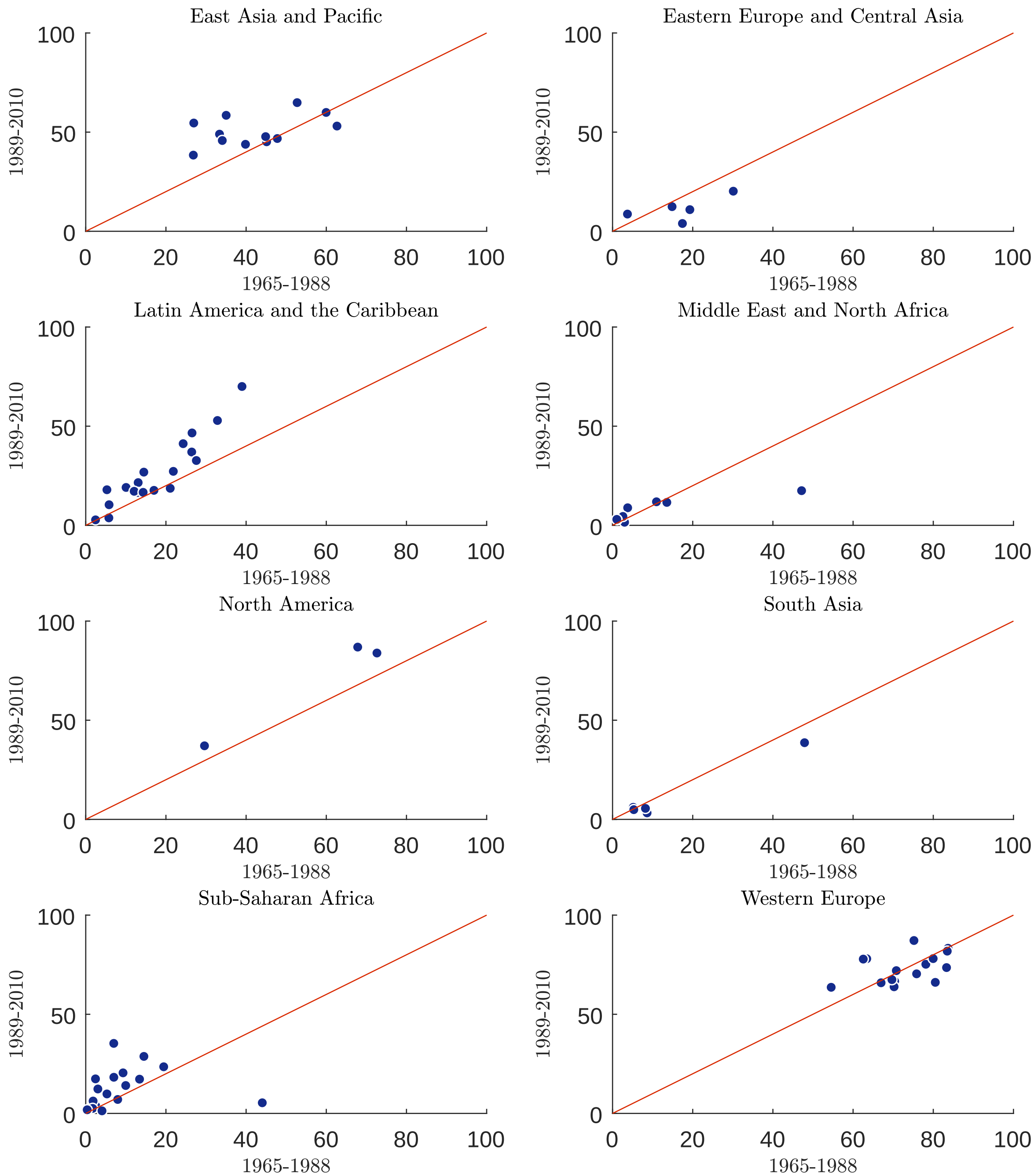


Figure 5: Export diversification over time and ordered by region.

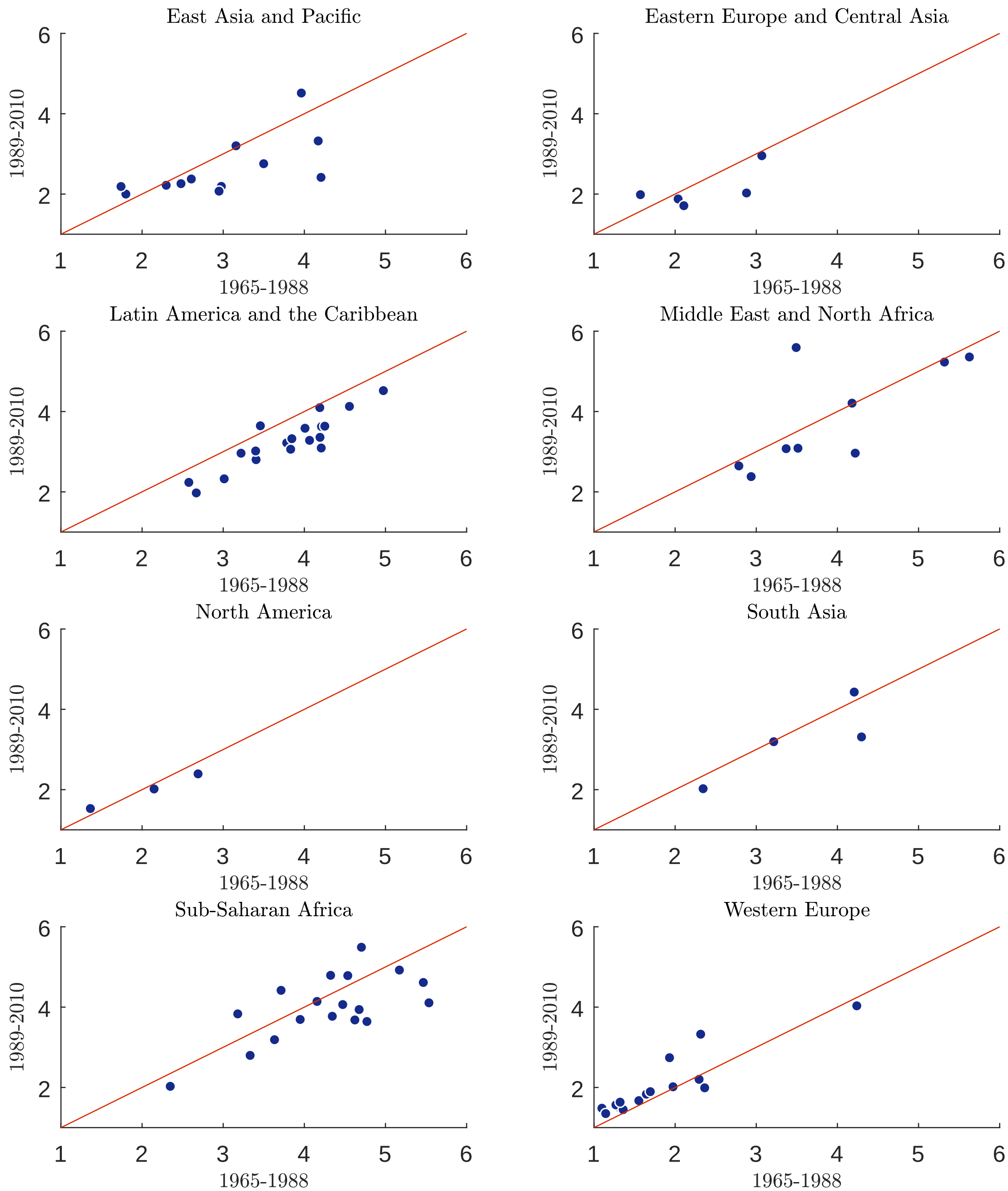
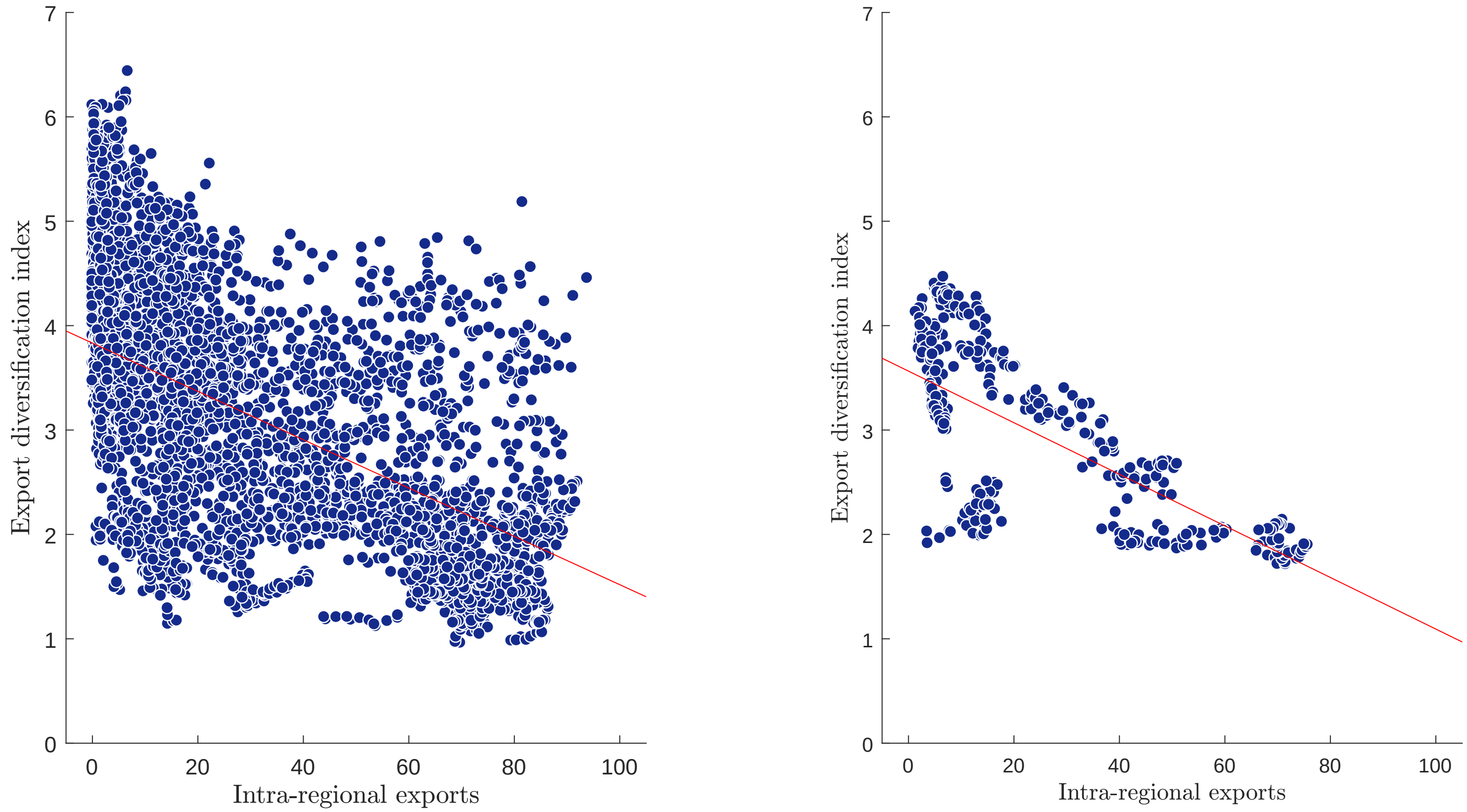


