KU LEUVEN

FACULTY OF ECONOMICS AND BUSINESS

Structural Ricardian Comparative Advantage and Network Centrality

Sergej Kaiser
0489053

Thesis submitted to obtain the degree of

Master of Science in Economics Research

Promotor: Dr. Liza Archanskaia

Academic year: 2015/2016

Contents

Abbreviations

ANBERD	Analytical Business Enterprise Research and Development
Backw.	Backward
BVAX	Backward value-added exports
F.O.B.	Free on Board
Forw.	Forward
FVAX	Forward value-added exports
I.I.D.	Independent and Identical Distributed
ISIC	International Standard Industry Classification
IV	Instrumental Variable
OECD	Organization for Economic Co-operation and Development
OLS	Ordinary Least Squares
MFN	Most Favorite Nation
MLE	Maximum Likelihood
PMM	Predictive Mean Matching
RCA	Ricardian Comparative Advantage
STAN	Structural Analysis Database
TiVA	Trade in Value-Added
WIOD	World Input-Output Database
WIOT	World Input-Output Table
WTO	World Trade Organization
VAX	Value-Added Exports

List of Figures

List of Tables

Symbols

Variable	Description
θ	Cost dispersion parameter
V	Domestic value-added share matrix
C^e	Eigenvector centrality
λ	Eigenvalue
n	A positive integer
$\delta_{i,j}$	Exporter-importer fixed effects
δ_i^k	Exporter-industry fixed effects
$F_{i,j}^k$	Exports for final demand
z_i^k	Fundamental productivity
A^k	Input-output coefficient matrix
$I_{i,j}^k$	Intermediate exports from industry k in country i to country j
a_i^k	Labour unit requirement in industry k in country i
B	Leontief inverse
g^k	Matrix of graph links
$\mathcal{G}^k$	Binary trade network in industry k
N	Graph nodes
$x_{i,j}^k$	Gross exports from industry k in country i to country j
$\tilde{x_i}^k$	Openness corrected gross exports from industry k in country i to country j
$\tilde{z_i}^k$	Observed productivity
ω	Variety
W^k	Weight matrix for industry k
ITN^k	International Trade Network in industry k

Preface

This Master thesis is submitted to partially fulfill the requirements to obtain the degree of Master of Science in Economics from the KU Leuven. Especially, I want to thank my promotor Liza Archanskaia for her helpful comments and discussions, which helped me to advance this thesis. Leuven, 15/03/2016.

1. Introduction

In the last four decades, international trade has become increasingly characterized by international production fragmentation (IPF). International production fragmentation describes that the steps to produce output are split across countries to minimize costs Johnson2012. Feenstra (1998) it marks a sharp departure from the past vertically integrated production process within a industry.

Johnson and Noguera (2012) studied the evolution of international product fragmentation over four decreased since the 1970s. The authors identify a first trend of international product fragmentation in the decade from the 1970s till the 1980s. In the following decade, the process stagnated. In contrast, for the period since the 1990's until 2008 they find that international product fragmentation accelerated strongly. Especially, according to a measure international product fragmentation, they document an increase of threefold since the 1990s. Moreover, evidence about the strength of international product fragmentation is documented by Timmer et al. (2014). The authors measured the difference of foreign content of goods for 560 products between 1995 to 2005. Their results showed that for 86 % of the products the foreign content increased. Moreover, Baldwin and Lopez-Gonzalez (2014) documented that the final goods share of exports declined for all 16 manufacturing sectors during the same period.

A consequence of IPF is that measures of international trade like gross exports including an increasing share of double counting, due to repeated border crossings of goods in the production process (Feenstra 1998). Further, the authors argued that IPF is fostering the trade in intermediate goods, which worsens the problem of double counting. Also, Johnson (2014) emphasized that gross exports are unreliable indicators of the domestic share of value-added in exports and about the origins of value-added embodied in final goods. Baldwin and Lopez-Gonzalez (2014) noted that traditional measures as gross exports of international trade do not give an accurate description of IPF. The criticism addressed about gross exports initiated a new literature on the correct measure of the value of exports.

In this literature several authors (Johnson 2014,Daudin, Rifflart, and Schweisguth 2011, Koopman, Wang, and Wei 2014) argued to focus on value-added exports instead of gross exports. An important motivation is that as Daudin, Rifflart, and Schweisguth (2011) noted, value-added exports allow to answer the policy question correctly "Who produces for whom" with international trade statistics. Koopman, Wang, and Wei (2014) contributed to the literature an accurate accounting framework to decompose gross exports into value-added exports and pure double counting, refining previous frameworks.

Hence the motivation for this thesis to study the impact of IPF on technological sources of comparative advantage. Related to the thesis, Koopman, Wang, and Wei (2014) analyzed the effects of international product fragmentation on RCA rankings with value-added exports. The authors concluded that the RCA ranking changed significantly when calculated with value-added exports. In this thesis, I contribute to the literature studying the impact of IPF on RCA by computing the structural RCA measure of Costinot, Donaldson, and Komunjer (2011) for both gross exports and value-added exports.

A limitation of the RCA ranking of Koopman, Wang, and Wei (2014) is their use of the Balassa (1965) index (BI) for the RCA ranking. The literature on the BI showed that the index has both empirical and theoretical limitations. First, from a theoretical perspective,Leromain and Orefice (2014) criticized that the BI is based on observed trade flows whereas comparative advantage in the Ricardo model is based on the fundamental productivity of countries before trade occurs (Leromain and Orefice 2014). Second, the empirical analysis of Yeats (1985) concluded that the BI has poor ordinal ranking qualities. Third, the statistical analysis of Hinloopen and Van Marrewijk (2001) found that the distribution of the BI shifted notably between countries. Therefore, the authors concluded that cross-country comparisons are problematic.In contrast, Leromain and Orefice (2014) found good statistical properties of the structural RCA. Especially, their results showed that the distribution of structural RCA is symmetric, has good ranking properties and is stable over time. Hence, the structural RCA of Costinot, Donaldson, and Komunjer (2011) is a more suitable indicator to analyze RCA.

The empirical analysis in this thesis showed that estimating the dispersion parameter from Costinot, Donaldson, and Komunjer (2011) based on backward value-added exports or gross exports yields to similar results. Regarding the first hypothesis about RCA, I find that the association of the RCA rankings based on backward value-added exports and gross exports are very similar. However, the picture changes for forward value-added exports. The rankings based on this indicator showed some differences.

A second objective of the thesis is to analyze a potential link between structural RCA and network centrality. The hypothesis is motivated by the literature on shock propagation in networks. In this literature Acemoglu et al. (2012) formulated a reduced form model of interactions among many economic actors in a network.They showed that network centrality is the first-order characteristic of an industry in one layer production network. Moreover, the authors showed that an industry with a higher network centrality contributes relative more the value added of the production network. The Ricardo model relates export shares state that trade shares reflect the pattern of cost advantages. The objective of this thesis is, therefore, to empirically analyze the association between relative network centrality and structural RCA.

On average I find that both measures are strongly associated. A more disaggregate view showed that there is considerable heterogeneity in the association. Especially, the analysis showed a high strength for countries as e.g. Chile or the USA, whereas for India the rankings were either nearly independent or showed an dissociation.

In chapter two, I describe the main assumptions in Costinot, Donaldson, and Komunjer (2011), which are necessary to obtain the relation between productivity and gross exports Moreover, I show that a simple extension of Costinot, Donaldson, and Komunjer (2011), which Further, I will sketch the interpretation and the construction of the measure of value-added exports. Further, in chapter two I present the results of the estimations and the correlation of structural RCA with gross exports and value added exports. In chapter three I describe the definition of the international trade network and define network centrality. Moreover, I present the empirical results of comparing relative network centrality and structural RCA. In chapter four I conclude.

2. Structural Ricardian comparative advantage for value-added trade

In this chapter I outline the necessary concepts to test the first hypothesis about RCA. The first hypothesis is that the RCA ranking is different if IPF is taken into account. First, I outline the concept of value-added exports, which were put forward to answer policy questions as "who is producing for whom" Daudin, Rifflart, and Schweisguth (2011). Moreover, I outline important aspects of the Ricardo model of Costinot, Donaldson, and Komunjer (2011) to explain the construction of the structural RCA measure. After outlining the model, I highlight with a simple extension how the structural RCA measure can be used to test the first hypothesis. Next, I discuss the estimation of the two components of the RCA measure. I highlight the construction of the sample and explain the data source choices. In the last part of this chapter, I compare the structural RCA ranking for value-added exports and gross exports. .

2.1. What for value-added exports and which indicator of value added exports ?

The literature on value-added exports is motivated by shortcomings of gross exports due to international production fragmentation (IPF). Before IPF emerged goods were produced in a industry with a vertical-integrated production process Feenstra et al. (2009), since IPF emerged goods are produced in several production sites in different countries Daudin, Rifflart, and Schweisguth (2011). Therefore intermediate goods frequently cross country borders in the production process. As a result of IPF gross exports include a large share of double counting, foreign value added and overstate the domestic value-added in exports Johnson (2014). Hence, they are not reliable to understand how value-added is traded between countries (Johnson 2014).

Value-added exports describe how much domestic value-added (GDP) is sold across destinations Johnson (2014). Specifically, they trace how much domestic-value added is included in the final expenditures across destination Johnson and Noguera (2012). Moreover, value-added exports is net of any double counting Koopman, Wang, and Wei (2014).

