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 Economics Research

Promotor: Dr. Liza Archanskaia

Academic year: 2015/2016

Contents

Abbreviations

ANBERD	Analytical Business Enterprise Research and Development
Backw.	Backward
F.O.B.	Free on Board
Forw.	Forward
G7	Group of Seven Leading Industry Nations
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, helping 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). Feenstra (1998) characterized IPF as the break down the vertically-integrated production process. The vertical integration production process decribes that an industry produces the output good in several production steps, whereas with IPF firms increasingly decide to source production steps. Moreover, Baldwin and Lopez-Gonzalez (2014) described IPF as an increase of complex trade flows of labor, human capital, and investment between countries.

The time evolution of IPF was studied by Johnson and Noguera (2012). They report that IPF started to increase in the 1970s, stagnated in the following decade and accelerated very strongly since the 1990s. Especially, they point out that IPF increased by threefold since the 1990s compared to the pre-1990s. They see this as an indication of acceleration. Further, evidence about the time evolution of IPF is documented by Timmer et al. (2014). The authors compare IPF in 1995 to 2005 by the difference of foreign content of goods for 560 products. The authors findings show that for 86 % of the products the foreign content increased. Moreover, Baldwin and Lopez-Gonzalez (2014) showed for the same time period that the final goods share of exports declined for all 16 manufacturing sectors, which they interpret as a sign of increased IPF.

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** critize 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**). 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 addition, **Leromain** noted in their analysis that the distribution of the BI has poor time stability. In contrast, **Leromain** found good empirical quality 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. Therefore I conclude that the structural RCA of Costinot, Donaldson, and Komunjer (2011) is empirically and theoretical reasons the more suitable measure to analyze RCA.

1. Introduction

The analysis conducted in this thesis showed that backward value-added exports do not substantially alter the estimates of the cost dispersion parameter or the structural RCA ranking compared to gross exprots. The assosciation of the rankings showed a very strong relationship and the coefficients were close to one. Further, I find that my results are robust to changes in the country coverages. However, the results for forward value-added exports The results are in contrast to the finding of Koopman, Wang, and Wei (2014).

A second objective of the thesis is to link the structural RCA to relative 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. In their setup of one production factor (labor) with industries using a Cobb-Douglas production technology, stochastic productivity shocks and consumers with one labor unit and Cobb-Douglas preferences they showed that network centrality is the first-order characteristic of an industry in one layer production network. In this model network centrality describes how much an industry contributes to a production network concerning $. Similar, the stochastic interpretation of trade shares states 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.

I find that both measures are strongly associated and that the result is robust to changes in the country coverages.

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 an 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 briefly describe the Ricardo model of Costinot, Donaldson, and Komunjer (2011). The contribution of the model is that it shows that relative cost differences predict the pattern of trade specialization at the country-industry level in a multiple country and multiple industries framework. Further, of more interest here, is the structural Ricardian comparative advantage measure developed in the model. I use this measure to compare RCA rankings advantage for value-added exports and gross exports.

The structure of the chapter is as follows, first I discuss the indicator of value-added exports. Next, I describe the construction of value-added exports and the sample. Further, I describe the empirical methods used. Then, I turn to the estimation of the productivity dispersion parameter and the exporter-industry fixed effects. In the last step I use correlation analysis to assess the association structural RCA for the different export measures.

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

The literature on value-added exports is motivated by the international production fragmentation (IPF). As a consequence of IPF intermediate inputs cross country boarders frequently in the production process and therefore goods are increasingly double counted in gross exports. This implies means that gross exports overestimate the domestic value-added content of goods (Johnson 2014). Another consequence of IPF is that the contribution to the production of goods by different countries in terms of value-added may be included in the gross exports of another country. Hence, gross exports overestimates and does not correctly attributes value-added. As gross exports are not able to answer the question "who is producing for whom" **daudin**, this motivated the development of new measurements.

The pioneer work of Hummels, Ishii, and Yi (2001) (HYI) defined the statistical indicators of vertical specialization to study IPF. According to HYI a country may engage in vertical specialization in two ways, either importing intermediate inputs to produce goods for export or by exporting intermediate inputs, which are subsequently used as inputs in the production of exports by other countries. Koopman, Wang, and Wei (2014) criticized this indicator of vertical trade for two reasons. First, it assumes that the amount of imported inputs used in domestic and exports industry is the same and certain types of exports violate this assumption. Second, the concept assumes that imports are completely sourced abroad. The authors argued that this assumption no longer necessarily holds for more than two countries. Therefore, they reasoned that a different indicator value-added exports is better suited to study IPF, as it does not include any double counting and correctly attributes value-added in the production process to .

At the country level the framework of Koopman, Wang, and Wei (2014) showed how to decompose gross exports into value-added. However, at the industry level decomposing gross exports into value-added exports leads to two different perspectives, the backward linkages perspective and the forward linkages perspectiveWang, Wei, and Zhu (2013). I will outline the different perspectives in the following and In the appendix I describe the math behind the decomposition for the 2 country and 2 goods case.

Backward linkage value-added exports of an industry include the direct domestic value-added of that industry and other upstream domestic industries in the gross exports of the exporting industry. This perspective is based on the destination 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, which is either directly or indirectly through other industry is 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). First a backward-linkages based view is useful to understand a country's domestic value added which is embodied in it's exports. In the context of RCA, the domestic value-added in gross exports, is consistent with a production based RCA, since it measures the 'total domestic factor content in exports' (p.490)Koopman note.

Second, Wang, Wei, and Zhu (2013) describe that the forward linkages perspective on value-added exports is helpful to understand how much value-added a given sector contributes to a country's exports. This indicator correctly attributes how much value-added of an industry is directly and indirectly through further downstream industries is exported across destinations. The RCA ranking with this indicator shows how efficiently an industry uses the domestic factors of production **bladwin**.

To obtain data on value-added exports various sources exist. Previous literature on value added trade mainly focused on the WIOD Timmer (2012) database. In this thesis I use the value-added export data from the TiVA OECD-WTO (2015) database. This choice is motivated at first that the estimation of the structural RCA requires matching different databases, which are based on a industry classification similar to the ISIC classification, which is used in TiVA. Further, from a substantial point TiVA data provides a larger country coverage with a more regional diverse focus. In addition, as a Johnson (2014) showed only one author has employed the TiVA dat.