Figure 6: Relationship between intraregional trade and diversification.



(a) Country level

(b) Regional level

Figure 7: Correlation between intraregional trade and diversification at country level.

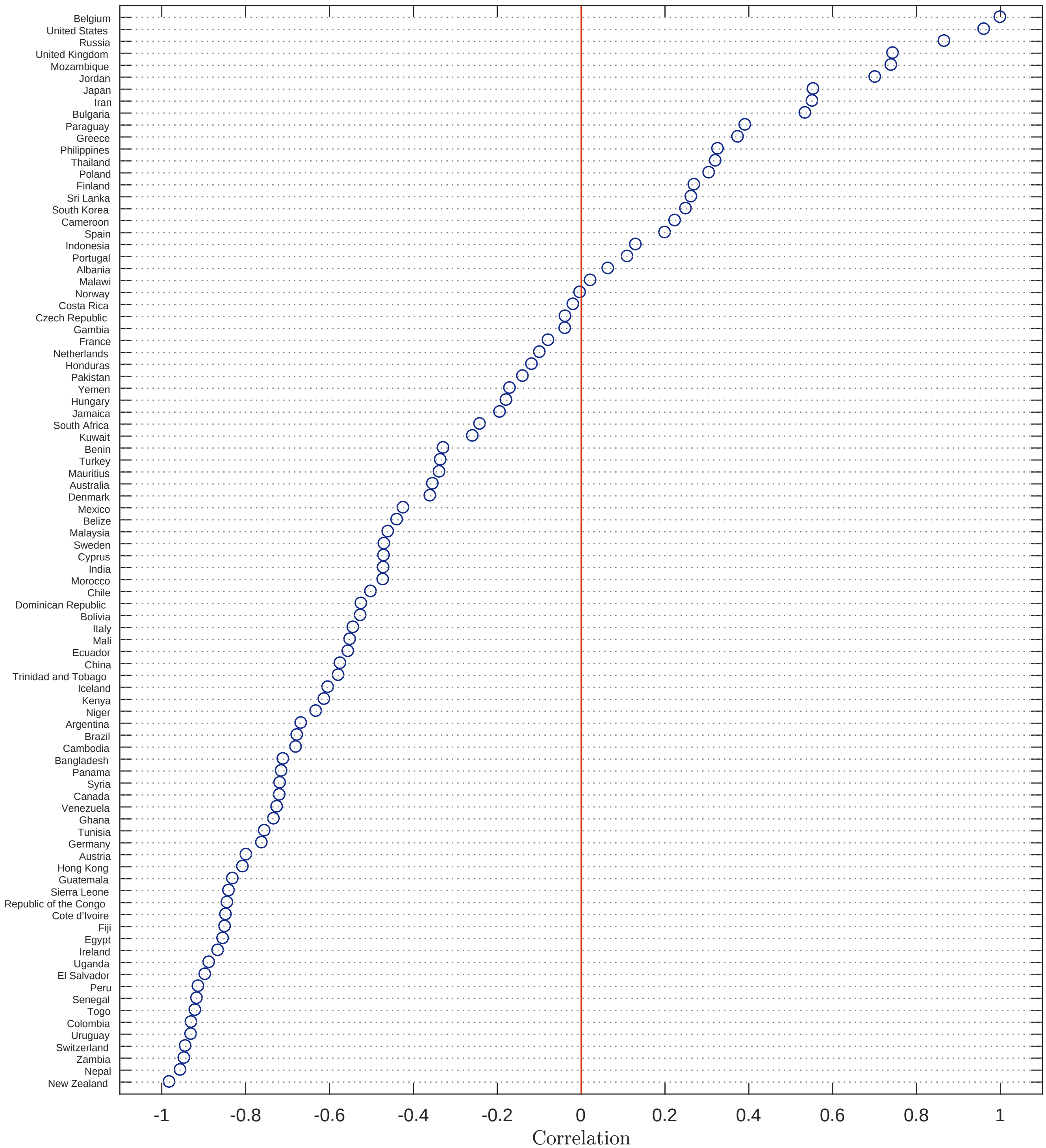


Figure 8: Correlation between intraregional trade and diversification at country level and ordered by region.

