Contents

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, 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). Feenstra (1998) characterized IPF as the break down the vertically-integrated production process. In the vertically-integrated process an industry produced a good in several production steps. Under IPF goods are increasingly produced combining domestic and international manufacturing and services.

A reference point of comparisions of IPF is 1970s, which is seen as begining. , Whereas the spread of IPF 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 and Orefice (2014) 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 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 problematicIn contrast, **Leromain** 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 assosciation 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. Further, I find that my results are robust to changes in the country coverages. .

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. 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 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 **johson**. 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, a backward-linkages based view is useful to understand how large the share is of a sector country domestic value added in it'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**. 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 WIOD **Timmer2012**. I choose the TiVA database as the main source as the TiVA data provides a larger country coverage with a more regional diverse focus. In addition, to my knowledge only one author has previously employed the TiVA data (Johnson 2014). In addition, the TiVA data is at a similar aggregate industry level as the other data sources.

I combine the TiVA data with other data sources to create the estimation sample. The sample includes R &D expenditure data from OECD (2013) ANBERD and international producer price data from the GGDC (Inklaar and Timmer 2014). It includes twenty nine countries and twenty two industries at the ISIC Rev.3 level.

Further, I create a large sample, which uses only the TiVA to estimate a fixed effects regression. I included in this larger dataset all countries, which had records on forward &backward value-added exports [1]. Further, to obtain a consistent sample across industries I excluded some countries because they had no exports recorded in at least one sector [2]. Another adjustment is that I excluded Saudi Arabia because its exports mainly consist of oil [3] (Organization of the Petroleum Exporting Countries 2008). The final sample includes twenty two industries at the ISIC Rev.3 classification and fifty six countries.

In order to create the estimation sample several data adjustments were necessary. I reconciled the industry coverage of TiVA and the GGDC international price data by aggregating the international price data using the ISIC Rev. 3.1 two digits classification. The aggregation extended the sample such that it includes a larger share of service sector industries. Specifically, I aggregated the prices using a weighted average with weights equivalent to the relative value-added of an industry across the aggregate industries. I obtained the value-added output data from OECD (2011) STAN database. Moreover in the appendix in table A.4 I report 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 ommited, 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**), which derives the first estimation equation to obtain the fundamental productivty.

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 a 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 productivitz 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 modeld 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. Formaly, $d_{i,l}^k \leq d_{i,j}^k d_{j,l}^k$ 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 competion 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 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. 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

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

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, it is necessarz 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 $(\frac{\tilde{x}_{i,j}^k}{\tilde{x}_{i'j}^{k'}})/\frac{tildex_{i'j}^k}{\tilde{x}_{i'j}^{k'}})$ accounts for differences in the expenditure shares α^k.

Moreover, productivity term in eq. (2.4) is 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 necessarz to specify the measure of observed productivity. In the context of the model assumptionm, the inverse ratio of the producer price fully reflects the observed ratio of relative productivity.

According to **Costinot** econometrically equivalent equation may be estimated instead.

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

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-varrying 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 to the effect of $\theta \ln z_i^k$ on the bilateral gross exports in the first equation. Therefore, the fundamental productivity may be obtained as follows.

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

2.4. Generalization

In the following, I discuss the effect of sector-specific use of production factors and including capital and intermediate inputs as production factors. The use of production facotrs may be more industy specific, because goods are increasingly produced with industry specific supply-chains. I will highlight that the effects of sector-specifc use of production factors iw similar to sector-specific international sourcing of inputs.

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).

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

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

. 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 $\sum_{m=1}^{K} \eta_{i,m} = 1, \forall i$. With the outlined cost function, the RCA measure would reflect productivity differences and differences in factor endowments.

A special case of the generalized cost function is the relation in theorem. If the production factor used are 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 clear that the cost function above in combination with eq. (1) and the assumption about the distrubtion of z_i^k can be used, to simplify the equation. The expression one obtains is exactly as in eq. (2).

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. Analog to the previous argument about sector specific production factors, the international sourcing pattern would additionally cofound 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 in 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 using 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 variable. Hence, the assumption requires that the inverse of producer prices should be uncorelated to variabel 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 estimation correctly estimates a causal effect if the instrument satisfies two assumptions. 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 endogeonous producer prices. Further, the relevance assumption requires that the instrument is sufficiently strong correlated with the endogenous regressor.

The motivation to use R & D expenditures as instrument are as follows. Firstly, I expect that higher R & D expenditures increase the productivity of a industry. In our model the lower production cost would be passed through to the producer pirce 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 statistical significant. Such an 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, which can not be empirically tested (()p.109)cameron2009. Secondly, R&D is used as an instrument in both Costinot, Donaldson, and Komunjer (2011) and Eaton and Kortum (2002).