Further, data however is necessary to be able to estimate the structural RCA indicator. In a first step Costinot, Donaldson, and Komunjer (2011) used a regression of gross exports on the inverse of producer prices instrumented by R&D to obtain an estimate of an dispersion parameter. To follow their approach I combine the (value-added) exports data of OECD-WTO (2015) TiVA, with R &D expenditure from OECD (2013) ANBERD and international producer price data from the GGDC (Inklaar and Timmer 2014). The sample I obtain by combining this sources is the estimation sample. It includes twenty nine countries and twenty two industries at the ISIC Rev.3 level. At the second stage I require only the value-added exports data of the TiVA. This larger dataset includes the twenty two industries at the ISIC Rev.3 classification and fifty six countries.

Further to create the first sample several data managements steps were necessary. To reconcile the industry coverage of TiVA and the GGDC international price data I aggregated the international price data using the ISIC Rev. 3.1 two digits classification. The data management steps enabled me to extend the sample to include a larger proportion of service sectors. In practice, I aggregated the prices using a weighted average. I specified as the weights the relative share of the industries in terms of its sector value-added share with country industry data from the OECD (2011) STAN database. Further, in case of missing data on value-added output data I aggregated the prices by a simple average. The resulting cross-section sample spans 22 industries and 29 countries for the year 2005. Moreover in the appendix in table A.4 I report the descriptive statistics of the sample.

Further, compared to the full country coverage of the TiVA November release I excluded some countries because they had no exports to any destination recorded in at least one sector [1]. Further, I excluded Saudi Arabia because its exports mainly consist of oil [2] (Organization of the Petroleum Exporting Countries 2008). Moreover, fifteen countries did not have records on forward linkages value-added exports and were therefore as well omitted from the sample [3]. Thus, the fixed-effects sample includes 22 industries and 43 countries for the year 2005.

1. Island, Costa Rica, Brunei Darussalam
2. 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)
3. Lithuania Latvia, Malta, Malaysia, Philippines, Romania, Rest of the World, Russia, Singapore, Thailand, Tunisia, Taiwan, Vietnam, South Africa, Costa Rica, Brunai Darussalem, Khambodia, Island

2.2. Ricardian model

In this subchapter I describe the major assumptions of the Ricardo model by (**Costinot**), which derives the first estimation equation to obtain the fundamental producitivty.

In general, the model considers a world economy of $i = 1, \dots, n$ countries and $k = 1, \dots, K$ industries. Further, there is only one factor of production, which is labour. Labour is perfectly mobile across industries and immobile between countries. L_i denotes the number of workers in each country i and w_i denotes their wage.

The authors model the production technology as follows. Each industry produces a good with a constant returns to scale technology. Further, each good there are indefinitely many varieties $\omega \in \Omega$. Furhter, they assume that stochastic productivity differences. The fundamental productivity $z_i^k(\omega)$, which denotes how much of a variety ω may be produced with one unit of labor, is for each triple of 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 of the following form. Of every 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 for a third country l importing a good k from country i through another country j is more costly than directly importing it. Thus $d_{i,l}^k \leq d_{i,j}^k d_{j,l}^k$ for any third country l.

The first assumption It is assumed that the market structure is perfect competition. The consumers therefore pay a price equal to the lowest combination of production and transport cost.

$$p_j^k(\omega) = \min_{1 \leq i \leq I} c_{i,j}^k(\omega)$$

with the unit production cost are equal to $c_{i,j}^k = \frac{d_{i,j}^k w_i}{z_i^k}$.

Moreover, the consumer preferences are modeled 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 consumer show a 'love-for-variety' property. The consumer welfare with this utility function 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 assumed structure of consumer preferences implies the following relation for the total expenditure of any country j on a variety ω of a good k.

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

Therefore the expenditure of a destination on a good 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

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

Therefore 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. Therefore the model makes Ricardian prediction at the industry level.

2.3. Empirical predictions

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

Costinot, Donaldson, and Komunjer (2011) showed that from assumption 1 the ratio of observed productivity $\tilde{z}_i^k/\tilde{z}_{i'}^k$ for a country pair i and i' links directly to the ratio of fundamental productivity.

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

The relation above links openness corrected exports $\tilde{x}_{i,j}^k$ to the observed productivity and trade cost. The first ratio of $\frac{\tilde{x}_{i,j}^k}{\tilde{x}_{i',j}^{k'}}$ accounts for income differences Y_j of the importing countries and wage differences w_i across exporting countries. Further, the second ratio $(\frac{\tilde{x}_{i,j}^k}{\tilde{x}_{i'j}^{k'}})/\frac{tildex_{i'j}^k}{\tilde{x}_{i'j}^{k'}})$ accounts for differences in the expenditure shares α^k

The productivity term in eq. (2.4) is net of specific trade barriers $\delta_{i,j}$ between country i and j like distance and net 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 and other components.

Further, to allow the estimation of the equation above the measure of observed productivity has to be specified. In their model setup the authors showed that the observed ratio of relative productivity is fully reflected by the inverse ratio of producer price indices. This result holds in the Ricardo world, however in the context of international product fragmentation, producer prices as well include foreign contributions.

According to **Costinot** following econometrically equivalent equation may be estimated instead of the previous equation.

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

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) interpret the equation as an analogue to a 'difference-in-difference' estimation. The productivity term in eq. (2.4) is that the first difference 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 [5]. Further the error term $\epsilon_{i,j}^k$ includes variable trade cost and other unobserved time-varrying components.

Further, structural RCA measure is obtained as follows. In the first step I estimate θ and in a second step I obtain the equation with the full set of fixed effects.

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

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

The δ_i^k in the second relation captures the effect of $\theta \ln z_i^k$ on the bilateral gross exports.

$$z_i^k = e^{\delta_i^k/\theta}$$

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).

5. 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 states 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).

2.4. Generalization

In the following, I discuss the effect of sector-specific use of production factors in the cost function. Moreover, I include capital and intermeidate inputs as production factor and I will argue that the effects of sector-specifc use of production factors are similar to sector-specific international sourcing of inputs. I regard both aspects of production as an effect of IPF. The effects of IPF, as generalizing the cost function, are that the ranking of RCA would not only reflect productivity differences but as well sector specfic factor usage and sourcing patterns. Therefore, I argue that differences of the RCA ranking for value-added exports or gross exports would reflect the effect of IPF.