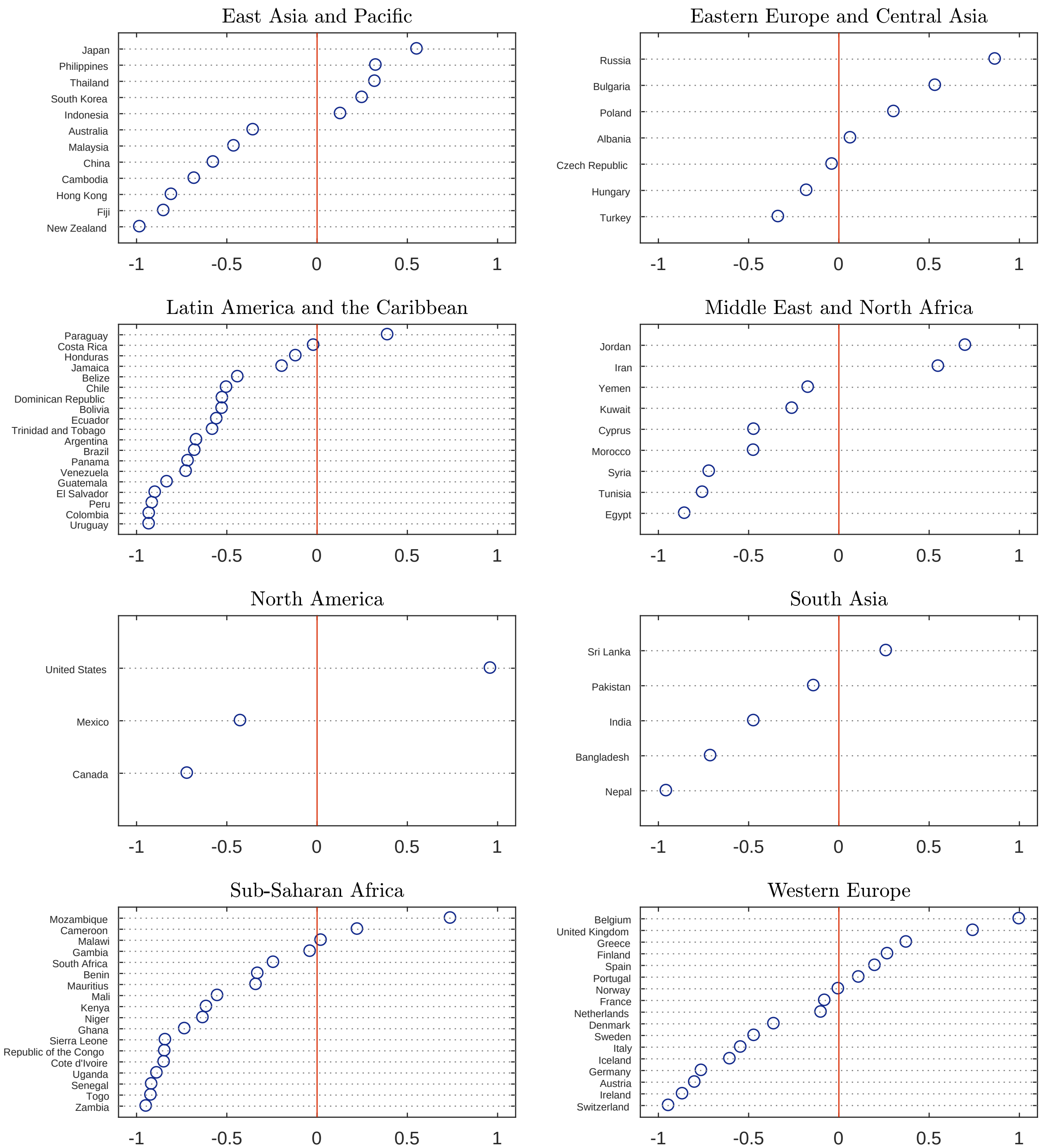


Table 1: List of countries included in the main sample and classified by region[a].

North America (NAM)		South Asia (SAR)		Eastern Europe and Central Asia (ECA)		Middle East and North Africa (MENA)	
Canada	●	Bangladesh	○	Albania	○	Cyprus	●
Mexico	●	India	●	Bulgaria	○	Egypt	●
United States	●	Nepal	○	Czech Republic	○	Iran	○
		Pakistan	●	Hungary	●	Jordan	●
		Sri Lanka	●	Poland	●	Kuwait	○
				Russia	○	Morocco	●
				Turkey	●	Syria	○
						Tunisia	●
						Yemen	○

East Asia and Pacific (EAP)		Western Europe (WEU)		Sub-Saharan Africa (SSA)		Latin America and the Caribbean (LAC)	
Australia	●	Austria	●	Benin	○	Argentina	●
Cambodia	○	Belgium	○	Cameroon	●	Belize	○
China	○	Denmark	●	Cote d'Ivoire	●	Bolivia	○
Fiji	○	Finland	●	Gambia	○	Brazil	●
Hong Kong	●	France	●	Ghana	●	Chile	●
Indonesia	●	Germany	●	Kenya	●	Colombia	●
Japan	●	Greece	●	Malawi	○	Costa Rica	●
Malaysia	●	Iceland	●	Mali	○	Dominican Republic	○
New Zealand	●	Ireland	●	Mauritius	●	Ecuador	○
Philippines	●	Italy	●	Mozambique	○	El Salvador	○
South Korea	●	Netherlands	●	Niger	○	Guatemala	●
Thailand	●	Norway	●	Republic of the Congo	○	Honduras	●
		Portugal	●	Senegal	●	Jamaica	●
		Spain	●	Sierra Leone	○	Panama	○
		Sweden	●	South Africa	○	Paraguay	○
		Switzerland	●	Togo	○	Peru	●
		United Kingdom	●	Uganda	○	Trinidad and Tobago	●
				Zambia	●	Uruguay	●
						Venezuela	●

[a] Full filled circles denote countries included also in the smaller, strongly balanced sample.

Table 2: Descriptive statistics of the gravity equation variables.

Variable	Obs.	Mean	Std. Dev.	Min.	Max.
Bilateral exports ($\chi_t^{i,j}$)	389,073	5.4e+08	5.4e+09	0.00	4.1e+11
Contiguity status ($adjacency^{i,j}$)	448,560	0.02	0.15	0.00	1.00
Crude death rate ($death_t^j$)	448,560	10.59	5.41	2.52	54.44
Droughts ($drought_t^j$)	448,560	0.04	0.19	0.00	1.00
Dry mass movements ($mass_t^j$)	448,560	0.00	0.01	0.00	1.00
Earthquakes ($earthquake_t^j$)	448,560	0.03	0.17	0.00	1.00
Financial remoteness ($financial_t^j$)	448,560	7.35	1.56	0.00	8.71
Floods ($flood_t^j$)	448,560	0.10	0.30	0.00	1.00
Geographical distance ($\ln distance^{i,j}$)	448,560	8.76	0.79	4.95	9.89
Landslides ($landslide_t^j$)	448,560	0.03	0.05	0.00	1.00
Language ($language^{i,j}$)	448,560	0.14	0.35	0.00	1.00
Natural disasters ($natural_t^j$)	448,560	0.22	0.56	0.00	5.00
Population size ($population_t^j$)	448,293	16.28	1.59	11.43	21.04
Storms ($storm_t^j$)	448,560	0.05	0.21	0.00	1.00
Surface area ($area_t^j$)	448,560	12.35	1.84	7.00	16.65
Volcanic activities ($volcanic_t^j$)	448,560	0.02	0.05	0.00	1.00
Wildfires ($wildifre_t^j$)	448,560	0.01	0.04	0.00	1.00

Table 3: Descriptive statistics of the main regression variables.

Variable	Obs.	Mean	Std. Dev.	Min.	Max.
Capital stock per capita ($infrastructure^i_\tau$)	870	0.65	1.32	-6.67	11.74
Exchange rate overvaluation ($overvaluation^i_\tau$)	853	-0.02	0.35	-2.93	2.35
Export diversification, extensive ($diversification^{x,i}_\tau$)	881	0.52	0.54	-0.04	3.21
Export diversification, intensive ($diversification^{i,i}_\tau$)	881	2.63	0.92	0.96	5.72
Export diversification, overall ($diversification^{o,i}_\tau$)	886	3.14	1.18	0.97	6.10
Freedom to trade internationally ($free\ to\ trade^i_\tau$)	728	6.14	2.20	0.00	9.97
Human capital ($education^i_\tau$)	900	5.79	3.31	0.00	13.42
Intra-regional exports (Ω^i_τ)	852	31.89	27.66	0.00	91.56
Intra-regional exports, Felbermayr-Gröschl I ($\dot{\Omega}^i_\tau$)	900	27.28	24.76	1.49	83.17
Intra-regional exports, Felbermayr-Gröschl II ($\dot{\Omega}^i_\tau$)	900	27.29	24.73	1.52	83.16
Intra-regional exports, Blanchard-Olney ($\dot{\Omega}^i_\tau$)	900	27.55	24.58	2.01	83.29
Legal system and property rights ($institutions^i_\tau$)	673	5.62	1.90	1.15	9.63
Net barter terms of trade ($terms\ of\ trade^i_\tau$)	890	4.62	0.11	4.11	4.87
Nominal exchange rate volatility ($volatility^i_\tau$)	760	1.2e-02	3.2e-02	3.7e-05	5.1e-01
Real gross domestic product per capita (gdp^i_τ)	828	8.35	1.53	4.93	11.39

Table 4: Impact of endogenous intraregional exports on overall Theil index (1965-2010), OLS[a].