To decompose gross exports into value-added exports it is sufficient to apply Leontief's insight Wang, Wei, and Zhu (2013). I describe the intuition behind the decomposition below and give more detail in the appendix. Leontief showed based on input-output tables, which collect input requirements at each stage of the production, that one can trace the type and amount of intermediate requirements to produce one unit of output across countries and industries. Initially, a firm producing an export of the value of one dollar, creates direct domestic value-added. In addition, the exported good is produced with intermediate goods. The production of the intermediate goods emobdied in the export created a first round of indirect value-added. Furthermore, the intermediate goods were also produced using intermediate goods. The production of those intermediate goods created as well indirect value-added. Keeping track of the production structure for the whole economy, it becomes clear that the total domestic value-added induced by the production of the one dollar export, is the sum of all direct and indirect value-added.

At the country level the accounting framework of Koopman, Wang, and Wei (2014) showed how to decompose gross exports into value-added exports and pure double counting. A further refinement of the framework by Wang, Wei, and Zhu (2013) extended the decomposition to the bilateral, sectoral, sectoral-country level. However, at this level two different perspectives emerge about value-added exports. Firstly, the backward linkages perspective and secondly the forward linkages perspective Wang, Wei, and Zhu (2013).

Backward linkage value-added exports of an industry include the direct domestic value-added of that industry and further upstream domestic industries in the gross exports of the exporting industry. This perspective is based on the importing country's view. It traces the sources of exports back to a country-sector Wang, Wei, and Zhu (2013).

The forward linkages perspective traces the value-added of an industry, whether it is directly or indirectly via other industry used to satisfy foreign final demand. This perspective is a supply side view. It describes how the value-added produced in one industry is used to satisfy foreign final demand through direct and indirect exports Wang, Wei, and Zhu (2013). Further, this perspective is in line with the factor content view of trade.

The two perspectives are useful for different purposes Wang, Wei, and Zhu (2013). Firstly, an indicator baed on backward-linkages is useful to understand how much domestic value added of a country is expored via the country-setor's exports. This indicator correctly attributes how much value-added an industry exports either directly or indirectly through further downstream industries. The RCA ranking with this indicator shows how efficiently an industry uses the domestic factors of production Baldwin and Lopez-Gonzalez (2014). Secondly, the forward perspective is useful to understand how much value-added of a sector is exported by itself and other domestic industries.

In this thesis I use the value-added export data from the TiVA OECD-WTO (2015) database and in a further step cross-validate the results with the input-output data from WIOD Timmer (2012). I choose the TiVA database as the main source it has a larger country coverage with a regionally more diverse focus. In addition, to my knowledge only one author has previously used the TiVA data (Johnson 2014). Finaly, I chose the TiVA data as they have a similar aggregate industry coverage as the other two data sources.

I constructed the estimation sample with value-added export data from TIVA, R &D expenditure data from OECD (2013) ANBERD and international producer price data from the GGDC (Inklaar and Timmer 2014). The estimation sample includes twenty nine countries and twenty two industries.

Further, I concstructed a large sample with value-added exports and gross exports from the TiVA database in order to estimate a fixed effects regression. The sample covers all countries from TiVA, which had records on forward &backward value-added exports [1]. Further, to obtain a consistent sample across industries I excluded some countries, which had no exports recorded in at least one sector [2]. Finaly, excluded Saudi Arabia because its exports mainly consist of oil [3] (Organization of the Petroleum Exporting Countries 2008). This sample includes twenty two industries and fifty six countries.

I conducted the following data reconciliations, to construct the estimation sample. To combine the TiVA data with the GGDC data, I aggregated the international price data for to create the manufacturing industry 17-19 (ISIC rev.3.1) from the industries 17 and 18-19. Further, I aggregated three service industries. First, I constructed the industry 50-55 from the industries 50, 51 ,52 and 55. Second, I took a weighted average of industry 60 and 64 for the industry 60-64. Finally, the international price data for industry 75-95 is constructed from the industries 75, 80, 90-93 and 95. Specifically, I aggregated the prices using a weighted average with weights equivalent to the share value-added of an industry among the industries aggregated in this step. I obtained the value-added output data from OECD (2011) STAN database. I report in the appendix a table A.4 with the descriptive statistics of the sample.

1. 15 countries did not have positive records on forward linkages value-added exports and were therefore omitted from the sample. The following countries were thus omitted, Lithuania Latvia, Malta, Malaysia, Philippines, Romania, Rest of the World, Russia, Singapore, Thailand, Tunisia, Taiwan, Vietnam, South Africa, Costa Rica, Brunai Darussalem, Khambodia, Island.

2. Island, Costa Rica, Brunei Darussalam

3. For 2005, the share of petroleum exports accounts for 90% of the fob exports. Fob denotes the price of a good at the factory excluding delivery and insurance costs (Combes, Mayer, and Thisse 2008, p.78)

2.2. Ricardian model

In this subchapter, I describe the main assumptions of the Ricardo model by (Costinot, Donaldson, and Komunjer 2011), which outlined the structural RCA measure used in this thesis.

In general, the model considers a world economy of $i = 1, \ldots, n$ countries and $k = 1, \ldots, K$ industries. The sole factor of production is labour, which is perfectly mobile across industries and immobile between countries. The number of workers is denoted with L_i for each country i and w_i denotes their wage.

The production technology is as follows. Each industry produces a good with constant returns to scale technology. Further, each good has indefinitely many varieties $\omega \in \Omega$. Moreover, the productivity differences are stochastic. The fundamental productivity $z_i^k(\omega)$ denotes how much of a variety ω may be produced with one unit of labor. The productivity is for each country, industry and variety (i, k, ω) a random draw from the Fréchet distribution. Therefore

$$F_i^k(z) = exp[-(z/z_i^k)^{-\theta}]$$

The transport cost are modeled as iceberg trade cost. For one unit of a good, which is shipped form industry k in country i to country j only a fraction $1/d_{i,j}^k \leq 1$ arrives. Further, the authors assume that no cross-country is possible. There for any third country l importing a good k from country i through another country j is more costly than directly importing it. Formally, $d_{i,l}^k \leq d_{i,j}^k d_{j,l}^k$ for any third country l.

Further, the model assumes that the market structure is perfect competition. Therefore each consumer seeks the lowest price of each variety of a good. The perfect competition assumption implies together with the constant returns to scale production technology that $p_{i,j}^k = min_{1 \leq i \leq I} c_{i,j}^k$. The cost of producing and shipping one unit of the variety are as follows $c_{i,j}^k = \frac{d_{i,j}^k w_i}{z_i^k}$. The price a customer pays for a variety of a good, is therefore a composite of the production cost $\frac{w_i}{z_i^k}$ and the shipping cost $d_{i,j}^k$.

Moreover, the consumer preferences are specified with a two-tier utility function. The upper-tier is a Cobb-Douglas function and lower tier CES function. The choice of the CES function implies that the consumers show a 'love-for-variety' property. It implies that consumer welfare increases monotonically with the number of goods for a given level of expenditures on goods and a given price of a variety Helpman and Krugman (1985, p. 118). The consumer preferences imply the following relation for the total expenditure of any country j on a variety ω of a good k.

$$x_j^k(\omega) = \left(\frac{p_j^k(\omega)}{\left(\sum_{\omega' \in \Omega} p_j^k(\omega')^{1-\sigma_j^k}\right)^{1/(1-\sigma_j^k)}} \right) \alpha_j^k \, Y_j \quad \text{where} 0 \leq \alpha_j^k < 1, \sigma_j^k < 1 + \theta \quad \text{and}$$

The equation describes that the expenditures of an importing country depends on the pattern of relative prices and the share of income it spends on the particular good.

The assumptions guarantee that bilateral trade satisfies the following condition

$$x_{i,j}^k = \frac{(c_{i,j}^k)^{-\theta}}{\sum_{i'=1}^{I} (c_{i',j}^k)^{-\theta}} \alpha_j^k Y_j \quad \text{and} \quad Y_j = w_j L_j \tag{2.1}$$

Moreover, it can be shown that the following Lemma holds.

$$\ln \left(\frac{x_{i,j}^k x_{i',j}^{k'}}{x_{i,j}^{k'} x_{i',j}^k} \right) = \theta \ln \left(\frac{z_i^k z_{i'}^{k'}}{z_i^{k'} z_{i'}^k} \right) - -\theta \ln \left(\frac{d_{ij}^k d_{i',j}^{k'}}{d_{i,j}^{k'} d_{i',j}^k} \right) \tag{2.2}$$

The first log difference of $x_{i,j}^k / x_{i,j}^{k'}$ accounts for differences in wages w_i across exporting countries and incomes Y_j across importing countries. Further, the second log difference $\left(x_{i,j}^k / x_{i,j}^{k'} \right) \left(x_{i',j}^k / x_{i,j}^{k'} \right)$ accounts for differences in the expenditure shares α_j^k across destinations. Therefore the ratio of relative exports of country i and i' to country j in industry k and k' is determined by the relative ratio of productivity and the relative ratio of trade cost. The lemma shows that the model makes a similar prediction as the original Ricardo model at the industry level.

2.3. Empirical predictions

However, the prediction above is based on fundamental productivity differences, which can not be empirically observed. To make the model empirically viable, a link between fundamental and observed productivity is necessary.