In the following I introduce international sourcing and sector-specific production factors based on the cost function in Shikher (2011).

$$c_{i,j}^k = \frac{d_{i,j}^k}{z_i^k Y_j} w_i^{\alpha^k} r_i^{\beta^k} \rho_i^{1-\alpha^k-\beta^k}$$

. Further, I assume that the industries mix intermediate inputs in fixed proportions. The price of inputs ρ_i is therefore a Cobb-Douglas function of industry prices:

$$\rho_i = \prod_{m=1}^{K} p_i^{\eta_{i,m}}$$

where $\eta_{i,m} \geq 0$ is the share of industry m goods in the intermediate inputs of industry k, such that $\sum_{m=1}^{K} \eta_{i,m} = 1, \forall i$. For this more general cost function the bilateral trade flows would know depend on the production factor usage in the industries and the usage of input factor prices. The ranking of countries based on the model specific productivity factor would now be confunded by the effects of different factor endowments.

From this general cost function one can still arrive at the relation in theorem if one assumes that the share of production factor used in the production of a good is not industry specific. Therefore the cost function simplifies as follows

$$c_{i,j}^k = \frac{d_{i,j}^k}{z_i^k} w_i^{\alpha} r_i^{\beta} \rho^{1-\alpha_i-\beta_i}$$

. where $\rho_i = \prod_{m=1}^{K} p_i^{\eta_{i,m}}$ It is easy to see that the cost function together with eq. (1) and assumption 1 would still imply eq. (2).

Therefore I can interpret the hypothesis that value-added exports affect the RCA ranking, as a test whether the factor shares of the inputs and other production factors are sector specific. Further, it is easy to extend the cost function to include international sourcing of inputs by introducing another subscript to denote whether the input production factor is sourced domestically or abroad. Similar to the previous argument about sector specific production factors, the international sourcing pattern would confound the picture of RCA rankings.

2.5. Estimation method

In this section I discuss the estimation of the dispersion parameter θ. The estimation of θ and the industry exporter fixed effect are the two necessary steps to construct the structural RCA measure. Furthermore, two important aspects of the estimation I discuss are the motivation of the instrumental variable (IV) estimation and the imputation of the missing data using multiple imputations. Finally I describe the construction of the sample.

The OLS estimate of θ is unbiased and consistent if the independent variable is uncorrelated with the error term, which includes variable trade cost and other unobserved time varying variables. However, Costinot, Donaldson, and Komunjer (2011) highlight two reasons, which motivate the IV estimation. The first reason is a simultaneity bias. The simultaneity bias arises if agglomeration effects [6] are present. The sign of the bias is a priori ambiguous (Costinot, Donaldson, and Komunjer 2011). The second reason is a measurement error. It is a source of bias if the measurement error is correlated with the true underlying variable. As a consequence of the measurement error, the estimate would be biased towards zero (Greene 2003, p.85). Both biases cause an endogeneity problem, which may be solved using an IV estimation (Dhaene 2014, p.139). A final motivation for the IV estimation is that under a more general cost function the inverse of producer prices may reflect other sources of comparative advantage than productivity differences. Using the variation of the observed producer prices only related to a instrument may lead to a better identification of the effects of productivity on gross exports.

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

I chose to instrument the producer prices with R& D expenditures as in Costinot, Donaldson, and Komunjer (2011) and Eaton and Kortum (2002). The relationship I hypothesize is that increased R & D expenditures increases the productivity of an industry. Such an relationship between productivity of industries and R&D is e.g. hypothesized in Griffith, Redding, and Reenen (2004). Increased productivity lowers the production cost of an exporter and is postively related to exports according to the Ricardo model of this chapter. According to this relationship, I expect that in the first stage the coefficient of R& D expenditures is positive and statistical significant. Moreover, under this hypothesis R&D as an instrument would also satisfy the second assumption, which can not be empirically tested (()p.109)cameron2009.

If an instrument violates the first assumption, it is termed a weak instrument. An issue arising from an weak instruments is that the IV estimator does not correctly identify the causal effect of the endogenous variable (Bound, Jaeger, and Baker 1993). Moreover, if a weak instrument is weakly correlation with the error term, the IV estimate becomes inconsistent (Bound, Jaeger, and Baker 1993). A test statistic to assess whether an instruments is weak, is 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 10 (Douglas Staiger 1997).

2.5.1. Missing data imputation

In the data set, I have missing data in the variables, R & D and producer prices. The 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). Further, in the IV estimation missing values of the instrument may reduce the strength of the first stage association between R&D and the inverse of producer prices. The IV estimates, which is the ratio of the reduced form and the first stage, would therefore show upward biased estimates. In the following I motivate and describe the employed missing data technique.

To impute the missing data I used the method of multiple imputation, which is a bayesian technique to impute missing data by simulated draws from a predictive posterior distribution. It was initally proposed in Rubin (1978) for nonresponse in surveys and the bayesian theory of multiple imputation has been shown in Rubin (1987). I chose the multiple imputations technique for mainly three reasons. Firstly, techniques ignoring the missing observations in the analysis as e.g complete case methods or case-wise deletion require stronger assumptions than imputation.

6. An example of agglomerations effects is e.g. if there are positive spillover from a spacial close exporting firm on another firms exports (Bernard and Jensen 2004).

Valid and unbiased statistical results using these technique are obtained only under the strong assumptions that the missing data is a random subset of all observations (Bhaskaran and Smeeth 2014). In the analysis I would therefore assume that the probability of missing values in the producer prices or R&D are independent of the observed data. Secondly, single imputation methods have the draw back that they do not take into account the uncertainty induced by the missing values. As a result the estimated variance of a single imputation would be downward biased (Imbens and Wooldridge 2007) . Thirdly, multiple imputation offers 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 mulitple imputation based on Little and Rubin (2002, p.209-211). To start with I assume that the indicators of missing values are random variables with a distribution described by a set of paramters κ. Additionally, I assume that the probability of missing observations depends only on observed data and is independent of unobserved data. 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 values from the joint posterior distrbution of the missing data and the observed data and in turn drawing the parameters of the distrbution from the coresponding complete data posterior. Importantly, the authors show that the mean of the posterior distrubution can be approximated by the simulated draws $ka\bar{p}pa = 1/M \sum_m = 1^M * \kappa$ and the variance can be expressed by the average of the simulated draws of the variance and the between imputations variance $Var(\kappa|X) = \sqrt{\bar{V} + 1/(m-1) * Bm\, whereB} = 1/M - 1 * \sum_m = 1^M (\kappa_m - ka\bar{p}pa_m)^2$ and $\bar{V} = 1/M * sum_m = 1^M V_m$ with$V_M = VarX_obs, X_mis^(d)$. Further, the flexibility of the multiple imputation method allows to use different models for the imputation and the analysis Little and Rubin (2002, p.217). 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. Therefore multiple imputation can be regarded as three step procedure. In a first step the data set is imputed using simulated draws and in the second step complete-data methods can be conducted on the M data sets and in the third step the results are pooled.