Dependent variable: $diversification_{\tau}^{o,i}$	(1)	(2)	(3)	(4)	(5)	(6)
Ω_{τ}^{i}	-0.013***	-0.007***	-0.010***	-0.007***	-0.004***	-0.005***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
$diversification_{\tau-1}^{o,i}$		0.676***		0.672***	0.665***	0.633***
		(0.041)		(0.042)	(0.041)	(0.039)
$education_{\tau}^{i}$			-0.013	-0.029	-0.002	-0.006
			(0.025)	(0.018)	(0.019)	(0.018)
$infrastructure_{\tau}^{i}$			-0.028	0.006	-0.048	-0.129
			(0.137)	(0.054)	(0.105)	(0.100)
$free\ to\ trade_{\tau}^{i}$			-0.092***		-0.036***	-0.035***
			(0.014)		(0.011)	(0.011)
$institutions_{\tau}^{i}$			-0.009		-0.020	-0.014
			(0.018)		(0.014)	(0.014)
$terms\ of\ trade_{\tau}^{i}$			1.019***		0.719***	0.597***
			(0.315)		(0.198)	(0.201)
$volatility_{\tau}^{i}$			-0.030		0.054	0.102
			(0.400)		(0.306)	(0.340)
$overvaluation_{\tau}^{i}$						0.277***
						(0.058)
Fixed-effects						
- country	Yes	Yes	Yes	Yes	Yes	Yes
- year	Yes	Yes	Yes	Yes	Yes	Yes
Countries	90	90	90	90	90	90
Observations	850	770	596	743	596	596
Adjusted R^2	0.877	0.944	0.907	0.944	0.952	0.955
F-stat			12.26	5.84	4.59	6.22
P-value of F-stat			0.000	0.001	0.000	0.000

[a] Robust standard errors are reported in parenthesis. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Fixed-effects are included but not reported. Averages of five years non-overlapping periods are computed to the whole set of variables.

Table 5: Felbermayr-Gröschl type gravity equation (1960-2015), PPML[a].

Dependent variable: $\chi_t^{i,j}$	Felbermayr-Gröschl I
$natural_t^j$	-1.376***
	(0.148)
$natural_t^j \times \ financial_t^j$	0.006
	(0.004)
$natural_t^j \times \ area^j$	-0.030***
	(0.011)
$natural_t^j \times \ population_t^j$	0.094***
	(0.013)
$natural_t^j \times \ adjacency^{i,j}$	0.055
	(0.043)
$population_t^i$	0.881***
	(0.235)
$population_t^j$	0.542***
	(0.190)
$distance^{i,j}$	-0.825***
	(0.033)
$adjacency^{i,j}$	0.384***
	(0.084)
$language^{i,j}$	0.192***
	(0.073)
Fixed-effects	
- importer and exporter	Yes
- year	Yes
Countries	90
Observations	388,568
R^2	0.867

[a] Robust standard errors clustered at country-pair level are reported in parenthesis. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Fixed-effects are included but not reported.

Table 6: Relationship between actual and constructed intraregional trade (1965-2010), OLS[a].

Dependent variable: Ω^i_τ	Felbermayr-Gröschl I	
	(1)	(2)
$\dot{\Omega}^i_\tau$	1.725***	0.910***
	(0.274)	(0.186)
$\Omega^i_{\tau-1}$		0.605***
		(0.043)
Fixed-effects		
- country	Yes	Yes
- year	Yes	Yes
Controls	Yes	Yes
Countries	90	90
Observations	598	595
Adjusted R^2	0.947	0.969
Instrument partial R^2	0.098	0.072
F-stat on excluded instrument	39.58	24.74
P-value of *F*-stat	0.000	0.000

[a] Robust standard errors are reported in parenthesis. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Fixed-effects and controls are included but not reported. Averages of five years non-overlapping periods are computed to the whole set of variables.

Table 7: Impact of intraregional exports on overall Theil index (1965-2010), 2SLS[a].

Dependent variable: $diversification_{\tau}^{o,i}$	Felbermayr-Gröschl I					
	(1)	(2)	(3)	(4)	(5)	(6)
$\dot{\Omega}_{\tau}^{i}$	-0.049***	-0.015***	-0.061***	-0.014***	-0.022***	-0.015***
	(0.010)	(0.005)	(0.015)	(0.005)	(0.008)	(0.005)
$diversification_{\tau-1}^{o,i}$		0.643***		0.639***	0.606***	0.597***
		(0.048)		(0.050)	(0.050)	(0.044)
$education_{\tau}^{i}$			-0.083*	-0.039**	-0.027	-0.020
			(0.045)	(0.019)	(0.023)	(0.019)
$infrastructure_{\tau}^{i}$			-0.524**	-0.003	-0.217*	-0.231**
			(0.218)	(0.050)	(0.120)	(0.106)
$free\ to\ trade_{\tau}^{i}$			-0.027		-0.018	-0.026**
			(0.028)		(0.013)	(0.011)
$institutions_{\tau}^{i}$			-0.046*		-0.032**	-0.020
			(0.026)		(0.015)	(0.013)
$terms\ of\ trade_{\tau}^{i}$			0.218		0.470**	0.447**
			(0.355)		(0.200)	(0.198)
$volatility_{\tau}^{i}$			1.090		0.432	0.313
			(0.823)		(0.412)	(0.371)
$overvaluation_{\tau}^{i}$						0.313***
						(0.058)
Fixed-effects						
- country	Yes	Yes	Yes	Yes	Yes	Yes
- year	Yes	Yes	Yes	Yes	Yes	Yes
Countries	90	90	90	90	90	90
Observations	850	770	596	743	596	596
Durbin's χ^2-stat	26.80	2.94	40.68	2.48	10.98	4.61
P-value of Durbin's χ^2-stat	0.000	0.086	0.000	0.115	0.001	0.031
Adjusted R^2	0.80	0.94	0.81	0.94	0.94	0.95
First-stage adjusted R^2	0.90	0.92	0.95	0.92	0.95	0.95
Instrument partial R^2	0.06	0.04	0.08	0.04	0.06	0.07
F-stat on excluded instruments	49.31	23.70	29.74	22.15	18.34	24.74
P-value of F-stat	0.000	0.000	0.000	0.000	0.000	0.000
χ^2-stat			80.48	8.86	42.93	59.92
P-value of χ^2-stat			0.000	0.031	0.000	0.000

[a] Adjusted robust standard errors are reported in parenthesis. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Fixed-effects are included but not reported. Averages of five years non-overlapping periods are computed to the whole set of variables.

Table 8: Impact of intraregional exports on Theil index in its extensive and intensive margins (1965-2010), 2SLS[a].

Dependent variable:	$diversification_{\tau}^{x,i}$						$diversification_{\tau}^{i,i}$					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\dot{\Omega}_{\tau}^{i}$	-0.031***	-0.007**	-0.021***	-0.007**	-0.005**	-0.004**	-0.017***	-0.009**	-0.040***	-0.008**	-0.015***	-0.009**
	(0.007)	(0.003)	(0.006)	(0.003)	(0.002)	(0.002)	(0.006)	(0.004)	(0.010)	(0.004)	(0.005)	(0.004)
$diversification_{\tau-1}^{\{x\|i\},i}$		0.630***		0.634***	0.740***	0.741***		0.611***		0.602***	0.593***	0.580***
		(0.071)		(0.075)	(0.053)	(0.052)		(0.050)		(0.051)	(0.039)	(0.036)
$education_{\tau}^{i}$			-0.019	-0.003	0.000	0.001			-0.063**	-0.040**	-0.025	-0.018
			(0.018)	(0.012)	(0.008)	(0.008)			(0.032)	(0.018)	(0.018)	(0.016)
$infrastructure_{\tau}^{i}$			-0.018	0.024	0.101*	0.100*			-0.506***	-0.031	-0.300***	-0.313***
			(0.094)	(0.026)	(0.052)	(0.053)			(0.174)	(0.057)	(0.102)	(0.091)
$free\ to\ trade_{\tau}^{i}$			-0.019		-0.002	-0.003			-0.008		-0.013	-0.019*
			(0.012)		(0.004)	(0.004)			(0.019)		(0.011)	(0.010)
$institutions_{\tau}^{i}$			-0.015		-0.008	-0.007			-0.030		-0.023*	-0.012
			(0.010)		(0.005)	(0.005)			(0.020)		(0.013)	(0.012)
$terms\ of\ trade_{\tau}^{i}$			-0.319**		-0.102	-0.104			0.531*		0.610***	0.592***
			(0.160)		(0.080)	(0.078)			(0.284)		(0.174)	(0.183)
$volatility_{\tau}^{i}$			0.651**		0.318**	0.302**			0.430		0.059	-0.041
			(0.284)		(0.147)	(0.143)			(0.594)		(0.332)	(0.310)
$overvaluation_{\tau}^{i}$						0.036						0.264***
						(0.026)						(0.046)
Fixed-effects												
- country	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
- year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Countries	90	90	90	90	90	90	90	90	90	90	90	90
Observations	851	770	597	743	596	596	851	770	597	743	596	596
Adjusted R^2	0.53	0.88	0.76	0.88	0.93	0.93	0.82	0.92	0.83	0.92	0.93	0.94
First-stage adjusted R^2	0.90	0.91	0.95	0.91	0.95	0.95	0.90	0.91	0.95	0.91	0.95	0.95
Instrument partial R^2	0.06	0.05	0.08	0.05	0.07	0.08	0.06	0.06	0.08	0.06	0.07	0.09
F-stat on excluded ins.	49.39	28.19	29.82	27.70	21.07	28.93	49.39	35.83	29.82	34.20	25.79	34.22
P-value of F-stat	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
χ^2-stat			53.09	7.93	23.98	25.18			64.77	7.18	50.37	74.09
P-value of χ^2-stat			0.000	0.047	0.001	0.001			0.000	0.066	0.000	0.000