Costinot, Donaldson, and Komunjer (2011) showed that based on the assumed distribution the ratio of observed productivities $\tilde{z}_i^k/\tilde{z}_{i'}^k$ for a country pair i and i' links directly to the ratio of fundamental productivities. Based on this insight they showed the following theorem.

$$\ln\left(\frac{\tilde{x}_{i,j}^k \tilde{x}_{i'j}^{k'}}{\tilde{x}_{i,j}^{k'} \tilde{x}_{i'j}^k}\right) = \theta \ln\left(\frac{\tilde{z}_i^k \tilde{z}_{i'}^{k'}}{\tilde{z}_i^{k'} \tilde{z}_{i'}^k}\right) - -\theta \ln\left(\frac{d_{ij}^k d_{i'j}^{k'}}{d_{i,j}^{k'} d_{i'j}^k}\right) \tag{2.3}$$

The equation above relates openness corrected exports $\tilde{x}_{i,j}^k$ to observed productivity and trade cost. The first ratio $\frac{\tilde{x}_{i,j}^k}{\tilde{x}_{i',j}^{k'}}$ accounts for income differences Y_j of the importing countries and wage differences w_i across exporting countries. Further, the second ratio $(\frac{\tilde{x}_{i,j}^k}{\tilde{x}_{i'j}^{k'}})/\frac{\tilde{x}_{i'j}^k}{\tilde{x}_{i'j}^{k'}})$ accounts for differences in the expenditure shares α^k. Moreover, productivity term is netted of specific trade barriers $\delta_{i,j}$ between country i and j like distance and of trade barriers δ_j^k specific imposed by the importing country$j$ on the k goods [4]. Further the error term $\epsilon_{i,j}^k$ includes variable trade cost.

To estimate the equation, it is necessary to specify the measure of observed productivity. In the context of the model assumptions, the inverse ratio of the producer price fully reflects the observed ratio of relative productivity.

Equivalently, the following equation may be estimated may be estimated instead (Costinot, Donaldson, and Komunjer 2011).

$$\ln \tilde{x}_{i,j}^k = \delta_{i,j} + \delta_j^k + \theta \ln \tilde{z}_i^k + \epsilon_{i,j}^k \tag{2.4}$$

The eq. (2.4) states that the openness corrected exports $\tilde{x}_{i,j}^k \equiv x_{i,j}^k - \tilde{\pi}_{i,i}$ from industry k in exporting country i to importing country j are predicted by the observed productivity $\ln \tilde{z}_i^k$, exporter-importer fixed-effects $\delta_{i,j}$ and importer-industry fixed-effects δ_j^k.

Costinot, Donaldson, and Komunjer (2011) highlight that the equation may be interpreted similar to a 'difference-in-difference' estimation. The productivity term in eq. (2.4) is first differenced of specific trade barriers $\delta_{i,j}$ between country i and j like distance and second differenced of trade barriers δ_j^k specific imposed by the importing country$j$ on the k goods. Further the error term $\epsilon_{i,j}^k$ includes variable trade cost and other unobserved time-varying components.

The structural RCA measure is obtained in two steps. In the first step I estimate θ and in the second step I estimate the full fixed effects regression to obtain the exporter-industry fixed effect.

$$\ln x_{i,j}^k = \delta_{i,j} + \delta_j^k + \theta \ln z_i^k + \epsilon_{i,j}^k \tag{2.5}$$

$$\ln x_{i,j}^k = \delta_{i,j} + \delta_j^k + \delta_i^k + \epsilon_{i,j}^k \tag{2.6}$$

The exporter-industry fixed effect δ_i^k in the second equation is equivalent is similar $\theta \ln z_i^k$ on the bilateral gross exports in the first equation. Specifically, the fixed effects regression identifies only $(\delta_i^k - \delta_i^{k'}) - (\delta_{i'}^k - \delta_{i'}^{k'})$, and therefore $(\ln z_i^k - \ln z_i^{k'}) - (\ln z_{i'}^k - \ln z_{i'}^{k'})$. However, the authors note that this difference-in-difference estimations capture the core of the Ricardian comparative advantage. The structural comparative advantage measure is than as follows $e^{\delta_i^k/\theta}$.

2.4. Generalization

In the following, I discuss the effect of the sector-specific use of production factors. The use of production factors may be more industry specific because goods are increasingly produced by specialized supply-chains.

4. The latter fixed-effect include as well trade protection in line with the most-favorite nation (MFN) clause of the WTO (Costinot, Donaldson, and Komunjer 2011). The MFN clause is that a country can not offer less favorable conditions to a party e.g. an investor of an agreement than to any other investor in the same specific matter from a third country (OECD 2004).

The extensions simply introduce another index to the factor price of the production factor.

$$c_{i,j}^k = \frac{d_{i,j}^k}{z_i^k Y_j} w_i^{\alpha^k}$$

. With the cost function above however the ratio of exports expression as in the lemma would include another term with the ratio of wages. Therefore, the structural RCA measure would then not only reflect productivity differences and differences in the factor price.

A special case of the generalized cost function is the relation in theorem. If the production factor used are not industry specific, I obtain the model predictions.

The argument above highlights that the first hypothesis, that the RCA ranking may be significantly different under value-added exports than under gross exports, can be interpreted as test whether the usage of production factors is sector specific.

Further, one could easily extend the cost function to include international sourcing of inputs. Analogue to the previous argument about sector-specific production factors, the international sourcing pattern would additionally confound the RCA ranking.

2.5. Estimation method

In this section, I describe the estimation procedure of θ and the industry exporter fixed effects to construct the structural RCA measure. Two aspects of the estimation are central to this section. Firstly, I motivate why I estimate θ with an instrumental variable (IV) regression. Secondly, I discuss the missing data techniques, which I employ in the estimation. In the last step, I describe the construction of the sample.

Estimating θ under OLS yields unbiased and consistent interference if the independent variable is uncorrelated with the error term. To assess this assumption it useful to highlight the interpretation of the error term in the model. The error term may be interpreted as variable trade cost and other unobserved time-varying variables. Hence, the assumption requires that the inverse of producer prices should be uncorrelated to variable trade cost.

Costinot, Donaldson, and Komunjer (2011) highlight two reasons, why they think that an IV estimation may be necessary. The first reason is a simultaneity bias. Agglomeration effects, e.g. positive spillover from an exporting firm on the exports of other firms in spatial proximity (Bernard and Jensen 2004)., could cause a simultaneity bias. The sign of the bias is a priori ambiguous (Costinot, Donaldson, and Komunjer 2011).

Second, a measurement error may cause the estimate to be biased toward zero. In general in can be shown that if the dependent variable is measured with a random error the estimated coefficient would be biased towards zero Angrist and Pischke (2008). Both biases cause an endogeneity problem, which may be solved using an IV estimation (Dhaene 2014, p.139).

Finally, under a more general cost function, the inverse of producer prices may reflect other sources of comparative advantage than productivity differences. The IV estimation of the producer prices with the instrument R & D expenditure may lead to a better identification of the effects of productivity.

The IV identification strategy rests on two assumptions about the instrument. The exclusion assumption requires that the instrument is uncorrelated with the error term (Cameron and Trivedi 2005). In other terms, the instrument R & D expenditures should only affect the independent variable gross exports through the endogenous producer prices. Further, the relevance assumption requires that the instrument is sufficiently strong correlated with the endogenous regressor.

I use R & D expenditures as an instrument for the following reasons. Firstly, I expect that higher R & D expenditures increase the productivity of an industry. In the model framework, a lower cost of producing a good would be directly passed through to the price and hence increase the exports. According to this relationship, I expect that in the first stage the coefficient of R& D expenditures is positive and statistically significant. Such a relationship between productivity of industries and R&D is e.g. hypothesized in Griffith, Redding, and Reenen (2004). Moreover, under this hypothesis R&D as an instrument would also satisfy the second assumption. This assumption, however can not be empirically tested (Cameron and Trivedi 2005, p.109). Secondly, Costinot, Donaldson, and Komunjer (2011) estimation strategy used R&D as instrument to .

However, if R&D would not be sufficiently strong correlated wit the endogenous regressor, it would be a weak instrument. A weak instrument causes two problems. Firstly, the IV estimator would not identify the causal effect of the endogenous variable (Bound, Jaeger, and Baker 1993). Moreover, the IV estimate becomes inconsistent if the weak instrument is correlated with the error term (Bound, Jaeger, and Baker 1993). In the estimation, I report the first-stage F-statistic of the excluded instrument. As a rule-of-thumb, a weak instrument can be ruled out if this F-statistic exceeds ten (Douglas Staiger 1997).

2.5.1. Missing data imputation

Concerns about missing data are (1) efficiency losses (2) complications in data handling and data analysis (3) bias due to differences between the observed and unobserved data (Schafer and Olsen 1998). In the first sample, I have missing data in the instrumental variables, R & D. The concern for the IV estimation here, are that efficiency losses in the first stage regression may reduce the strength of the first stage association between R&D and the inverse of producer prices. The second stage IV estimates would, therefore, show an upward biased estimate of θ. In the following, I motivate and describe the employed missing data technique.

To impute the missing data I used the method of multiple imputations, which is a Bayesian technique to impute missing data by simulated draws from the posterior predictive distribution[5] Rubin (1987). It was initially

5. The posterior distribution is in Bayesian interference obtained by dividing the product of the assumed prior distribution and the Likelihood by normalizing constant. The posterior predictive distribution describes the

proposed in Rubin (1978) for nonresponse in surveys, and it's statistical properties were developed in Rubin (1987). I chose multiple imputations for the following reasons. Firstly, techniques ignoring the missing observations in the analysis as, e.g., complete case methods or case-wise deletion require stronger assumptions about the missing data. Ignoring the missing observations leads to unbiased interference if the missing data is a random subset of all observations (Bhaskaran and Smeeth 2014). Secondly, single imputation methods do not take into account the uncertainty induced by the missing values and as a result, the estimated variances would be downward biased (Imbens and Wooldridge 2007). Multiple imputation offer a simple and general approach to deal with missing data, which correctly accounts for the uncertainty induced by missing observations (Schafer and Olsen 1998).

In the following, I outline multiple imputation based on Little and Rubin (2002, p.209-211). Initially, I describe certain assumptions and some notation, which I use in the outline of multiple imputations. To start with I assume that the indicators of missing values are random variables with distribution. Additionally, I assume the probability of missingness depends only on observed data and is independent of unobserved data. Further, I denote with κ the parameter of interest.