Another multivariate imputation method, which can be used with multiple imputation, is the fully conditional specification. This approach, which is also known as sequential regression or chained equaitions, splits the problem of imputing a joint distribution into several univariate problems Van Buuren (2007). I choose the fully conditional specification (FCS) due to the flexibility. Further, using FCS I specify the imputation model at the variable level, which allows a more credible imputation Buuren and Oudshoorn (2000). A draw back of the FCS is that the method is theoretically not well understood (Van Buuren (2007)). However, several simulation studies indicate that for a wide range of applications the approach yields unbiased estimates (Van Buuren (2007),Van Buuren et al. (2006)).

The description of the approach is founded on the outlines in Morris, White, and Royston (2014) and Royston and White (2011). In the inital step of FCS each variable with missing values is regressed on each other. In the next step the missing values are replaced by draws from the predictive posterior distribution and the initial step is repeated. The two steps are called a cycle. Usually, several cycles are conducted for each of the M imputations until the algorithm converged to the implicit joint distribution (Van Buuren (2007)).

Moreover, I combine the FCS with predictive mean matching (PMM). PMM is a nearest neighbor matching technique suggested by Rubin (1986). In the framework of MI the PMM technique replaces the missing values with draws from the observed data instead of draws from a posterior distribution. As a consequence, the distribution of the imputed variables closely resembles the observed variables. The imputation method may be used to impute skewed variables for which normality assumptions are not tenable (White, Royston, and Wood 2011). Moreover, the PMM technique showed good results in the simulation study of Morris, White, and Royston (2014).

In a first step each missing variable is regressed on the imputation model and other variables with missing values. From the regression I obtain a set of parameters β and responding variances V. Further, a simulated β^* is than obtained from a multivariate normal distrubiton The imputation of the missing values is based on a random draw from the q [7] closest observations, which minimize the distance between the observed parameter multiplied by the covariates $\beta * x_h \quad h = 1, \ldots n_obs$ and the simulated parameter multiplied by the covariates $\beta^* * x_p \quad p = 1, \ldots n_obs$. In the following step, the q observations with the smallest difference between the β and

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

β^* are identified and from the q closest observations one observations is randomly chosen as imputed value.

I specified the imputation model as follows. I decided to impute the R& D and producer prices since the focus of the imputation is the analysis model of the first stage regression. Further, a simulation study of Moons et al. (2006) suggests that multiple imputation using the outcome leads to unbiased results whereas ignoring the outcome leads to biased results. Specifically, the imputation for both the log of R& D and the log of the inverse of producer prices are based on the country and industry fixed effects dummies. In the case here, I argue that the assumption of ignorable missigness is plausible given that I specify country and industry fixed effect as covaraites to account for unobserved variables. The imputation model is motivated by a similar specification a regression imputation for the log of R&D in Costinot, Donaldson, and Komunjer (2011).

After the multiple imputation the complete-data techniques can be used on each of the m data sets. The m results of this step are pooled using Rubins Rules Rubin (1987, p.77). These rules suggest that to obtain parameter estimate of θ I will take the simple average, therefore $\theta = 1/m \sum_{imp=1}^{m} \theta^{imp}$. Further, the standard error is computed as follows $SE = \sqrt{W + 1/(m-1) * B}$. Moreover, I apply the Fisher transformation to the m R-squared and then average them, and reverse the transformation after averaging. This step is suggested to improve the asymptotic normality, as Rubins Rules are obtained for asymptotically normal distributed variables.

2.5.2. Sample

I have created two samples for the estimation the RCA. The first smaller estimation sample, constis of the international price data from the GGDC (Inklaar and Timmer 2014), the value-added exports and gross exports data from the OECD-WTO 2015 database, and R &D expenditure data from the OECD (2013) ANBRED. To construct the sample I merged the data sources using the ISIC Rev.3.1 two digits classification. Especially, I to allow merging of the international price data I aggregated several service industries using a weighted average. I chose the weights to be equal to the sectoral share of value added output constructed with the data from the OECD (2011) STAN database.

Further, I constructed a larger fixed effect sample to obtain the exporter-industry fixed efffect. The sample includes nearly the complete countries of the TiVA, as it only requires gross exports and value-added export data. I made several adjustments to the sample industry and country coverage due to missing observations. First, concerning the industry coverage I excluded the "Utilities" industry from all estimations due to missing observations. Further I excluded certain countries (Malta, Island, Costa Rica, Brunei Darussalam, Cambodia) because they had no positive exports in at least one sector in 2005. Especially, the last two countries recorded many zero exports. The data recorded for both countries positive exports in less than fifty percent of the cases. Further, I excluded Saudi Arabia from the estimations because it exports mainly oil. Hence, I expect that for Saudi Arabia the soruces of CA are rather factor endowments than productivity differences. E.g. in 2005 the share of petroleum exports accounts for 90% of the fob exports of Saudi Arabia (Organization of the Petroleum Exporting Countries 2008).

2.6. Empirical results structural RCA

In this section, I describe the empirical results of the two step procedure to estimate the structural RCA measure. The first step, is to estimate θ with OLS and IV methods. In the second step I obtain the structural RCA indicator form a fixed effects regression. I discuss the association of the RCA indicator for the three indicators in two steps. First, I will compare the structural RCA ranking for the country pair China and the USA in all sectors. Second, I will present the association of structural RCA against GDP plot, which will show whether the RCA ranking is stronger for high income countries.