[a] Adjusted robust standard errors are reported in parenthesis. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Fixed-effects are included but not reported. Averages of five years non-overlapping periods are computed to the whole set of variables.

Table 9: Impact of intraregional exports on overall Theil index using the strongly balanced sample (1965-2010), 2SLS[a].

Dependent variable: $diversification^{o,i}_{\tau}$	Felbermayr-Gröschl I					
	(1)	(2)	(3)	(4)	(5)	(6)
$\dot{\Omega}^{i}_{\tau}$	-0.104***	-0.025***	-0.127***	-0.024***	-0.031**	-0.018**
	(0.022)	(0.010)	(0.035)	(0.009)	(0.012)	(0.009)
$diversification^{o,i}_{\tau-1}$		0.716***		0.724***	0.693***	0.668***
		(0.040)		(0.039)	(0.046)	(0.040)
$education^{i}_{\tau}$			-0.174**	-0.040*	-0.050*	-0.031
			(0.068)	(0.023)	(0.026)	(0.021)
$infrastructure^{i}_{\tau}$			-1.214**	-0.091	-0.301	-0.194
			(0.546)	(0.120)	(0.183)	(0.142)
$free\ to\ trade^{i}_{\tau}$			0.011		-0.008	-0.019
			(0.044)		(0.016)	(0.013)
$institutions^{i}_{\tau}$			-0.012		-0.023	-0.011
			(0.038)		(0.017)	(0.016)
$terms\ of\ trade^{i}_{\tau}$			0.218		0.395**	0.386**
			(0.481)		(0.192)	(0.191)
$volatility^{i}_{\tau}$			0.771		-0.077	-0.159
			(1.506)		(0.422)	(0.263)
$overvaluation^{i}_{\tau}$						0.295***
						(0.066)
Fixed-effects						
- country	Yes	Yes	Yes	Yes	Yes	Yes
- year	Yes	Yes	Yes	Yes	Yes	Yes
Countries	58	58	58	58	58	58
Observations	578	522	449	495	449	449
Adjusted R^2	0.61	0.94	0.48	0.94	0.92	0.94
First-stage adjusted R^2	0.97	0.97	0.98	0.97	0.98	0.98
Instrument partial R^2	0.07	0.05	0.05	0.06	0.05	0.07
F-stat on excluded instruments	24.32	13.70	12.94	14.00	12.77	18.55
P-value of F-stat	0.000	0.000	0.000	0.000	0.000	0.000
χ^2-stat			34.85	6.89	24.90	43.58
P-value of χ^2-stat			0.000	0.075	0.001	0.000

[a] Adjusted robust standard errors are reported in parenthesis. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Fixed-effects are included but not reported. Averages of five years non-overlapping periods are computed to the whole set of variables.

Table 10: Impact of intraregional exports on Herfindahl-Hirschman index (1965-2010), 2SLS[a].

Dependent variable: $diversification_{\tau}^{hh,i}$	Felbermayr-Gröschl I					
	(1)	(2)	(3)	(4)	(5)	(6)
$\dot{\Omega}_{\tau}^{i}$	-0.006***	-0.002***	-0.006***	-0.002***	-0.003***	-0.002**
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
$diversification_{\tau-1}^{hh,i}$		0.587***		0.576***	0.507***	0.513***
		(0.051)		(0.053)	(0.057)	(0.053)
$education_{\tau}^{i}$			-0.017***	-0.011***	-0.008**	-0.007**
			(0.006)	(0.004)	(0.004)	(0.004)
$infrastructure_{\tau}^{i}$			-0.049	-0.002	-0.016	-0.017
			(0.032)	(0.006)	(0.024)	(0.022)
$free\ to\ trade_{\tau}^{i}$			-0.010***		-0.005**	-0.007***
			(0.003)		(0.002)	(0.002)
$institutions_{\tau}^{i}$			-0.011***		-0.006**	-0.004
			(0.004)		(0.003)	(0.003)
$terms\ of\ trade_{\tau}^{i}$			0.021		0.049	0.046
			(0.050)		(0.036)	(0.034)
$volatility_{\tau}^{i}$			0.012		0.014	-0.006
			(0.119)		(0.069)	(0.060)
$overvaluation_{\tau}^{i}$						0.051***
						(0.012)
Fixed-effects						
- country	Yes	Yes	Yes	Yes	Yes	Yes
- year	Yes	Yes	Yes	Yes	Yes	Yes
Countries	90	90	90	90	90	90
Observations	852	772	598	745	598	598
Adjusted R^2	0.74	0.88	0.82	0.89	0.90	0.91
First-stage adjusted R^2	0.90	0.92	0.95	0.92	0.95	0.95
Instrument partial R^2	0.06	0.05	0.08	0.06	0.07	0.08
F-stat on excluded instruments	49.49	32.35	29.89	30.66	22.82	31.30
P-value of F-stat	0.000	0.000	0.000	0.000	0.000	0.000
χ^2-stat			80.87	11.86	33.92	46.14
P-value of χ^2-stat			0.000	0.008	0.000	0.000

[a] Adjusted robust standard errors are reported in parenthesis. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Fixed-effects are included but not reported. Averages of five years non-overlapping periods are computed to the whole set of variables.

Table 11: Panel type gravity equations (1960-2015), PPML[ab].

Dependent variable: $\chi_t^{i,j}$	Felbermayr-Gröschl II	Blanchard-Olney
	(1)	(2)
$\mathcal{D}_t^j$		0.441***
		(0.130)
$drought_t^j$	-1.965***	
	(0.256)	
$wildfire_t^j$	-1.857***	
	(0.227)	
$earthquake_t^j$	-1.832***	
	(0.235)	
$mass_t^j$	-0.499	
	(0.320)	
$volcanic_t^j$	-1.762***	
	(0.252)	
$flood_t^j$	-1.774***	
	(0.222)	
$landslide_t^j$	-1.955***	
	(0.262)	
$storm_t^j$	-1.775***	
	(0.224)	
$\mathcal{D}_t^j \times financial_t^j$	0.013***	-0.007
	(0.004)	(0.007)
$\mathcal{D}_t^j \times area^j$	-0.034***	0.025
	(0.013)	(0.024)
$\mathcal{D}_t^j \times population_t^j$	0.119***	-0.040**
	(0.018)	(0.017)
$\mathcal{D}_t^j \times adjacency^{i,j}$	0.053	-0.019
	(0.043)	(0.019)
$population_t^i$	0.877***	0.881***
	(0.234)	(0.249)
$population_t^j$	0.509***	0.996***
	(0.188)	(0.213)
$distance^{i,j}$	-0.825***	-0.820***
	(0.033)	(0.034)
$adjacency^{i,j}$	0.384***	0.578***
	(0.084)	(0.182)
$language^{i,j}$	0.192***	0.185**
	(0.073)	(0.072)
Fixed-effects		
- importer and exporter	Yes	Yes
- year	Yes	Yes
Countries	90	90
Observations	388,568	388,568
R^2	0.868	0.865

[a] Robust standard errors clustered at country-pair level are reported in parenthesis. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Fixed-effects are included but not reported.