Multiple imputation is a simulation method to impute missing values by draws based on the predictive posterior distribution. The basic idea is to relate the observed posterior distribution to the complete-data posterior, which would be observed in the absence of missing data. It can be shown that the complete-data posterior distribution can be simulated by drawing the missing observations from the joint posterior distribution of the missing data and the observed data. The predicted missing values are used to impute the data set. The estimate of the parameter is obtained by drawing from the complete data posterior distribution. Moreover, if the complete data posterior mean and variance are of interest only, these quantities can be obtained by simulated draws from the predictive posterior distribution. Specifically, the posterior mean is obtained by taking the mean of simulated draws from the predictive posterior distribution $\kappa = 1/m \sum_{imp=1}^{m} \kappa^{imp}$. Slightly more complex, the variance is a combination of the average of the variance of each imputed data set and the between imputations variance $Var(\kappa|X) = \sqrt{\bar{V} + 1/(M) * B}\, where B = 1/M - 1 * \sum_m = 1^M (\kappa_m - k\bar{appa}_m)^2$ and $\bar{V} = 1/M * \sum_m = 1^M V_m$ with$V_M = Var X_{obs}, X_{mis}^(d)$.

The strength of the multiple imputation method allows is that the imputation and the analysis model can be different Little and Rubin (2002, p.217). The interference obtained is valid and leads only in extreme cases to problems.

After the imputation complete-data methods can be used independently on the M data-sets and the mean and the variance can be pooled based on simulated draws as described. In a first step the data set is imputed using simulated draws and in the second step, complete-data methods can be used on the M data sets and in the third step the results are pooled. Moreover, I combine multiple imputation with predictive mean matching (PMM). PMM is a nearest neighbour matching technique outlined by Rubin (1986). In the framework of MI the PMM technique replaces imputes the missing data with draws from the observed data, which are the closest to the values of a simulated regression model. Because the missing data is imputed from the observed data, the distribution of the imputed variables is similar to the observed variables.

The imputation method is suited to impute skewed variables, which would violate the normality assumption invoked by multiple imputation using regression techniques (White, Royston, and Wood 2011). Both imputed variables R&D and producer prices are highly skewed variables. Moreover, simulation studies analyzing MI with PMM found that the imputation results were efficient and unbiased Morris, White, and Royston (2014).

In the following, I outline PMM based on (White, Royston, and Wood 2011). Initially, each missing variable is regressed on the imputation model and the other variables with missing values. From this regression a set of estimates β and corresponding variances V are obtained. Further, a simulated β^* is than obtained from a multivariate normal distribution. The imputation of the missing values is based on a random draw from the q [6] closest observations, which minimize the distance between the product of the predicted value of the regression βx_h and the predicted values based on the simulated parameter $\beta^* x_h$.

I implemented the imputation as follows. First, I decided to impute both independent and the dependent variable of the first stage regression. Therefore both variables R& D and producer prices are imputed. In this choice, I follow the recommendation in the simulation study Moons et al. (2006), which found that multiple imputation without imputing the outcome leads to biased results.

predicted value average over the posterior distribution.

6. I chose $q = 10$. Therefore a missing value is imputed by a random draw from the ten closest observations. This choice rests on the recommendations in the simulation study of Morris, White, and Royston (2014).

Moreover, I included in the regression imputation model country and industry fixed effects dummies. The country and industry fixed effect as covariates account for the variation at country and industry level. The fixed effects variables should, therefore, account for the time-invariant determinants of both variables. Moreover, the imputation of the log of R&D in Costinot, Donaldson, and Komunjer (2011) was based on the same covariates.

2.6. Empirical results structural RCA

In this section, I describe the empirical results of the intermediate step of estimating the parameter θ. Further, I show the results of structural RCA measure. Further, I analyze the association of the RCA indicator for the three indicators in two steps. First, I present the association of structural RCA against GDP plot, which shows whether the association of RCA ranking is related to GDP per capita. Second, I compare the structural RCA ranking for the country pair Belgium and the Germany across the manufacturing industries.

Table one shows the cross-sectional estimation results for the year 2005. The columns (2)-(4) report the IV estimates of θ for different samples regarding the industry coverage and country coverage. The results show that estimates are robust. First, I note that in both tables as expected the point estimates of θ are positive and significant. In column (4) I reduce the sample countries to include only high-income countries based on the world bank classification for 2005.

The OLS estimates for gross exports, and backward value-added exports show a small yet statistical strongly significant coefficient. The IV estimates in the columns (2)-(4) are significantly increased with an estimated θ between 12.63 and 14.68. I interpret the increase of the IV estimate as an indicator that the independent variable is endogenous since otherwise, both estimates should show no significant difference Hausman (1978). From a substantial point of view, there are mainly two reasons, why I use an instrument to account for the potential endogeneity of $\tilde{z}_i^k$. First, the estimates of θ might be biased because of measurement error in the international price data. The measurement error would bias the estimate towards zero (Angrist and Krueger 2001).

The results of the first stage regression address two concerns about the validity of the IV regression, first the relevance of the instrument and whether the causal effect of the instrument is as expected. Firstly, concerning the relevance of the instrument, the table (see appendix) shows the F-statistic of the excluded instrument in the first stage is across the specifications very high and hence the instrument is highly relevant. Further, the first stage shows a statistically significant positive effect of R&D on the inverse of producer prices, which confirms the hypothesis of a positive causal effect of R&D on the producer prices.

The IV estimates of θ show following results. First, the comparison between the full and the sample excluding primary industries shows that the estimates are similar for θ for both backward value-added and gross exports. In general comparing the estimates of θ for both dependent variables, I conclude that the estimates are statistically not significantly different. Second, the pattern for both dependent variables shows that the estimate from the sample excluding the primary industries is statistically not significantly decreased. Third, the sample excluding no high-income countries and primary industries shows a statistically significantly increased estimate compared to the full sample [7]. From a substantial point, a higher point estimate of θ implies a decreased dispersion. The effect is expected, as the sample includes only high-income countries.

The table for the dependent variable forward value-added exports shows noticeable differences. In the first column, the OLS estimate shows a negative sign and is not significant. The negative sign of the OLS estimate may be caused by bias. However, since the OLS estimate is not significant, I forgo a further discussion. The IV estimates for the different samples are statistically significant and show positive estimates. Surprisingly, the estimate for the last sample, which includes only high-income countries, is not increased. Overall the estimates for forward value-added exports are very stable and change little across the different samples.

Compared to Costinot, Donaldson, and Komunjer (2011) I find that the estimates without primary industries, which is the closest matching sample, show comparable estimates.However, the authors best estimate of 6.58 is significantly lower than the results I obtained. The authors obtained the estimate with the dependent variables openness corrected gross exports. They argued that openness corrected gross exports are necessary to account

7. Based on an extension of the t-test to the multiple imputation setting. The distribution of test statistic is a t-distribution with v degrees of freedom, where $v = (m-1) * (1 + ((1 + M{-}1) * B/\bar{U}){-}1)^2$ and $\bar{U}$ denotes the average within-imputation variance and $B = 1/m - 1 * \sum_i = 1^m \theta_i - \bar{\theta})$ denotes the between imputation variation of the estimated parameter Rubin (1987, p.77)

for trade selection [8] downward biases the differences in observed productivity compared to the fundamental productivity. Therefore, they reasoned that the estimates of θ with gross exports are upward biased.

In this thesis, I use gross exports and value added exports without an openness correction. First, the authors used the import penetration ratio to correct for openness. This measure is only available for the manufacturing industries and hence applying the correction would reduce my sample coverage. Second, it is not clear, how to correct for openness for the dependent variable value-added exports [9].

8. Trade selection denotes that a country does not produce certain goods for which they receive a low productivity draw and instead imports them (Costinot, Donaldson, and Komunjer 2011).

9. A possible definition openness for value-added exports might be the ratio of re-imported value added from domestic industries to value-added exports . However, the estimates of θ with this correction showed implausible values.

(a) Cross-section results I

	(1) OLS	(2) Full Sample	(3) Without primary industries	(4) Without primary industries high [10]
Dependent variable log gross exports in 2005				
Log productivity	0.43	12.65	11.42	14.69
	(0.067)	(1.331)	(1.422)	(2.13)
Exporter-Importer Fixed Effects	YES	YES	YES	YES
Importer-Industry Fixed Effects	YES	YES	YES	YES
Observations	18143	18143	16582	14449
R-squared*	0.77	0.20	0.32	0.14
First-stage F-statistic exc. instrument		151.41	125.60	85.24

Heteroscedasticity robust standard errors in parentheses
Log Productivity is instrumented in columns 2-4 with log of R&D expenditures
Without primary industries excludes the industries agriculture and mining
* avg. of imputed R-squareds

(b) Cross-section results II

	(1) OLS	(2) Full Sample	(3) Without primary industries	(4) Without primary industries high
Dependent variable log backward value-added exports in 2005				
Log Productivity	0.48	12.91	11.76	15.08
	(0.066)	(1.340)	(1.447)	(2.180)
Exporter Importer Fixed Effects	YES	YES	YES	YES
Importer Industry Fixed Effects	YES	YES	YES	YES
Observations	18085	18085	16538	14412
R-squared*	0.78	0.18	0.30	0.13
First-stage F-statistic of exc. instrument		151.41	125.60	85.24

Heteroscedasticity robust standard errors in parentheses
Log Productivity is instrumented in columns 2-4 with log of R&D expenditures
Without primary industries excludes mining and agriculture industry
* avg. of imputed R-squareds