In the table 1 I show the cross-sectional results for the year 2005 of OLS and IV estimation. In the IV estimations, I instrumented the regressor productivity with research and development expenditures as in Costinot, Donaldson, and Komunjer (2011). First, I note that in both tables as theoretically predicted the point estimates of θ are positive and significant. In the columns 2-5 I report the IV estimates of θ for different samples regarding the industry coverage. and country coverage to show that estimates are robust. 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 shows a small yet statistically 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 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 bias of an measurement error would cause the estimate to be biased towards zero (Angrist and Krueger 2001). Further, another reason is that a simultaneity bias may occur due to agglomeration effects (Costinot, Donaldson, and Komunjer 2011). Agglomeration effects describe that firms locate geographically close to each other and learn about exports opportunities. This creates a link from higher export levels to increased the productivity.

The IV regression requires that the instrument is valid, which means that it satisfies the exclusion restriction and that it is relevant. The exclusion restriction requires that the instrument R& D has no explanatory power for exports except through productivity. The restriction is plausible Costinot, Donaldson, and Komunjer (2011) showed that for their sample the estimates of θ were not sensitive, to changes of productivity as total factor productivity. The interpretation is that the variations of the prices, which are explained by R & D are orthogonal to factor endowment trade motifs.

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 instrument affects the endogenous regressor in the hypothesized way. First, 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. This implies that the instrument is highly relevant. Further, the first stage shows a statistical significant positive effect of R&D on the inverse of producer prices. Concluding, the first stage confirms the expected positive effect of R& D on the inverse of producer prices and confirms the statistical validity of the IV regression.

The IV estimates of θ across the samples and the dependent variables gross exports and backward value-added exports show following results. First, the comparison between the full and the sample excluding primary industries shows that the estimates are similar for θ for the dependent variable backward value-added and gross exports. In general the estimates of θ for both dependent variables are very close and the difference is statistically indistinguishable. Second, the pattern for both dependent variables shows that the estimate base on the sample excluding the primary industries is decreased, however statistically not significant decrease compared to the full sample. Third, the sample excluding no high-income countries and primary industries shows an statistically significantly increased estimate [8]. The magnitude of the increase of θ is about 28 %. A higher estimate of θ implies a decreased dispersion of relative cost, this is what I expected for the sample with high income countries. The estimates of θ for forward value-added exports do not show a clear pattern.

Comparing the IV results to the estimates to the results of Costinot, Donaldson, and Komunjer (2011), I obtain a similar estimate to the authors results for gross exports as dependent variable (θ11.1 SE 0.981). However, the authors favorite estimate uses openness corrected exports, for which they obtained an estimate of 6.58. They motivated using openness corrected gross exports to account for trade selection [9] 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.

For two reasons I decided to use gross exports and value added exports without correcting for openness. First the data on the import penetration ratio is only available for the manufacturing industries, which would reduce the sample considerably. Second, I was unable to obtain a similar correction for VAX [10]

8. I performed an significance based on the t-test. 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 denotes the between imputation variation of the estimated parameter (p.77)Rubin1987

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

10. A possible definition openness for value-added exports might be the ratio of re-imported value added from domestic industries to VAX . However, this measure was not bounded between 0 and 1 and the results of estimating θ with this correction showed .

(a) Cross-section results I

	(1) OLS	(2) Full Sample	(3) Without primary industries	(4) Without primary industries high [11]
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 2-6 with log of R&D expenditures
Without primary industries excludes the industries mining and agriculture
* based on Fisher's z transformation

(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 2-6 with log of R&D expenditures
Without primary industries excludes mining and agriculture industry
* based on Fisher's z transformation

(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	−0.01908	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 2-6 with log of R&D expenditures
Without primary industries excludes the industries mining and agriculture
* based on Fisher's z transformation

Table 2.1.: Cross-section Results OLS and IV

2.7. Ranking of structural Ricardian comparative advantage

In this section I present the results of the structural RCA ranking for both value-added exports measures and gross exports. I will give two views on the structural RCA ranking a global view and a local view. First, the global view will shows a scatter plot of the association of the RCA ranking for gross exports and value-added exports and a country's per capita GDP. The global view investigates the hypothesis that country's with a higher GDP show a higher similarity between the rankings and hence their sourcing and factor usage are not strongly sector specific.

Second, the local view focuses on comparing the structural RCA ranking for Belgium and Germany of forward and backward value-added exports to gross exports. In this way, I analyze two aspects, first whether value-added exports alter the picture of comparative advantage and whether the different perspectives of value-added exports affect the results.