[b] $\mathcal{D}_t^j$ variable denotes $natural_t^j$ term in column (1) and $death_t^j$ in column (2).

Table 12: Impact of intraregional exports on overall Theil index using alternative instruments (1965-2010), 2SLS[a].

Dep.var.: $diversification^{o,i}_{\tau}$	Felbermayr-Gröschl II						Blanchard-Olney					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\dot{\Omega}^{i}_{\tau}$	-0.046***	-0.013**	-0.058***	-0.013**	-0.020***	-0.014***	-0.058***	-0.015**	-0.069***	-0.015**	-0.021**	-0.011*
	(0.010)	(0.005)	(0.015)	(0.005)	(0.007)	(0.005)	(0.013)	(0.008)	(0.016)	(0.008)	(0.008)	(0.006)
$diversification^{o,i}_{\tau-1}$		0.649***		0.645***	0.612***	0.602***		0.641***		0.635***	0.608***	0.613***
		(0.047)		(0.049)	(0.049)	(0.043)		(0.055)		(0.057)	(0.052)	(0.045)
$education^{i}_{\tau}$			-0.078*	-0.037**	-0.024	-0.018			-0.094*	-0.040*	-0.026	-0.013
			(0.044)	(0.018)	(0.022)	(0.018)			(0.048)	(0.021)	(0.023)	(0.019)
$infrastructure^{i}_{\tau}$			-0.493**	-0.002	-0.198*	-0.216**			-0.608***	-0.004	-0.211*	-0.186*
			(0.212)	(0.050)	(0.118)	(0.105)			(0.235)	(0.050)	(0.122)	(0.107)
$free\ to\ trade^{i}_{\tau}$			-0.031		-0.020	-0.027**			-0.016		-0.019	-0.030***
			(0.027)		(0.013)	(0.011)			(0.029)		(0.013)	(0.011)
$institutions^{i}_{\tau}$			-0.044*		-0.031**	-0.019			-0.052*		-0.031**	-0.017
			(0.025)		(0.014)	(0.013)			(0.027)		(0.015)	(0.013)
$terms\ of\ trade^{i}_{\tau}$			0.269		0.498**	0.468**			0.083		0.479**	0.514**
			(0.348)		(0.198)	(0.198)			(0.404)		(0.208)	(0.204)
$volatility^{i}_{\tau}$			1.020		0.390	0.282			1.280		0.419	0.219
			(0.795)		(0.395)	(0.362)			(0.900)		(0.413)	(0.355)
$overvaluation^{i}_{\tau}$						0.308***						0.297***
						(0.058)						(0.057)
Fixed-effects												
- country	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
- year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Countries	90	90	90	90	90	90	90	90	90	90	90	90
Observations	850	770	596	743	596	596	850	770	596	743	596	596
Adjusted R^2	0.81	0.94	0.83	0.94	0.94	0.95	0.76	0.94	0.78	0.94	0.94	0.95
First-stage adjusted R^2	0.90	0.92	0.95	0.92	0.95	0.95	0.90	0.91	0.94	0.91	0.94	0.95
Instrument partial R^2	0.06	0.04	0.08	0.04	0.06	0.07	0.04	0.02	0.06	0.02	0.04	0.05
F-stat on excluded inst.	49.94	24.54	30.45	22.83	18.94	25.39	23.09	10.27	19.62	10.15	11.82	16.80
P-value of F-stat	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.002	0.001	0.000
χ^2-stat			81.65	7.67	42.90	59.05			70.71	5.25	40.67	55.59
P-value of χ^2-stat			0.000	0.053	0.000	0.000			0.000	0.155	0.000	0.000

[a] Adjusted robust standard errors are reported in parenthesis. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Fixed-effects are included but not reported. Averages of five years non-overlapping periods are computed to the whole set of variables.

Table 13: Impact of lagged intraregional exports on overall Theil index (1965-2010), 2SLS[a].

Dependent variable: $diversification_{\tau}^{o,i}$	Felbermayr-Gröschl I					
	(1)	(2)	(3)	(4)	(5)	(6)
$\Omega_{\tau-1}^{i}$	-0.055***	-0.054***	-0.060***	-0.061***	-0.072***	-0.052***
	(0.013)	(0.013)	(0.018)	(0.018)	(0.023)	(0.014)
$education_{\tau-1}^{i}$		-0.090**	-0.067	-0.063	-0.080	-0.060
		(0.044)	(0.051)	(0.052)	(0.063)	(0.045)
$infrastructure_{\tau-1}^{i}$		-0.123***	-0.436	-0.449*	-0.609*	-0.576**
		(0.033)	(0.273)	(0.271)	(0.332)	(0.244)
$free\ to\ trade_{\tau-1}^{i}$			-0.007	0.001	0.017	-0.003
			(0.026)	(0.029)	(0.036)	(0.024)
$institutions_{\tau-1}^{i}$			-0.046*	-0.053**	-0.059**	-0.034
			(0.026)	(0.027)	(0.030)	(0.023)
$terms\ of\ trade_{\tau-1}^{i}$					-0.174	-0.039
					(0.511)	(0.380)
$volatility_{\tau-1}^{i}$				1.871**	2.153**	1.602**
				(0.824)	(0.990)	(0.680)
$overvaluation_{\tau-1}^{i}$						0.606***
						(0.116)
Fixed-effects						
- country	Yes	Yes	Yes	Yes	Yes	Yes
- year	Yes	Yes	Yes	Yes	Yes	Yes
Countries	90	90	90	90	90	90
Observations	760	734	541	523	515	515
Adjusted R^2	0.77	0.77	0.81	0.80	0.75	0.85
First-stage adjusted R^2	0.91	0.91	0.95	0.95	0.95	0.95
Instrument partial R^2	0.04	0.04	0.05	0.06	0.05	0.06
F-stat on excluded instruments	29.47	28.66	19.44	19.70	15.83	24.99
P-value of F-stat	0.000	0.000	0.000	0.000	0.000	0.000
χ^2-stat		31.19	32.92	32.48	31.56	56.19
P-value of χ^2-stat		0.000	0.000	0.000	0.000	0.000

[a] Adjusted robust standard errors are reported in parenthesis. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Fixed-effects are included but not reported. Averages of five years non-overlapping periods are computed to the whole set of variables.

Table 14: Impact of intraregional exports on overall Theil index using annual data (1965-2010), 2SLS[a].

Dependent variable: $diversification_t^{o,i}$	Felbermayr-Gröschl I					
	(1)	(2)	(3)	(4)	(5)	(6)
$\dot{\Omega}_t^i$	-0.054***	-0.007***	-0.065***	-0.007***	-0.008***	-0.006***
	(0.006)	(0.002)	(0.009)	(0.002)	(0.002)	(0.002)
$diversification_{t-1}^i$		0.843***		0.847***	0.842***	0.832***
		(0.016)		(0.016)	(0.020)	(0.020)
$education_t^i$			-0.076***	-0.016**	-0.010	-0.008
			(0.026)	(0.008)	(0.009)	(0.008)
$infrastructure_t^i$			-0.543***	0.007	-0.056	-0.062
			(0.149)	(0.023)	(0.057)	(0.055)
$free\ to\ trade_t^i$			-0.028*		-0.007*	-0.010**
			(0.017)		(0.004)	(0.004)
$institutions_t^i$			-0.042***		-0.011*	-0.006
			(0.016)		(0.006)	(0.005)
$terms\ of\ trade_t^i$			0.172		0.225***	0.231***
			(0.184)		(0.064)	(0.063)
$volatility_t^i$			0.010		0.001	0.030
			(0.154)		(0.047)	(0.041)
$overvaluation_t^i$						0.109***
						(0.018)
Fixed-effects						
- country	Yes	Yes	Yes	Yes	Yes	Yes
- year	Yes	Yes	Yes	Yes	Yes	Yes
Countries	90	90	90	90	90	90
Observations	3,862	3,784	2,711	3,651	2,711	2,711
Adjusted R^2	0.78	0.97	0.79	0.97	0.97	0.98
First-stage adjusted R^2	0.90	0.91	0.95	0.91	0.95	0.95
Instrument partial R^2	0.04	0.02	0.05	0.02	0.03	0.04
F-stat on excluded instruments	170.69	77.87	95.97	76.28	53.95	68.12
P-value of F-stat	0.000	0.000	0.000	0.000	0.000	0.000
χ^2-stat			65.43	17.42	49.52	72.40
P-value of χ^2-stat			0.000	0.001	0.000	0.000

[a] Adjusted robust standard errors are reported in parenthesis. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Fixed-effects are included but not reported. Estimations are performed with the dataset at an annual frequency.