(c) Cross-section results III

	(1) OLS	(2) Full Sample	(3) Without primary industries	(4) Without primary industries high
Dependent variable log forward value-added exports in 2005				
Log Productivity	−0.0191	9.29	10.33	10.22
	(0.0454)	(0.868)	(1.291)	(1.199)
Exporter Importer Fixed Effects	YES	YES	YES	YES
Importer Industry Fixed Effects	YES	YES	YES	YES
Observations	16727	16727	15271	14095
R-squared*	0.88	0.48	0.43	0.49
First-Stage F-statistic of exc. instrument		151.41	125.60	85.24

Heteroscedasticity robust standard errors in parentheses
Log Productivity is instrumented in columns 2-4 with log of R&D expenditures
Without primary industries excludes the industries mining and agriculture
* avg. of imputed R-squareds

Table 2.1.: Cross-section Results OLS and IV

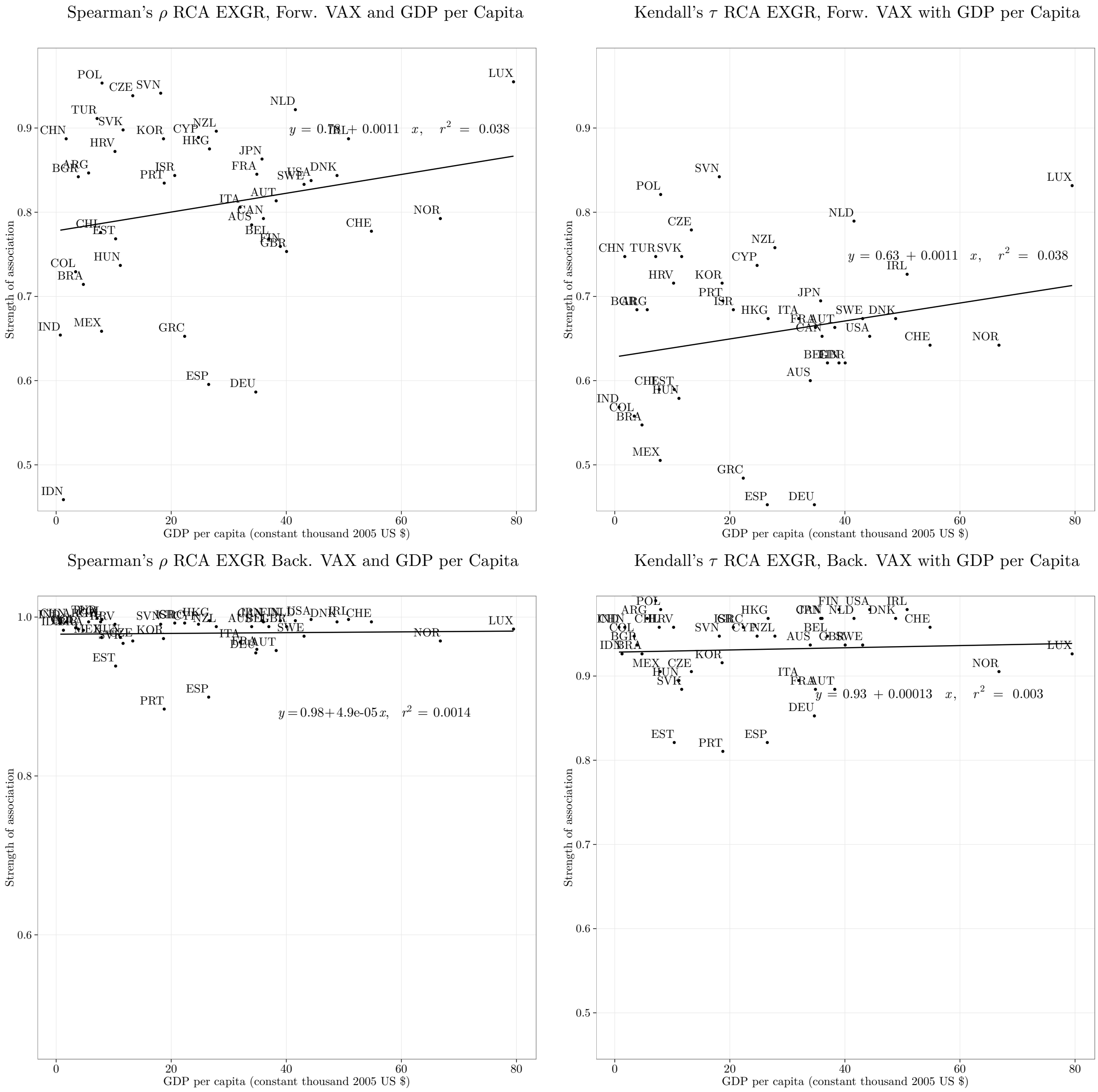
In the following, I discuss the results of the structural RCA ranking for both value-added exports measures and gross exports. Especially, I will show two applications of the structural RCA ranking. First, I show a scatter plot of the association of the RCA ranking for gross exports and value-added exports and a country's GDP per capita. The plot is used to analyse the hypothesis that countries with a higher GDP show a greater similarity between the rankings. The reason may be that more developed countries use the production factors in a less sector-specific manner.

The second application compares the structural RCA ranking for Belgium and Germany based on forward and backwards value-added exports and gross exports. By analysing the pattern of RCA for a country pair, it can be directly seen whether value-added exports change the conclusions about a country's comparative advantage. Moreover, I research whether both perspectives show a similar story about countries comparative advantage industries.

2.6.1. Structural Ricardian comparative advantage based gross exports and value-added exports

To compare the RCA rankings I have chosen the following association. First of all, I chose the Spearman's ρ since I focus on the similarity of the rankings. Moreover, I chose Kendall's τ as it computes the similarity of the two rankings, by counting the number of country pairs, which are different between two rankings. Also, Kendall's τ may be interpreted as the difference between the probability of concordance and discordance of two variables (**newson2parameters**). Finally, both measures are invariant to any rescaling of the ranking variable, which preserves the ordering of the countries. The property is useful, as the estimates of θ are in the upper range of previous results.

Figure 2.1.: Association RCA based on VAX & EXGR and GDP per capita



I conclude four findings from the figures above. Firstly, the RCA rankings based on gross exports and backward value-added exports show a high degree of similarity for all countries. Secondly, the association between forward value-added exports and gross exports is substantially lower than the association of backwards value-added and gross exports. Thirdly, the association of gross exports and forward value-added exports show a weak positive association with a country's GDP per capita. However, the positive relation in the graph is not robust to the exclusion of the country with the lowest and the highest RCA from the graph. Such a change would alter the graph to a flat line. As a final point, I find that the overall strength of the associations is higher using Spearman's ρ compared to Kendall's τ.

The first and second finding are similar to the results in the estimation of the θ parameter, which showed similar estimates using gross exports and backward value-added exports while the estimates using forward value-added exports showed differences. The finding of a smaller association for Kendall compared to Spearman is unsprising, given that the population analog of Spearman's ρ and Kendall's τ is equal to three half (Fredricks and Nelsen 2007).

The country pair graph comparing RCA for the three indicators backward/forward value-added exports and gross exports present a more local view of RCA. below I present the normalized RCA based on both value-added export measures and gross exports for the industries of the manufacturing sector. the RCA [11] as in Leromain and Orefice (2014). The normalized RCA has the following interpretation, a value above (below) one indicates a comparative (dis)advantage of a country in a specific industry.

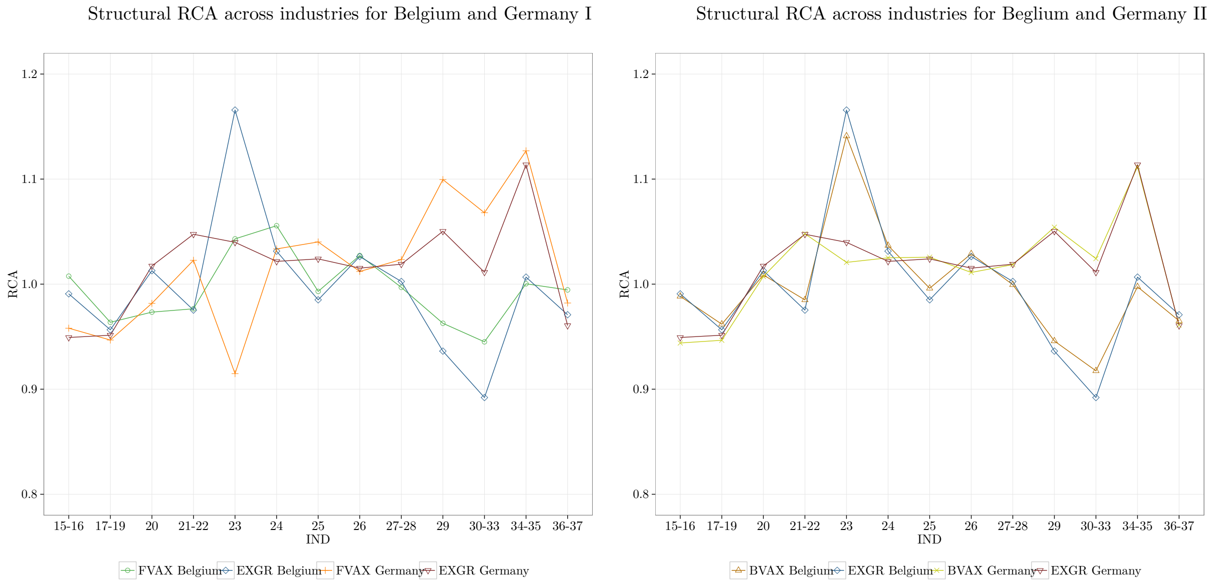
To start with the RCA ranking results, I discuss the results for backward value-added exports and gross exports. I find that for both countries backward value added closely traces the RCA pattern of gross exports.