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

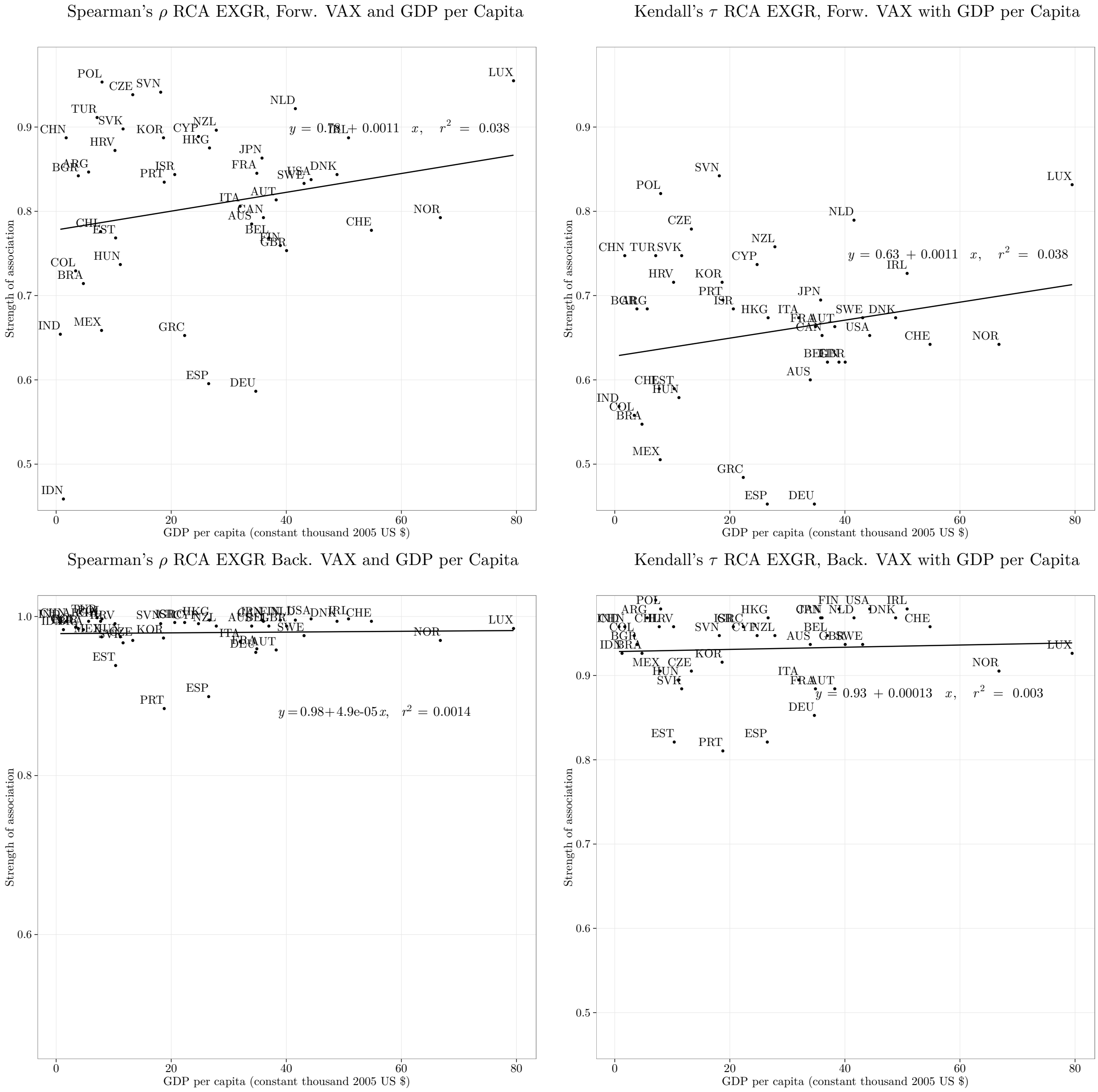
I will discuss the choice of the association measures shortly. First of all, I chose the Spearman's ρ since I focus on the similarity of the rankings and the strength of the monotonic association between them. Moreover I chose Kendall's τ as it computes the similarity of the two rankings, by the means of counting the number of country pairs, which are different between two rankings.

I outline the construction of Kendall's τ based on Abdi (2007). The outline makes the simple interpretation of Kendall's τ in terms of probabilities more clear. The basic idea behind the measure is to count the number of different pairs of two sets of ordered objects, which include the same objects Abdi (2007). I illustrate this idea in the context of the RCA rankings. For two RCA rankings, the measure is based on counting the number of different ordered country pairs, which I denote as $d(P_1, P_2)$, where $P_i \quad i = 1, 2$ indicates the two ordered set of pairs obtained from the country rankings. In the next step, this number is normalized such that is bounded by -1 and 1, where -1 reflects the largest differences and 1 is equal to the smallest difference. Kendall's τ is then defined as follows

$$\tau = \frac{1/2N(N-1) - d(P_1, P_2)}{1/2N(N-1)}.$$

Moreover, Kendall's τ has an intuitive stochastic interpretation based on the idea of drawing orderd pairs (Abdi 2007). In the context of two country rankings, the interpretation is that if a country pair is randomly drawn from each ranking, Kendall's τ is the difference between the probability that the draws have the same order and the probability that the country pairs have a different order. The focus of Kendall's τ on country pairs is especially useful for RCA, as the main focus of RCA is to compare pairs of countries and industries.

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



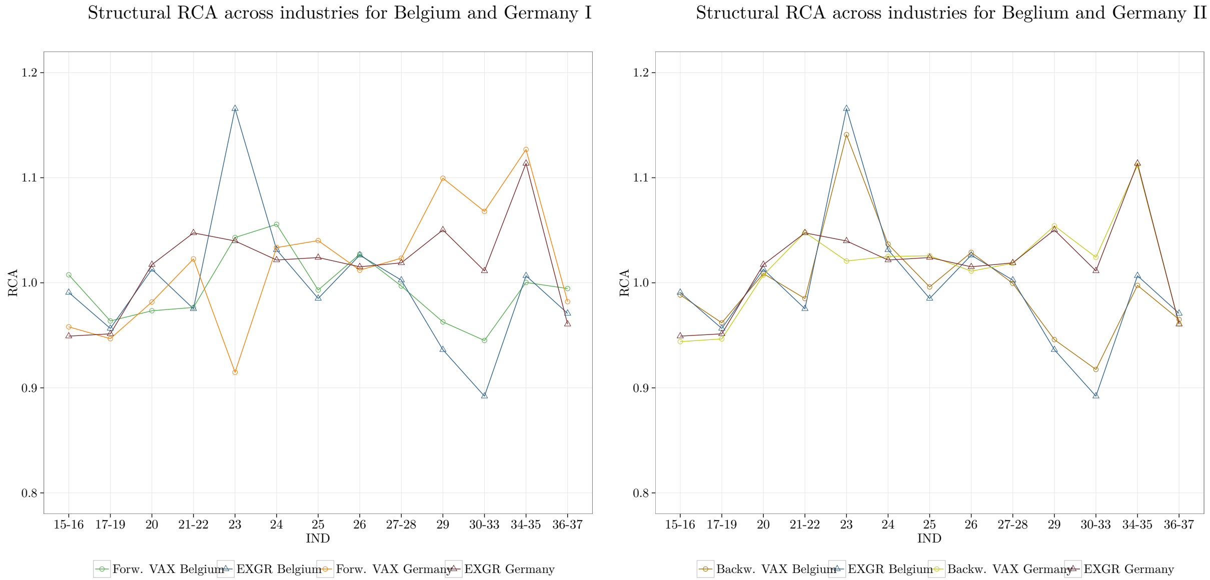
I conclude four findings from the figures above. Firstly, the RCA rankings based on gross exports and backward value-added exports show a high degree of similarity for all countries. Secondly, the association between forward value-added exports and gross exports is substantially lower than the association of backward value-added and gross exports. Thirdly, the association of gross exports and forw. VAX shows a weak positive relation with a country's GDP per capita. As a final point I observe that the overall strenght of the associations as measured by Spearman's ρ is higher than Kendall's τ.