Table 15: Impact of intraregional exports on overall Theil index (1965-2010), 2SLS[a].

Dependent variable: $diversification_{\tau}^{o,i}$	Felbermayr-Gröschl I					
	(1)	(2)	(3)	(4)	(5)	(6)
$\dot{\Omega}_{\tau}^{i}$	-0.027***	-0.008**	-0.029***	-0.007**	-0.009**	-0.008**
	(0.005)	(0.004)	(0.006)	(0.003)	(0.004)	(0.004)
$diversification_{\tau-1}^{o,i}$		0.629***		0.628***	0.567***	0.568***
		(0.044)		(0.045)	(0.043)	(0.042)
$education_{\tau}^{i}$			-0.038	-0.036**	-0.014	-0.014
			(0.023)	(0.018)	(0.017)	(0.017)
$infrastructure_{\tau}^{i}$			-0.270**	0.013	-0.168*	-0.199**
			(0.115)	(0.048)	(0.093)	(0.093)
$free\ to\ trade_{\tau}^{i}$			-0.025		-0.016	-0.020*
			(0.016)		(0.011)	(0.011)
$institutions_{\tau}^{i}$			-0.030*		-0.029**	-0.023*
			(0.017)		(0.013)	(0.013)
$terms\ of\ trade_{\tau}^{i}$			0.742***		0.729***	0.657***
			(0.246)		(0.178)	(0.185)
$volatility_{\tau}^{i}$			0.267		0.032	0.048
			(0.520)		(0.338)	(0.344)
$overvaluation_{\tau}^{i}$						0.198***
						(0.052)
gdp_{τ}^{i}	-2.498***	-1.037***	-3.624***	-1.085***	-1.773***	-1.406***
	(0.382)	(0.263)	(0.413)	(0.263)	(0.292)	(0.288)
$(gdp_{\tau}^{i})^2$	0.149***	0.069***	0.222***	0.073***	0.115***	0.092***
	(0.022)	(0.015)	(0.024)	(0.015)	(0.017)	(0.017)
Fixed-effects						
- country	Yes	Yes	Yes	Yes	Yes	Yes
- year	Yes	Yes	Yes	Yes	Yes	Yes
Countries	90	90	90	90	90	90
Observations	804	735	590	713	590	590
Adjusted R^2	0.88	0.95	0.91	0.95	0.96	0.96
First-stage adjusted R^2	0.92	0.93	0.95	0.93	0.95	0.95
Instrument partial R^2	0.10	0.08	0.12	0.08	0.10	0.10
F-stat on excluded instruments	72.38	49.83	49.05	53.83	36.44	39.09
P-value of F-stat	0.000	0.000	0.000	0.000	0.000	0.000
χ^2-stat			219.91	36.78	85.82	89.58
P-value of χ^2-stat			0.000	0.000	0.000	0.000

[a] Adjusted robust standard errors are reported in parenthesis. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Fixed-effects are included but not reported. Averages of five years non-overlapping periods are computed to the whole set of variables.

Table 16: Impact of intraregional exports of non-homogeneous products on Theil index in its extensive and intensive margins (1965-2010), 2SLS[a].

Dependent variable:	$diversification_{\tau}^{x,i}$						$diversification_{\tau}^{i,i}$					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\dot{\Omega}_{\tau}^{nh,i}$	-0.055**	-0.009*	-0.023***	-0.008*	-0.005**	-0.004**	-0.038**	-0.014	-0.046***	-0.013	-0.015**	-0.009*
	(0.027)	(0.005)	(0.007)	(0.005)	(0.002)	(0.002)	(0.017)	(0.009)	(0.015)	(0.009)	(0.007)	(0.005)
$diversification_{\tau-1}^{\{x\|i\},i}$		0.662***		0.664***	0.777***	0.773***		0.632***		0.620***	0.596***	0.580***
		(0.070)		(0.073)	(0.048)	(0.047)		(0.054)		(0.055)	(0.041)	(0.037)
$education_{\tau}^{i}$			-0.029	-0.012	-0.001	-0.000			-0.090**	-0.063**	-0.037*	-0.029
			(0.022)	(0.015)	(0.008)	(0.008)			(0.044)	(0.026)	(0.021)	(0.018)
$infrastructure_{\tau}^{i}$			-0.061	0.019	0.098*	0.098*			-0.617***	-0.038	-0.322***	-0.326***
			(0.128)	(0.029)	(0.058)	(0.058)			(0.224)	(0.048)	(0.118)	(0.102)
$free\ to\ trade_{\tau}^{i}$			-0.017		-0.001	-0.002			-0.003		-0.010	-0.016
			(0.014)		(0.005)	(0.005)			(0.025)		(0.013)	(0.011)
$institutions_{\tau}^{i}$			-0.019		-0.008	-0.007			-0.043*		-0.025*	-0.014
			(0.013)		(0.006)	(0.005)			(0.026)		(0.014)	(0.012)
$terms\ of\ trade_{\tau}^{i}$			-0.410*		-0.110	-0.110			0.308		0.558***	0.562***
			(0.242)		(0.092)	(0.088)			(0.394)		(0.202)	(0.195)
$volatility_{\tau}^{i}$			0.873**		0.349**	0.330**			0.924		0.186	0.036
			(0.436)		(0.169)	(0.160)			(0.871)		(0.386)	(0.322)
$overvaluation_{\tau}^{i}$						0.032						0.262***
						(0.027)						(0.051)
Fixed-effects												
- country	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
- year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Countries	90	90	90	90	90	90	90	90	90	90	90	90
Observations	877	788	604	761	603	603	877	788	604	761	603	603
Adjusted R^2	0.64	0.85	0.61	0.85	0.92	0.93	0.64	0.90	0.74	0.91	0.92	0.94
First-stage adjusted R^2	0.85	0.87	0.92	0.87	0.92	0.92	0.85	0.87	0.92	0.87	0.92	0.92
Instrument partial R^2	0.01	0.01	0.04	0.01	0.04	0.05	0.01	0.01	0.04	0.01	0.04	0.04
F-stat on excluded inst.	6.17	7.48	16.78	7.48	18.11	23.35	6.17	5.42	16.78	5.29	15.81	20.32
P-value of F-stat	0.013	0.006	0.000	0.006	0.000	0.000	0.013	0.020	0.000	0.022	0.000	0.000
χ^2-stat			28.58	4.05	19.45	20.84			37.84	6.36	38.43	61.50
P-value of χ^2-stat			0.000	0.256	0.007	0.008			0.000	0.095	0.000	0.000

[a] Adjusted robust standard errors are reported in parenthesis. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Fixed-effects are included but not reported. Averages of five years non-overlapping periods are computed to the whole set of variables.

Table 17: Impact of intraregional exports of non-primary products on Theil index in its extensive and intensive margins (1965-2010), 2SLS[a].