However, if R&D would not be sufficiently strong correlated wit the endogeneous regressor, it would be a weak instrument. A weak instruments 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 10 (Douglas Staiger 1997).

2.5.1. Missing data imputation – work in progress

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 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 imputation, 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 distrubiton is in Bayesian interference obtained by divding the product of the assumd prior distribution and the Likelihood by normalizing constant. The posterior predicitive 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 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 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 imputation. To start with I assume that the indicators of missing values are random variables with a distribution. Additionally, I assume 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 $Var(\kappa|X) = \sqrt{\bar{V} + 1/(m-1) * B} m\, where B = 1/M - 1 * \sum_m = 1^M (\kappa_m - k\bar{a}ppa_m)^2$ and $\bar{V} = 1/M * sum_m = 1^M V_m$ with$V_M = Var X_obs, X_m is(d)$.

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

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 neighbor matching technique suggested 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 closested 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 regresion 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.

Moreover, I included in the regression imputation model country and industry fixed effects dummies. The

predicted value average over the posterior distribution.

6. 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).

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 interdmediate of estimating the paramater θ to obtain the 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 RCA ranking is stronger for high income countries. Second, I compare the structural RCA ranking for the country pair Belgium and the Germany across the manufacturing industries.

In table 1 I show the cross-sectional results for the year 2005 of OLS and IV estimation. The columns (2)-(4) report the IV estimates of θ for different samples regarding the industry coverage. and country coverage to 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 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).

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 statistical 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 pirces.

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 somewhat decreased, yet statisically not significant. Third, the sample excluding no high-income countries and primary industries shows an statistically significantly increased estimate compared to the full sample [7]. A higher estimate of θ implies a decreased dispersion of relative cost, which is not surprising for the sample with high income countries.

The estimates of θ for forward value-added exports do not show a clear pattern. The estimates of θ with forward value-added exports are reduced compared to the other estimates and show only little variation across the samples. Especially, I note that the estimate for the forth sample does not show an inceased point estimate.

Comparing to the estimates to the results of Costinot, Donaldson, and Komunjer (2011), I find that the estimates are comparable.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 [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.

7. Based on a 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 denotes the between imputation variation of the estimated parameter (p.77)Rubin1987

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).

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 [9]

(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
* pooled using 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 2-4 with log of R&D expenditures
Without primary industries excludes mining and agriculture industry
* pooled using 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 2-4 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

9. 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 estimates of θ with this correction showed unplausible values.

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 two views on the structural RCA ranking a global view and a local view. First, the global view I show 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