The first and second finding are similar to the results in the estimation of the θ parameter. The results showed that the estimates for backw. VAX and gross exports showed a similar pattern, while the estimates of forw. VAX were reduced and did not follow a pattern. The third finding is consistent with the hypothesis that countries with a higher GDP have less sector specific input and sourcing patterns. The forth finding is consistent with the result that asymptotically the ratio of the population analog of Spearman's ρ and Kendall's τ is equal to three half (Fredricks and Nelsen 2007).

Turning to the local view I present below the normalized RCA based on both value-added export measures and gross exports for the industries of the manufacturing sector. the RCA [12] as in Leromain and Orefice (2014).

12. ormally, I define it as follows $RCA_i^k = \frac{z_i^k * \bar{z}}{\bar{z}_i * \bar{z}^k}$, where $\bar{z}$ denotes the grand mean, $\bar{z}^k$ denotes the sector specific mean and $\bar{z}_i$ denotes the country specific mean.

Figure 2.2.: Country pair RCA based on for- and backward VAX & EXGR



According to the normalization, an RCA value above (below) 1 indicates a comparative (dis)advantage of a country in the particular industry.

To start with the RCA ranking results, I discuss the results for backward value-added exports and gross exports. The major point which I conclude is that for both countries backward value added closely traces the RCA pattern of gross exports. This results resembles the result of the estimation of θ, where I observed a similar pattern.

The graph highlights that both countries highlights have an comparative advantage in the following sectors 20 23 24 26. Germany has an higher comparative advantage in seven industries namely 20 21-22 25 27-28 29 30-33 34-35. On the other hand, Belgium has an comparative advantage in six industries namely the food industry, 17-19 23 24 26 36-37.

In contrast, I observe that the RCA rankings are different for forward value-added gross exports. Initially, I observe a largest decrease of RCA under forw. VAX compared to gross exports in the sector 23 and 20. Additionally for Germany large decrease of RCA for the industry 21-22. The largest decrease of RCA induces a change of 15% in the industry 23 for both countries. As a result, the industry 23 changes from a comparative advatange to a comparative disadvantage for Germany. The industry changes from an comparative advantage to comparative disadvantage, whereas for Belgium the industry remains a comparative advantage. Additionally, in industry 20 both countries show no longer an comparative advantage under forw. VAX, whereas under gross exports they show an comparative advantage. Moreover, I observe the largest increase of RCA under VAX compared to EXGR in the industry 29 and 30-33. For both countries I observe an increase of about 5% in the industry 30-33 for forw. VAX, and a somewhat smaller increase in the industry 29. However, the pattern of RCA remains unchanged for both countries, for Belgium an comparative disadvantage and Germany has an comparative advantage. Concluding, the graphs show that forward value-added exports changes the pattern of RCA, whereas backward value-added exports traces the pattern of RCA under gross exports.

3. Relative network centrality and structural Ricardian comparative advantage

In this chapter, I analyze the association between the structural RCA indicator and network centrality. The motivation is as follows. First according to Ricardo model I expect that a industry within a country with relative lower cost to produce and export more a good. The eigenvector centrality of a country is high in a industry network if it exports to destinations which are important exporters themselves. A higher centrality of a country corresponds to higher trade shares. Therefore we hypothesize that there may be a link between both measures. The literature propagation of shocks in network Acemoglu et al. 2012 showed that network centrality Further, I analyze the robustness of the correlation results to changes of the normalization or changes of the sample.

3.1. International trade network

In the following I define the trade network as directed and weighted network. The definition is based on Jackson (2010) and De Benedictis and Tajoli (2010).

In the trade network each node represents a particular industry k in different countries $N^k = 1, \ldots, n$, to simplify notation I omit high script k in the following. In the network terminology, nodes are connected by edges.

Each edge $g^k_{i,j}$ represents a trade relation between an industry k in exporting country i to importing country j. Further, in the trade network exports are not equal to imports $g^k_{i,j} \neq g^k_{j,i}$, therefore I specify a directed network. Further, the variable $g^k_{i,j}$ is equal to one if there are positive exports in industry k in country i to country j and zero otherwise. Thus, $g^k_{i,j} = \begin{cases} 1 & \text{if} \quad x^k_{i,j} \neq 0 \\ 0 & \text{if} \quad x^k_{i,j} = 0 \end{cases}$ Moreover, I collect all $g^k_{i,j}$ in k symmetric matrices g^k of the dimensions $n \times n$. The binary trade network for all industries is represented by k tuples of nodes and trade relationships $\mathcal{G}^k(N, g^k)$.

In the next step I define the weighted international trade network. To start with, the weight variable $W^k_{i,j}$ which represents the value of exports from industry k in country i to country j. To simplify we represent the weight variable, in k matrices W^k, which I denote as weight matrices. The international trade network in each industry is therefore the binary trade network combined with the $ITN^k = (\mathcal{G}^k(N, g^k), W^k)$.

3.2. Network centrality

After having defined the international trade network, in turn I outline the concept of eigenvector network centrality. The definition is based on the textbook of Jackson (2010) The outline of network centrality of Jackson (2010) is based on the mathematical outline in Bonacich (1972). The outline will be based on the binary trade network, it can be applied without changes to both weighted and directed networks (Jackson 2010).

The idea of centrality is as follows, a node is more central if it is connected to more central nodes. The centrality of the connected nodes is as well determined by how central the nodes are they are connected to. This recursion can be mathematical represented

$$\lambda C^e_i(g^k) = \sum_{j \neq i} g^k_{i,j} C^e_{i,j}(g^k)$$

Restating the equation in matrix notation and solving it

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

where λ is the corresponding eigenvalue to the eigenvector $C^e(g^k)$. Following the convention I use the leading eigenvector, which corresponds with the largest eigenvalue.

To compute the relative network centrality I calculate the eigenvector centrality for the weight matrix W^k for each industry k. In addition, in the computation of the eigenvector centrality I column normalize the weight matrix, which therefore the export shares of each exporting country. Further important for the computations, is the Perron-Frobenius Theorem. As the weight matrix is nonzero and for some power n, where n is a positive integer, there exists according to the theorem a unique right-hand eigenvector $C^e(g^k)$, which solves the centrality equation with the largest eigenvalue equal to one $\lambda = 1$. The interpretation of centrality of an industry k in country i in this , how important in terms of $ it is in the network of imports from all other countries j.