The figure highlights that both countries highlights have a comparative advantage in the following industries wood industry (ISIC Rev.3 20), fuel products (ISIC Rev. 3, 23), chemical (ISIC Rev. 3 24) and other non-metalic mineral products (ISIC Rev. 3 26). Germany has a higher comparative advantage in seven industries namely the wood industry, paper and printing(ISIC. Rev. 3 21-22), rubber and plastics (ISIC Rev. 3, 25), basic metals and metal products (ISIC Rev. 3, 27-28), machinery and equipment (ISIC Rev. 3, 29), electrical and optical equipment (ISIC Rev. 3, 30-33) and transport equipment (ISIC Rev.3, 34-35). On the other hand, Belgium has a higher comparative advantage in six industries namely the food industry (ISIC Rev. 3 15-16), textiles, leather and footwear industry (ISIC Rev.3, 17-19), fuel products, chemical, manufacturing and recycling (ISIC Rev. 3 ,36-37).

In contrast, the RCA rankings based on forward value-added gross exports show larger differences . The figure shows the largest decrease of RCA in the industries paper products (ISIC. Rev. 3, 21) and wood (ISIC. Rev. 3 20). Additionally Germany shows a large decrease of RCA for the paper and printing products industry . The largest decrease of RCA induces a change of 15% in the industry chemical products industry (ISIC. Rev. 3 23) for both countries. As a result, the chemical products industry changes from a comparative advantage to an industry of comparative disadvantage for Germany. Additionally, in wood industry both countries no longer show a comparative advantage under forward value-added exports whereas under gross exports they show a comparative advantage. Moreover, I observe the largest increase of RCA under forward value-added exports compared to gross exports in the industry 29 and 30-33. For both countries, I observe an increase of about 5% in the industry 30-33 for forward value-added exports and a somewhat smaller increase in the industry 29. However, the pattern of RCA remains unchanged for both countries, for Belgium, a comparative disadvantage and Germany has a comparative advantage. Concluding, the graphical analysis show that forward value-added exports changes the pattern of RCA, whereas the RCA pattern under backward value-added exports and gross exports is similar.

11. Formally, I define it as follows $RCA_i^k = \frac{z_i^k * \bar{z}}{\bar{z}_i * \bar{z}^k}$, where $\bar{z} = 1/NK * \sum_{i=1}^{N}\sum_{k=1}^{K} z_i^k$ denotes the grand mean, $\bar{z}^k = 1/N\sum_{i=1}^{N} z_i^k$ denotes the industry specific mean and $\bar{z}_i = 1/K\sum_{k=1}^{K} z_i^k$ denotes the country specific mean.

Figure 2.2.: Comparative advantage based on for- and backward VAX & EXGR for Belgium and Germany



3. Relative network centrality and structural Ricardian comparative advantage

In this chapter, I analyze the association between structural RCA and network centrality. The motivation is as follows. First according to Ricardo model an industry within a country with a comparative cost advantage is exporting more of the good and hence has a higher exports share. The eigenvector centrality of a country-industry in the network is high if it exports to destinations which are important exporters themselves. A higher centrality of a country corresponds to higher trade shares. The similarity motivates the hypothesize that both measures may be linked. Furthermore, the literature on shock propagation in network Acemoglu et al. (2012) provided a microfoundation for network centrality. The literature showed that an industry with a higher network centrality contributes relative more the economies value-added. Moreover, it highlighted that network centrality is related to microshocks. Finally, if comparative cost advantage and network centrality are similar, the latter is a simpler measure.

3.1. International trade network and network centrality

In the following, I outline definition of the trade network as directed and weighted network. The definition is based on Jackson (2010) and De Benedictis and Tajoli (2010). proceed by first defining the binary directed trade network and next the weighted trade network.

I define for each industry k a trade network of $N = 1, \ldots, n$ exporting countries i and N importing countries j. The countries in the network are the nodes. Each edge $g^k_{i,j}$ represents a trade relationship between an exporting industry k in country i and importing country j. Since the trade relationships in each trade network are not symmetric, the trade network is a directed network.

Specifically, I define a trade relationship as positive exports from industry k in country i to country j. Each edge is represented by $g^k_{i,j}$, which is equal to one if there are positive exports in industry k in country i to country j and zero otherwise. Thus, $g^k_{i,j} = \begin{cases} 1 & \text{if} \quad x^k_{i,j} \neq 0 \\ 0 & \text{if} \quad x^k_{i,j} = 0 \end{cases}$ For each industry k the edges $g^k_{i,j}$ are recorded in a symmetric matrix g^k of the dimensions $n \times n$. The binary trade network for each industry k is then the tuple of nodes and trade relationships $\mathcal{G}^k(N, g^k)$.

Extending the binary trade network to its weighted version, I introduce the weight variable $W^k_{i,j}$, which records the value of exports from industry k in country i to country j. The weight of an edge is thus denoted as the dollar value of the exports in a trade relationship. Analog to the matrix of trade relationships for each binary trade network, I define for each industry k the weight matrix W^k, which records the weight of each trading relationship.

The weighted trade network is a combination of each binary trade network and the corresponding weight matrix $ITN^k = (\mathcal{G}^k(N, g^k), W^k)$.

3.2. Network centrality

The outline of eigenvector network centrality is based on the textbook of Jackson (2010), who attributes the original mathematical exposition to Bonacich (1972). I first describe eigenvector centrality for a binary trade network. The concept extends without modifications to a directed and weighted network (Jackson 2010).

Intuitively, eigenvector centrality describes the idea that a node is more central if it is connected to other central nodes. The centrality of the other nodes is in turn determined by the centrality of the nodes are they are connected to. The eigenvector centrality $C^e_i(g^k)$ associated with the network g^k is proportional to the centrality of the nodes it is connected to.

$$\lambda C_i^e(g^k) = \sum_{j \neq i} g_{i,j}^k C_{i,j}^e(g^k)$$

, where λ is a proportionality constant. Restating the equation in matrix notation and solving it

$$\lambda C^e(g^k) = g^k C^e(g^k)$$
$$(I\lambda - g^k)C^e(g^k) = 0$$

where λ is the corresponding eigenvalue to the eigenvector $C^e(g^k)$. In general, this equation has n solutions. However, it is a convention to use the eigenvector corresponding to the largest eigenvalue. The Perron-Frobenius theorem implies an important property of the eigenvector. It states that for a non-negative column stochastic matrix [1], the right-hand eigenvector corresponding to the largest eigenvalue is positive and equivalently for a non-negative row stochastic matrix that the left-hand eigenvector for the largest eigenvalue is positive. Further, the theorem implies that if for some power the matrix g^k is positive, then the largest eigenvalue is equal to one and all other eigenvalues are smaller.

Extending the eigenvector centrality concept to the weighted and directed trade network is straightforward. Instead of computing the eigenvector for each adjacency matrix g^k, I calculate it for each weight matrix W^k. Moreover, take advantage of the Perron-Frobenius Theorem I row-normalize the weight matrix so that each cell is divided by the sum of exports of a country. Each cell of the matrix thus records the share of exports of the country in the exporting industry across destinations.

In the steps I outlined to construct, I described the weighted out-eigenvector centrality. However, it would be possible to instead column normalize the weight matrix and obtain the left eigenvector, which is the in-eigenvector centrality.

My analysis is based on the out-eigenvector, due to the following consideration. The out-eigenvector centrality is computed for the row-normalized matrix, where each cell records the export shares. A higher out centrality denotes that a country is exporting to countries with relative higher export shares. On the other hand, for the in-eigenvector, a higher centrality would describe that a country is importing from countries with high shares of imports. Therefore, I focus on the out-eigenvector centrality due to its similarity to the structural RCA.

3.3. Network centrality and structural Ricardian comparative advantage

In this subsection, I analyze the association between relative network centrality and structural RCA based on domestic value added export. I applied two normalizations to the eigenvector centrality as to the RCA before computing the association between them. Firstly, I normalised both measures to the grand mean, the mean across industries and the average across countries. Secondly, I normalized both measures relative to a benchmark country (the USA) and a benchmark industry (the food and beverages industry).

In the introduction, I outlined the similarity of network centrality and the RCA, which motivates the empirical analysis. Given the similarity, if structural RCA and network centrality show comparable results, the later may be preferred as a simpler measure.

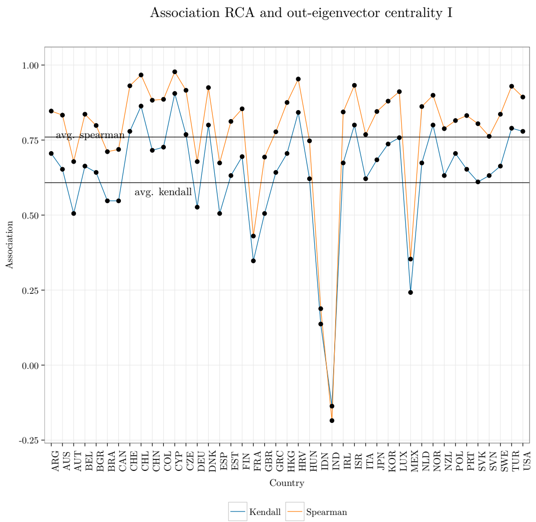
In the following, I first describe the left figure, which shows the association between network centrality and RCA according to the first normalization. First interpreting the left graph I see that the association of the rankings are high for most countries. An indicator of this is the average of Spearman's ρ 0.76 and Kendall's τ 0.58. Both China and the USA have a high similarity between the rankings. On the other hand, Finland and Mexico show high relative dissimilarity between the rankings. I obtain strong results for India and Indonesia. For India is that Spearman's ρ and Kendall's τ are negative. For Indonesia, the results show a small yet positive coefficients for both measures.