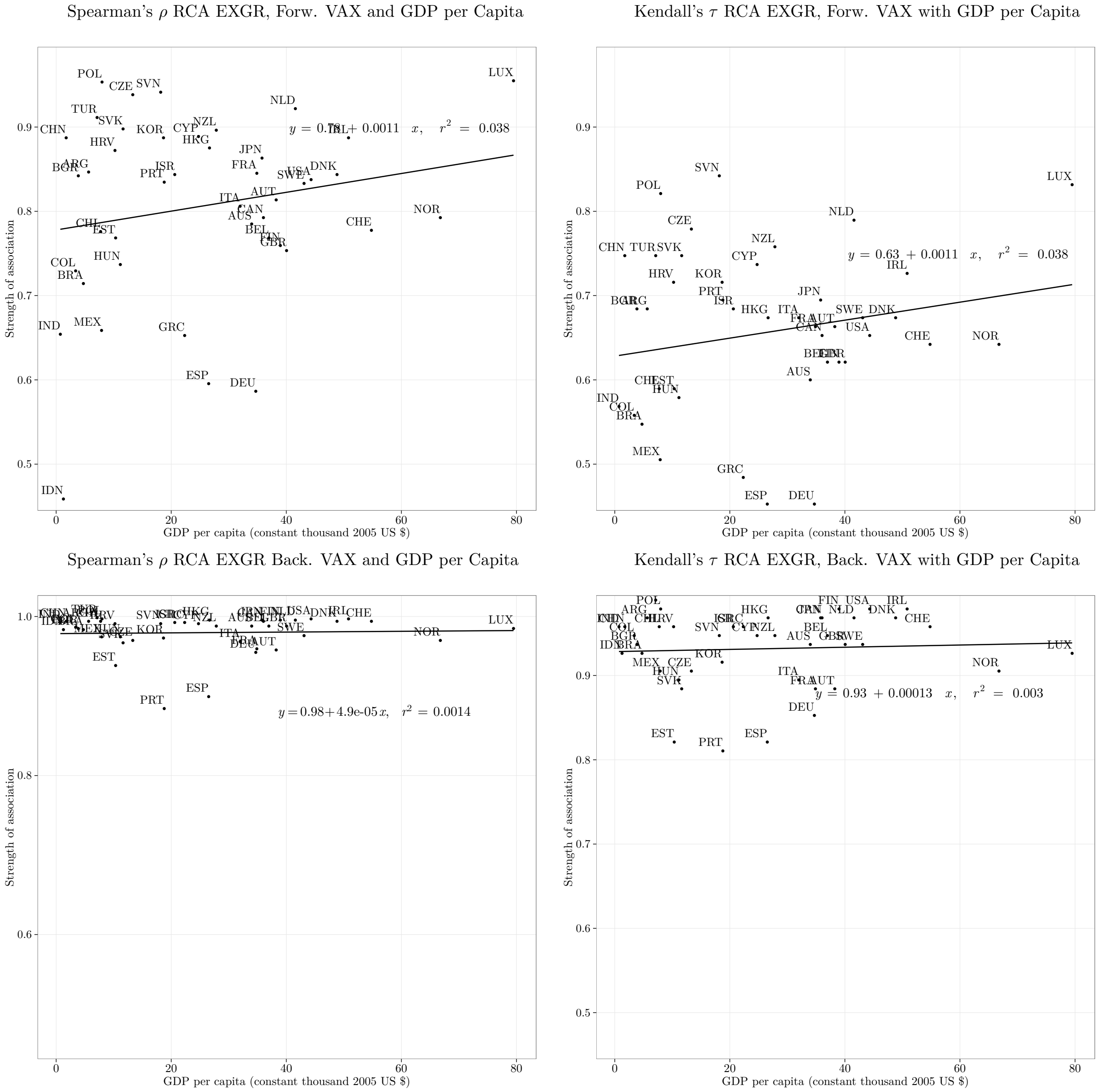
I discuss the choice of the association measures shortly. First of all, I chose the Spearman's ρ since I focus on the similarity of the rankings and the strength of the monotonic association between them. Moreover I chose Kendall's τ as it computes the similarity of the two rankings, by the means of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



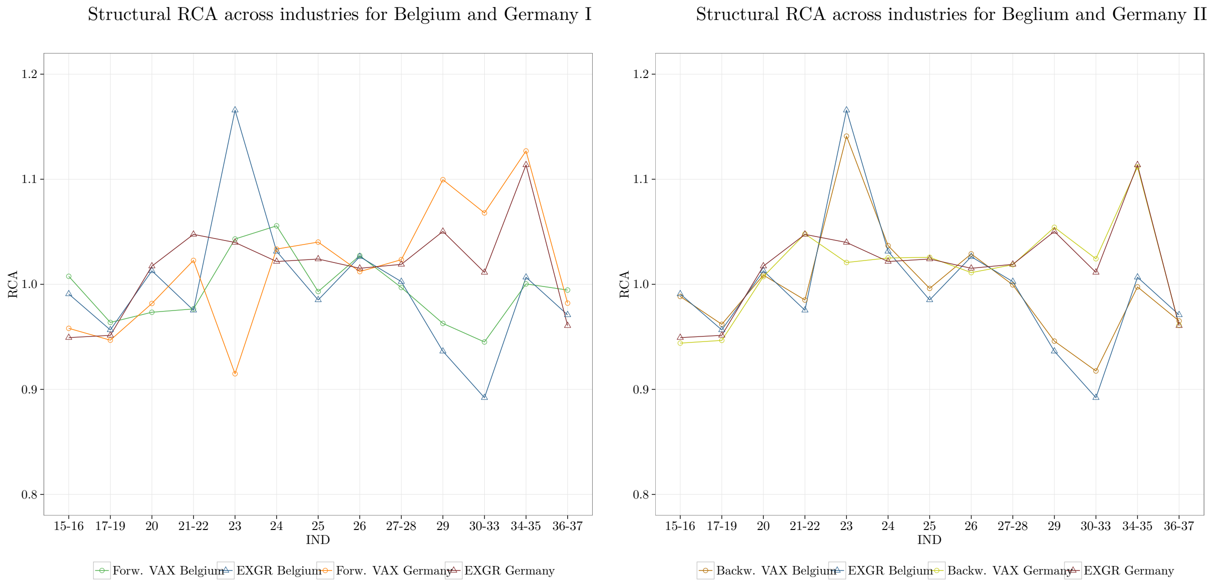
I conclude four findings from the figures above. Firstly, the RCA rankings based on gross exports and backward value-added exports show a high degree of similarity for all countries. Secondly, the association between forward value-added exports and gross exports is substantially lower than the association 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 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. 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 [11] as in Leromain and Orefice (2014).