3.3. Network centrality and structural Ricardian comparative advantage

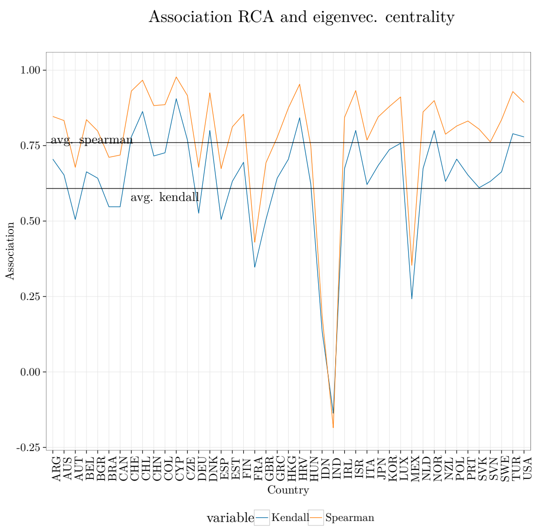
In this subsection, I analyze the association between relative network centrality and structural RCA based on domestic value added exports. In the introduction I outlined the similarity of centrality and the stochastic interpretation of trade shares. Network centrality may be interpreted as how much each industry contributes in terms of dollars to the exporting network. The stochastic interpretation of trade shares reflects productivity draws. This similarity is the motivation behind the analysis conducted.

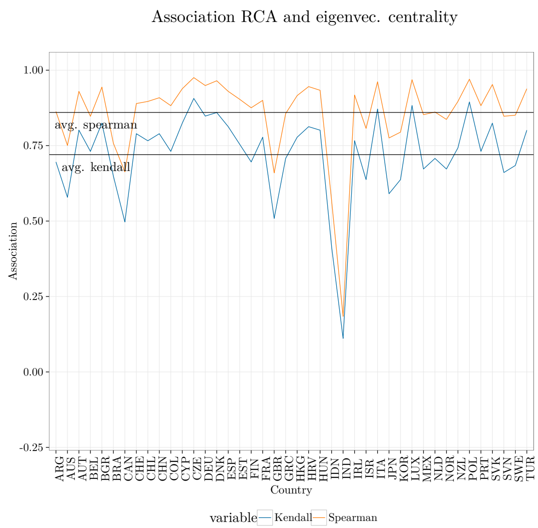
I compared the association of network centrality and structural RCA based on simple and rank correlations. The simple correlation should indicate a stronger association than the rank correlation if the relationship is linear, when both measures are approximately normal distributed and outliners are absent. The second method is more robust to outliners and does not assume that the measures analyzed are normally distributed. Further, a stronger rank correlation than simple correlation would imply that the relation between the measures is monotone instead of linear.

In table **??** the correlation between both measures shows is between 0.51-0.87. The lowest correlation is in the construction industry. The highest coefficient is in the food and beverages industry (ISIC 15-16). In column two, which shows rank correlation, most coefficients are concentrated around 0.9. Further the coefficents span a smaller interval from 0.76 to 0.95. The highest coefficient is in the machinery and equipment industry and the other non-metal mineral products industry. The lowest rank correlation is in the food and beverages industry.

The rank correlation is for most industries larger than the simple correlation coefficients. The relative magnitude of the differences between the coefficients is between 67% for social services and -12% for the food industry. The food industry shows the only decreased coefficient, another industry, the petroleum industry, shows an rank correlation coefficient, which is only slightly higher.

Overall the results in the table **??** show that between relative centrality and structural RCA a higher rank correlation, and therefore the relation between both measures is rather monotone instead of linear. The economic interpretation of this is that the sector-specific input and sourcing structures do not vary strongly across sectors.





??

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 estimated the measure for both domestic value-added and gross exports and compared the results using simple and rank correlations.

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 this regression, 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 used the complete TiVA sample with 56 countries.

In the first regression, I obtained an estimate for θ, which may be interpreted as the inverse of the productivity dispersion. Comparing the point estimates of the productivity dispersion parameter to the results in Eaton and Kortum (2002) and in Costinot, Donaldson, and Komunjer (2011), I find that my point estimates were at the upper bound of the 95 % confidence interval of the first and my estimates were significantly higher than the results of the latter. Moreover, I found that the estimates of θ using value-added exports showed increased values. This results may be explained by the construction of domestic value-added, which are net of double counting, foreign value-added and domestically absorbed exports. The variance of Domestic value-added exports band compared to gross exports. If fewer variations of the regressand are explained by the same variations of the observed productivity regressor, the estimated inverse of productivity dispersion θ should be increased.

Further, to assess the robustness of the estimates I reestimate the regression in the multiplicative form with PPML methods, as suggested by Silva and Tenreyro (2006). The estimates of the dispersion parameter showed a statistical not significant decreased estimate compared to the log-linear specification. An explanation of the result is that in the levels specification there is more variation of the regressand and, therefore, the estimates decreased. The results confirmed that the estimated values are not sensitive to the sample or the estimation technique.

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. Moreover, I compared the ranking results for the real estate industry to the results of Koopman, Wang, and Wei (2014). In contrast to the authors, I find that domestic value-added exports do 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

Abdi, H. 2007. “The Kendall rank correlation coefficient.” In *Encyclopedia of Measurement and Statistics,* edited by N. Salkind, 508–510. Sage.

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.

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.

Buuren, S. v., and C. Oudshoorn. 2000. *Multivariate imputation by chained equations: MICE V1. 0 user’s manual.* Technical report. TNO.

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.

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.

Greene, W. H. 2003. *Econometric analysis.* Pearson Education India.

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.

Hummels, D., J. Ishii, and K.-M. Yi. 2001. "The nature and growth of vertical specialization in world trade." *Journal of International Economics* 54 (1): 75–96.

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.*

Royston, P., I. R. White, et al. 2011. "Multiple imputation by chained equations (MICE): implementation in Stata." *Journal of Statistical Software* 45 (4): 1–20.

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.

Shikher, S. 2011. "Capital, technology, and specialization in the neoclassical model." *Journal of International Economics* 83 (2): 229–242.

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.

Van Buuren, S. 2007. "Multiple imputation of discrete and continuous data by fully conditional specification." *Statistical methods in medical research* 16 (3): 219–242.

Van Buuren, S., J. P. Brand, C. Groothuis-Oudshoorn, and D. B. Rubin. 2006. "Fully conditional specification in multivariate imputation." *Journal of statistical computation and simulation* 76 (12): 1049–1064.

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
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$