Dependent variable:	$diversification^{x,i}_{\tau}$						$diversification^{i,i}_{\tau}$					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\dot{\Omega}^{np,i}_{\tau}$	-0.014**	-0.001	-0.010***	-0.001	-0.002**	-0.002*	-0.016***	-0.004	-0.022***	-0.004	-0.005	-0.003
	(0.006)	(0.002)	(0.003)	(0.002)	(0.001)	(0.001)	(0.006)	(0.004)	(0.006)	(0.004)	(0.004)	(0.003)
$diversification^{\{x\|i\},i}_{\tau-1}$		0.659***		0.661***	0.772***	0.771***		0.640***		0.636***	0.604***	0.587***
		(0.069)		(0.072)	(0.046)	(0.046)		(0.052)		(0.052)	(0.038)	(0.036)
$education^{i}_{\tau}$			0.010	0.005	0.007	0.007			-0.011	-0.034**	-0.011	-0.012
			(0.012)	(0.010)	(0.007)	(0.007)			(0.025)	(0.017)	(0.016)	(0.015)
$infrastructure^{i}_{\tau}$			0.111	0.032	0.131**	0.128**			-0.281**	-0.011	-0.191**	-0.239***
			(0.086)	(0.025)	(0.052)	(0.052)			(0.139)	(0.055)	(0.090)	(0.086)
$free\ to\ trade^{i}_{\tau}$			-0.034***		-0.004	-0.004			-0.035***		-0.022**	-0.023**
			(0.009)		(0.004)	(0.004)			(0.014)		(0.010)	(0.009)
$institutions^{i}_{\tau}$			-0.017*		-0.009	-0.008			-0.044**		-0.020	-0.013
			(0.011)		(0.006)	(0.006)			(0.021)		(0.013)	(0.013)
$terms\ of\ trade^{i}_{\tau}$			-0.248		-0.085	-0.089			0.596*		0.713***	0.664***
			(0.167)		(0.081)	(0.079)			(0.339)		(0.199)	(0.203)
$volatility^{i}_{\tau}$			0.280		0.229**	0.230**			-0.259		-0.231	-0.223
			(0.196)		(0.109)	(0.109)			(0.528)		(0.280)	(0.279)
$overvaluation^{i}_{\tau}$						0.016						0.227***
						(0.026)						(0.044)
Fixed-effects												
- country	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
- year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Countries	90	90	90	90	90	90	90	90	90	90	90	90
Observations	877	788	604	761	603	603	877	788	604	761	603	603
Adjusted R^2	0.58	0.88	0.77	0.88	0.93	0.93	0.79	0.92	0.86	0.92	0.94	0.94
First-stage adjusted R^2	0.79	0.82	0.88	0.81	0.88	0.88	0.79	0.82	0.88	0.81	0.88	0.88
Instrument partial R^2	0.03	0.04	0.05	0.04	0.05	0.05	0.03	0.03	0.05	0.04	0.05	0.05
F-stat on excluded inst.	41.05	54.43	42.48	52.90	42.30	41.85	41.05	45.68	42.48	44.37	34.26	34.10
P-value of F-stat	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
χ^2-stat			45.78	1.97	21.88	22.54			59.97	5.28	38.14	61.23
P-value of χ^2-stat			0.000	0.578	0.003	0.004			0.000	0.152	0.000	0.000

[a] Adjusted robust standard errors are reported in parenthesis. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Fixed-effects are included but not reported. Averages of five years non-overlapping periods are computed to the whole set of variables.

Appendix B

Table 18: Description of variables and sources

Variable	*Definition*	*Source*
Bilateral exports $\left(\chi_t^{i,j}\right)$	Good's bilateral export flows between country-pair (i, j). Figures are reported on free on board basis and measured in current USD. Deeper analysis are performed using bilateral flows disaggregated at commodity level according to the four digit SITC rev. 2 classification. These values are also measured in current USD.	Direction of Trade Statistics, IMF and Feenstra et al. (2005).
Capital stock per capita $\left(infrastructure_\tau^i\right)$	Stock of physical capital per capita. It is computed as the sum of general government and private capital stock adjusted by population size. Figures are measured in billions of constant 2005 international USD per capita.	Investment and Capital Stock Dataset, IMF.
Contiguity status $\left(adjacency^{i,j}\right)$	Dummy that takes the value of one if the country-pair (i, j) share a common land border, and zero otherwise.	Geographic and Bilateral Distance Database, CEPII.
Crude death rate $\left(death_t^j\right)$	Number of deaths occurring during year t in country i per 1,000 of its population.	World Development Indicators, World Bank.

continues in next page

Table 18 – continued from previous page

Variable	*Definition*	*Source*
Exchange rate overvaluation $\left(overvaluation_\tau^i\right)$	Based on Rodrik (2008), it is a real exchange rate adjusted by the Balassa-Samuelson effect: we adjust the price of tradables to non-tradables. The index is computed regressing the log of real exchange rate with the log of gross domestic product per capita; the predicted value is the Balassa-Samuelson adjusted rate. The procedure is completed taking the difference between the logs of the actual real exchange rate and the Balassa-Samuelson adjusted rate.	Darvas (2012).
Export diversification $\left(diversification_\tau^{o,i}\right)$	Overall Theil index computed as the sum of its extensive $(diversification_\tau^{x,i})$ and intensive $(diversification_\tau^{i,i})$ components. While the extensive margins reflects variations in the number of lines opened, the intensive margin reflects variations in the distribution of export values among those lines. Higher values for all the three indexes denote lower diversification level. See Cadot et al. (2011) for more details.	The Diversification Toolkit: Export Diversification and Quality Databases, IMF.
Financial remoteness $\left(financial_t^j\right)$	Financial remoteness is computed as the log of the great circle distance to the closest major financial center (London, New York, or Tokyo). Since log of zero value is not defined, it is imposed zero financial remoteness for those economies where financial hubs are located (United Kingdom, United States, and Japan, respectively).	Rose and Spiegel (2009).

continues in next page

Table 18 – continued from previous page

Variable	*Definition*	*Source*
Freedom to trade internationally $\left(free\ to\ trade^{i}_{\tau}\right)$	Index designed to measure restraints that affect international exchange. It considers the following areas: tariffs, quotas, hidden administrative restraints, and controls on exchange rates and capital movements. Higher ratings on the index are related to low tariffs, easy clearance and efficient administration of customs, a freely convertible currency, and few controls on the movement of physical and human capital. It ranges from zero to ten. For more details, see Gwartney et al. (2016).	Economic Freedom, Fraser Institute.
Geographical distance $\left(distance^{i,j}\right)$	Geographical distance between country-pair (i, j) based on weighted bilateral distances between the biggest cities of those two countries. Country position is computed weighting the share of each main agglomeration in the overall country's population. For more details, see Mayer and Zignago (2011).	Geographic and Bilateral Distance Database, CEPII.
Human capital $\left(education^{i}_{\tau}\right)$	Average years of schooling for the population aged 25 and over in country i.	Barro and Lee (2013).
Language $\left(language^{i,j}\right)$	Dummy that takes the value of one if the country-pair (i, j) share a common official language spoken by at least 20% of the population, and zero otherwise.	Geographic and Bilateral Distance Database, CEPII.

continues in next page

Table 18 – continued from previous page

Variable	*Definition*	*Source*
Legal system and property rights $\left(institutions^{i}_{\tau}\right)$	Index designed to measure the degree of protection of persons and their acquired property. It considers the following areas: rule of law, security of property rights, an independent and unbiased judiciary system, and the impartiality and effective enforcement of the law. Higher ratings denotes higher levels of protection of persons and their property. It ranges from zero to ten. For more details, see Gwartney et al. (2016).	Economic Freedom, Fraser Institute.
Natural disasters $\left(natural^{j}_{t}\right)$	Based on Felbermayr and Groschl (2013), number of "large" natural disasters -presumably orthogonal to economic factors- that hit country j. It includes: droughts ($drought^{j}_{t}$), wildfires ($wildfire^{j}_{t}$), earthquakes ($earthquake^{j}_{t}$), dry mass movements ($mass^{j}_{t}$), volcanic activities ($volcanic^{j}_{t}$), floods ($flood^{j}_{t}$), landslides ($landslide^{j}_{t}$) and storms ($storm^{j}_{t}$). Here, large is defined as disasters that: (i) caused 1,000 deaths; or (ii) injured 1,000 persons; or (iii) affected 100,000 persons, at least.	Emergency Events Database, CRED.
Net barter terms of trade $\left(terms\ of\ trade^{i}_{\tau}\right)$	Log of the percentage ratio of the export unit value index to the import unit value index, measured relative to the base year 2000 (i.e., 2000=100).	World Development Indicators, World Bank.

continues in next page

Table 18 – continued from previous page

Variable	*Definition*	*Source*
Nominal exchange rate volatility $\left(volatility_{\tau}^{i}\right)$	Based on Agosin et al. (2012), it is computed as the standard deviation of monthly values in nominal exchange rate over the complete five year non-overlapping periods involved in each average observation τ.	Darvas (2012).
Population size $\left(population_{t}^{j}\right)$	Log of total mass population in country j, it counts all residents regardless their legal status or citizenship.	World Development Indicators, World Bank.
Real gross domestic product per capita $\left(gdp_{\tau}^{i}\right)$	Sum of the gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products, divided by midyear population. Data is measured in constant 2010 USD.	World Development Indicators, World Bank.
Surface area $\left(area^{j}\right)$	Log of land surface area of country j measured in square kilometers.	Geographic and Bilateral Distance Database, CEPII.