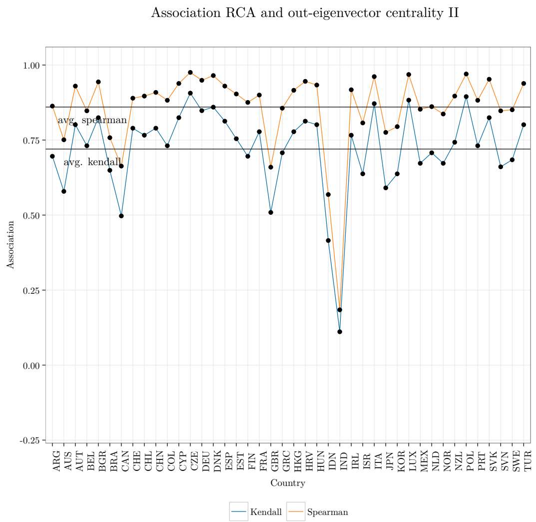
The right figure shows overall a higher association between RCA and network centrality. The average of Spearman ρ is 0.85 and Kendall's τ is 0.71. Like before the graph shows that India and Indonesia have the lowest similarity between both measures. However, the strength of association for India is slightly increased so that the

1. A column stochastic matrix is a matrix, where each column sum is equal to one. Analogue a row stochastic matrix has a row sum of one.

sign of the association is positive. Both Finland and Mexico show an increased similarity in the graph. Further, it is noteworthy that for Canada the strength of both associations is reduced, indicating that the rankings differ more for Canada.

Overall, the results suggest that the ranking of RCA and network centrality are similar for backward value-added exports. However, noteworthy for India and Indonesia I observed that both rankings are very different. Hence, future research should analyse the connection between RCA and network centrality from a theoretical perspective.





4. Conclusion

The objectives of this thesis were two-fold. First, assess whether the impact of IPF on the production process is such that traditional export measures no longer provide a reliable picture of technological comparative advantage. A second objective was to analyze the association between relative network centrality and technological comparative advantage. The hypothesis was motivated by the similarity of the interpretation of network centrality as how important an industry in a country in the export network is regarding $ to the stochastic interpretation of trade shares resulting from productivity draws.

To analyze technological comparative advantage, I used an structural RCA measure based on the methodology ofCostinot, Donaldson, and Komunjer (2011). The authors developed a theoretically consistent measure in a setup with imperfect specialization, multiple industries, and multiple countries. I constructed the measure based on forward and backward value-added exports and gross exports and compared the results using two association measures.

I proceeded in two steps to construct the structural RCA. In a first step I estimated a regression of the log of bilateral trade flows on the log of observed productivity, the inverse of international prices, an exporter-importer fixed effect and an importer-industry pair fixed effect. For the first step, I created a sample combining international relative price data from the GGDC Inklaar and Timmer (2014), R&D data from the ANBERD OECD database and gross exports and value-added data from the OECD-WTO (2015). In the second step, I regressed the log of bilateral trade flows on the full set of export-importer, importer-industry and export-industry fixed effects. For the second estimation, I constructed a sample, which the countries with observations on all three concepts.

In the first regression, I obtained an estimate for θ, which may be interpreted as the inverse of the cost dispersion. Comparing the point estimates of the cost dispersion parameter to the results in Costinot, Donaldson, and Komunjer (2011), I find that my point estimates were similar to one estimate of the authors. However, the favorite estimate of the authors was significantly lower. Moreover, the dispersion parameter showed similar estimates for backward value-added exports and gross exports. The estimates based on forward value-added were reduced compared to the other estimates.

Comparing the results of the structural RCA with gross exports and domestic value-added exports, I found that the simple and rank correlation coefficients showed very high coefficients. The result suggests that the sector-specific input and sourcing patterns are similar do not vary strongly across sectors and, therefore, cleaning gross exports of foreign value-added does not change the ranking significantly.

The second objective of this thesis was to analyze the empirical relation between relative network centrality and structural RCA. In network propagation of shocks literature, it was shown that for a single layer production network that network centrality is a first-order characteristic of how many actors contribute in $ terms to the network. This interpretation is similar to the stochastic interpretation of trade shares as a result of productivity draws. Hence, I analyzed the association between relative centrality and structural RCA. The results of a correlation analysis showed a stronger rank correlation than simple correlation and hence pointed out that the association is rather monotone than linear between both measures. In conclusion, the strong empirical correlation supports the hypothesis.

A direction for future work is to establish a theoretical model that explains the strong association of relative network centrality of an industry in a country in the international production network and structural RCA. Moreover, future work may analyze the association between a RCA ranking based on domestic value-added exports and the ranking predicted by the Heckscher-Ohlin model.

Bibliography

Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi. 2012. "The network origins of aggregate fluctuations." *Econometrica* 80 (5): 1977–2016.

Angrist, J. D., and A. B. Krueger. 2001. "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments." *Journal of Economic Perspectives* 15 (4): 69–85.

Angrist, J. D., and J.-S. Pischke. 2008. *Mostly harmless econometrics: An empiricist's companion.* Princeton university press.

Balassa, B. 1965. "Trade liberalisation and 'revealed' comparative advantage." *The Manchester School* 33 (2): 99–123.

Baldwin, R., and J. Lopez-Gonzalez. 2014. "Supply-chain Trade: A Portrait of Global Patterns and Several Testable Hypotheses." *The World Economy:* 1–40.

Bernard, A. B., and J. B. Jensen. 2004. "Why some firms export." *Review of Economics and Statistics* 86 (2): 561–569.

Bhaskaran, K., and L. Smeeth. 2014. "What is the difference between missing completely at random and missing at random?" *International Journal of Epidemiology* 43 (4): 1336–1339.

Bonacich, P. 1972. "Factoring and weighting approaches to status scores and clique identification." *The Journal of Mathematical Sociology* 2 (1): 113–120.

Bound, J., D. Jaeger, and R. Baker. 1993. "The Cure Can Be Worse than the Disease: A Cautionary Tale Regarding Instrumental Variables." *NBER Working Papers,* no. 0137.

Cameron, A., and P. Trivedi. 2005. *Microeconometrics: Methods and Applications.* Cambridge: Cambridge University Press.

Combes, P.-P., T. Mayer, and J.-F. Thisse. 2008. *Economic geography: The integration of regions and nations.* Princeton, N.J.: Princeton University Press.

Costinot, A., D. Donaldson, and I. Komunjer. 2011. "What goods do countries trade? A quantitative exploration of Ricardoś ideas." *The Review of Economic Studies* 79:581–608.

Daudin, G., C. Rifflart, and D. Schweisguth. 2011. "Who produces for whom in the world economy?" *Canadian Journal of Economics/Revue canadienne d'économique* 44 (4): 1403–1437.

De Benedictis, L., and L. Tajoli. 2010. "Comparing sectoral international trade networks." *Aussenwirtschaft* 65 (2): 53–73.

Dhaene, G. 2014. *Advanced Econometrics.* Leuven: Leuven Ekonomika.

Douglas Staiger, J. H. S. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65 (3): 557–586.

Eaton, J., and S. S. Kortum. 2002. "Technology, geography, and trade." *Econometrica* 70 (5): 1741–1779.

Feenstra, R. C. 1998. "Integration of Trade and Disintegration of Production in the Global Economy." *Journal of Economic Perspectives* 12 (4): 31–50.

Feenstra, R. C., A. Heston, M. P. Timmer, and H. Deng. 2009. "Estimating Real Production and Expenditures across Nations: A Proposal for Improving the Penn World Tables." *Review of Economics and Statistics* 91 (1): 201–212.

Fredricks, G. A., and R. B. Nelsen. 2007. "On the relationship between Spearman's rho and Kendall's tau for pairs of continuous random variables." *Journal of Statistical Planning and Inference* 137 (7): 2143–2150.

Griffith, R., S. Redding, and J. V. Reenen. 2004. "Mapping the Two Faces of R&D: Productivity Growth in a Panel of OECD Industries." *The Review of Economics and Statistics* 86, no. 4 (November): 883–895.

Hausman, J. A. 1978. "Specification tests in econometrics." *Econometrica: Journal of the Econometric Society:* 1251–1271.

Helpman, E., and P. R. Krugman. 1985. *Market structure and foreign trade: Increasing returns, imperfect competition, and the international economy.* Cambridge, Mass.: MIT press.

Hinloopen, J., and C. Van Marrewijk. 2001. "On the empirical distribution of the Balassa index." *Weltwirtschaftliches Archiv* 137 (1): 1–35.

Imbens, G., and J. Wooldridge. 2007. *What is New in Econometrics? Lecture Notes 12, Missing Data.*

Inklaar, R., and M. P. Timmer. 2014. "The Relative Price of Services." *Review of Income and Wealth* 60 (4): 727–746.

Jackson, M. O. 2010. *Social and Economic Networks.* Princeton University Press.

Johnson, R. C. 2014. "Five facts about value-added exports and implications for macroeconomics and trade research." *The Journal of Economic Perspectives:* 119–142.

Johnson, R. C., and G. Noguera. 2012. *Fragmentation and Trade in Value Added over Four Decades.* Working Paper, Working Paper Series 18186. National Bureau of Economic Research.

Koopman, R., Z. Wang, and S.-J. Wei. 2014. "Tracing Value-Added and Double Counting in Gross Exports." *The American Economic Review* 104 (2): 459–494.

Leontief, W. W. 1936. "Quantitative input and output relations in the economic systems of the United States." *The review of economic statistics:* 105–125.

Leromain, E., and G. Orefice. 2014. "New revealed comparative advantage index: dataset and empirical distribution." *International Economics* 139:48–70.

Little, R. J. A., and D. B. Rubin. 2002. *Statistical Analysis with Missing Data.* Second Edition. Wiley series in probability and statistics. John Wiley & Sons, Inc.

Moons, K. G. M., R. A. R. T. Donders, T. Stijnen, and F. E. J. Harrell. 2006. "Using the outcome for imputation of missing predictor values was preferred." *Journal of clinical epidemiology* 59, no. 10 (October): 1092–1101.

Morris, T. P., I. R. White, and P. Royston. 2014. "Tuning multiple imputation by predictive mean matching and local residual draws." *BMC Medical Research Methodology* 14 (1): 1–13.

OECD. 2004. "Most-Favoured-Nation Treatment in International Investment Law."

———. 2011. *STAN Database for Structural Analysis.*

———. 2013. *STAN R & D expenditures in Industry.*

OECD-WTO. 2015. *Trade in Value Added (TiVA).*

Organization of the Petroleum Exporting Countries. 2008. *Annual Statistical Bulletin.*

Rubin, D. B. 1978. "Multiple imputations in sample surveys-a phenomenological Bayesian approach to nonresponse." In *Proceedings of the survey research methods section of the American Statistical Association. Vol. 1.*

———. 1986. "Statistical Matching Using File Concatenation with Adjusted Weights and Multiple Imputations." *Journal of Business & Economic Statistics* 4 (1): 87–94.

———. 1987. *Multiple Imputation for Nonresponse in Surveys.* 99th ed. Wiley.

Schafer, J. L., and M. K. Olsen. 1998. "Multiple imputation for multivariate missing-data problems: A data analyst's perspective." *Multivariate behavioral research* 33 (4): 545–571.

Silva, J. M. C. S., and S. Tenreyro. 2006. "The log of gravity." *The Review of Economics and statistics* 88 (4): 641–658.

Timmer, M. P. 2012. "The world input-output database (WIOD): contents, sources and methods." *WIOD Working Paper No. 10.*

Timmer, M. P., A. A. Erumban, B. Los, R. Stehrer, and G. J. de Vries. 2014. "Slicing up global value chains." *The Journal of Economic Perspectives:* 99–118.

Wang, Z., S.-J. Wei, and K. Zhu. 2013. "Quantifying International Production Sharing at the Bilateral and Sector Levels." *NBER Working Papers,* no. w19677.

White, I. R., P. Royston, and A. M. Wood. 2011. "Multiple imputation using chained equations: Issues and guidance for practice." *Statistics in Medicine* 30 (4): 377–399.

Yeats, A. 1985. "On the appropriate interpretation of the revealed comparative advantage index: Implications of a methodology based on industry sector analysis." *Weltwirtschaftliches Archiv* 121 (1): 61–73.

A. Appendix

A.1. Decomposition of gross exports into value-added exports

I use the indicators from the OECD-TiVA database later to compute the RCA ranking based on the two step method described in Costinot, Donaldson, and Komunjer (2011) and compare the rankings to an RCA ranking based on gross exports.

Input-Output tables are models of the economy based on the work of Leontief (1936). The intuitive insight of Leontief was to record the usage of intermediate inputs and production factors to produce units of output in a matrix. He showed that the recursion can be mathematically solved until one accounts for the complete set of intermediate inputs to produce one unit of output. Leontief's insight of modeling input-output relations is sufficient to decompose gross exports into domestic value-added exports (Koopman, Wang, and Wei 2014).

At this point, a note on the data limitations concerning Input-Output tables is necessary. Ideally value-added exports data would be decomposed from global Input-Output tables provided by national statistical agencies. Yet, global input-output data do not exist, and therefore scientist and different international organizations construct synthetic global Input-Output tables based on National Input-Output tables (Johnson 2014).

In the following I will illustrate the decomposition of gross exports into forward and backward vale-added exports for two simple cases. Interested readers may be referred to the work of Koopman, Wang, and Wei (2014) for a more general treatment.

A.2. ISIC and ISO 3 Alpha Classification

ISIC Code	Short	TiVA Description
01-05	Agriculture	Agriculture, hunting, forestry and fishing
10-14	Mining	Mining and quarrying
15-16	Food	Food products, beverages and tobacco
17-19	Textiles	Textiles, textile products, leather and footwear
20	Wood	Wood and products of wood and cork
21-22	Paper	Pulp, paper, paper products, printing and publishing
23	Fuel	Coke, refined petroleum products and nuclear fuel
24	Chemicals	Chemicals and chemical products
25	Plastic	Rubber and plastics products
26	Minerals	Other non-metallic mineral products
27-28	Metals	Basic metals and fabricated metal products
29	Machinery	Machinery and equipment, nec
30-33	Electrical	Electrical and optical equipment
34-35	Transport	Transport equipment
36-37	Misc. Manufacturing	Manufacturing nec; recycling
40-41	Electricity	Electricity, gas and water supply
45	Construction	Construction
50-52	Trade	Wholesale and retail trade; repairs
55	Gastronomy	Hotels and restaurants
60-64	Communication	Transport and storage, post and telecommunication
65-67	Finance	Financial intermediation
70-74	Real estate	Real estate, renting and business activities
75-95	Social	Community, social and personal services

Table A.1.: ISIC Revision 3.1

COU	Country	COU	Country
ARG	Argentina	ITA	Italy
AUS	Australia	JPN	Japan
AUT	Austria	KOR	Korea
BEL	Belgium	LTU	Lituhania
BGR	Bulgaria	LUX	Luxembourg
BRA	Brazil	LVA	Latvia
CAN	Canada	MEX	Mexico
CHE	Switzerland	MYS	Malaysia
CHL	Chile	NLD	Netherlands
CHN	China	NOR	Norway
COL	Colombia	NZL	New Zeeland
CYP	Cyprus	PHL	Philippiens
CZE	Czech Republic	POL	Poland
DEU	Germany	PRT	Portugal
DNK	Denmark	ROU	Romania
ESP	Spain	ROW	Rest of the World
EST	Estonia	RUS	Russian Federation
FIN	Finland	SGP	Singapore
FRA	France	SVK	Slovakia
GBR	United Kingdom	SVN	Slovenia
GRC	Greece	SWE	Sweden
HKG	Hong Kong	THA	Thailand
HRV	Croatia	TUN	Tunisia
HUN	Hungary	TUR	Turkey
IDN	India	TWN	Taiwan
IND	Indonesia	USA	United States of America
IRL	Ireland	VNM	Vietnam
ISR	Israel	ZAF	South Africa

Table A.2.: ISO 3 Alpha Code

A.3. Data Appendix

A.3.1. Sample Statistics

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
Log Backward Value Added Exports	2.44	−4.60517	10.75	2.87	17453
Log Gross Exports	2.74	2.87	−4.60517	11.11	17505
Log Forward Value-Added Exports	3	2.35	−4.605	10.74	15999
Log Productivity	0.27	0.27	−.6715387	1.17	18444
Log R&D	17.80	2.44	10.74	24.76	17313

Table A.4.: Summary statistics in estimation sample

Variables	Log gross exports	Log backward value-added exports	Log forward value-added exports	Log Productivity	Log R&
Log gross exports	1.00				
Log backward value-added exports	1.00	1.00			
Log forward value-added exports	0.87	0.89	1.00		
Log productivity	−0.092	−0.100	−0.211	1.00	
log R&D	0.43	0.45	0.49	−0.200	1.0

Table A.5.: Pairwise correlation in estimation sample

IND	N	IND	N
01-05	841	10-14	841
15-16	841	17-19	841
20	841	21-22	841
23	841	24	841
25	841	26	841
27-28	841	29	841
30-33	841	34-35	841
36-37	841	45	841
50-52	841	55	841
60-64	841	65-67	841
70-74	841	75-95	841
Total	18502		

Table A.6.: N. obs / IND in estimation sample

COU	N	COU	N
AUS	572	AUT	572
BEL	572	CAN	572
CZE	572	DEU	572
ESP	572	EST	572
FIN	572	FRA	572
GBR	572	GRC	572
HUN	572	IRL	572
ITA	572	JPN	572
KOR	572	LUX	572
MEX	572	NLD	572
POL	572	PRT	572
SVK	572	SVN	572
TUR	572	USA	572
Total	114872		

Table A.8.: N. obs / COU in estimation sample

COU	N	COU	N
ARG	946	AUS	946
AUT	946	BEL	946
BGR	946	BRA	946
CAN	946	CHE	946
CHL	946	CHN	946
COL	946	CYP	946
CZE	946	DEU	946
DNK	946	ESP	946
EST	946	FIN	946
FRA	946	GBR	946
GRC	946	HKG	946
HRV	946	HUN	946
IDN	946	IND	946
IRL	946	ISR	946
ITA	946	JPN	946
KOR	946	LUX	946
MEX	946	NLD	946
NOR	946	NZL	946
POL	946	PRT	946
SVK	946	SVN	946
SWE	946	TUR	946
USA	946	Total	40678

Table A.10.: N. obs / COU in structural RCA sample

IND	N	IND	N
01-05	1849	10-14	1849
15-16	1849	17-19	1849
20	1849	21-22	1849
23	1849	24	1849
25	1849	26	1849
27-28	1849	29	1849
30-33	1849	34-35	1849
36-37	1849	45	1849
50-52	1849	55	1849
60-64	1849	65-67	1849
70-74	1849	75-95	1849
Total	40678		

Table A.12.: N. obs / IND in structural RCA sample

A.3.2. First Stage

	(1) Full Sample	(2) Without primary industries	(3) Without primary industries high[1]
Log of R&D	0.02^{***}	0.02^{***}	0.02^{***}
	(0.002)	(.0020154)	(.0022056)
Exporter Importer FE	Yes	Yes	Yes
Importer Industry FE	Yes	Yes	Yes
N	19343	17661	15283
F (excluding dummies)	125.60	88.17	85.24
Imputations	29	29	29

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$