11. Formally, I define it as follows $RCA_i^k = \frac{z_i^k * \bar{z}}{\bar{z}_i * \bar{z}^k}$, where $\bar{z}$ 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. I find 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.

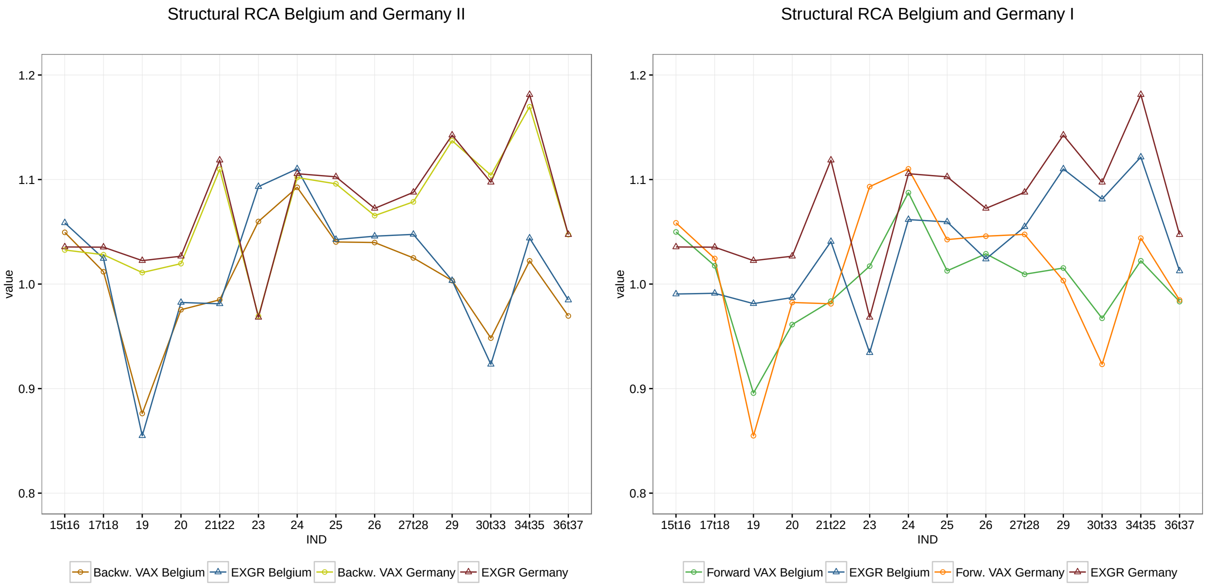
The graph highlights that both countries highlights have an comparative advantage in the following sectors 20 23 24 26. Germany has an higher comparative advantage in seven industries namely 20 21-22 25 27-28 29 30-33 34-35. On the other hand, Belgium has an comparative advantage in six industries namely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

2.7.2. RCA based on WIOD

In this subsection I compare the previous results about the RCA for BEL and GER to the results based on the WIOD data. Firstly, compared to the TiVA data the WIOD data comes at a greater level of dissagregation.

Figure 2.3.: Country pair RCA based on for- and backward VAX & EXGR



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 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 results to changes of the normalization and 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). I 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 $n \times n$. Now, based on the notation I define the binary trade network for each industry k as the tuple of nodes and trade relationships $\mathcal{G}^k(N, g^k)$.

To extent the binary trade network to its weighted version, I define 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 value of the exports in the trade relationship. Analog to the matrices of trade relationships for the binary trade networks, I define for each industry k a matrix, which records the weight of each trade relationship. This matrix W^k is denoted as weight matrix.

er interactions have taken place (De Benedictis and Tajoli 2010).

Based on the notation the weighted trade network for each industry is combination of the binary trade network and the 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 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. Mathematically, defining the eigenvector centrality $C^e_i(g^k)$ associated with the network g^k, the centrality of an actor 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. An important property of the eigenvector is implied by the Perron-Frobenius theorem. 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 eignevector for the largest eigenvalue is positive. Further, the theorem implies that if for some power the matrix g^k is positive, than 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 Peron-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, likewise one could instead of 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, 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-eignevector 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 normalized both measures to the grand mean, the mean across industries and the mean across countries. Secondly, I normalized both measures relative to a benchmark country (the USA) and a benchmark industry (the food and beverages industry).

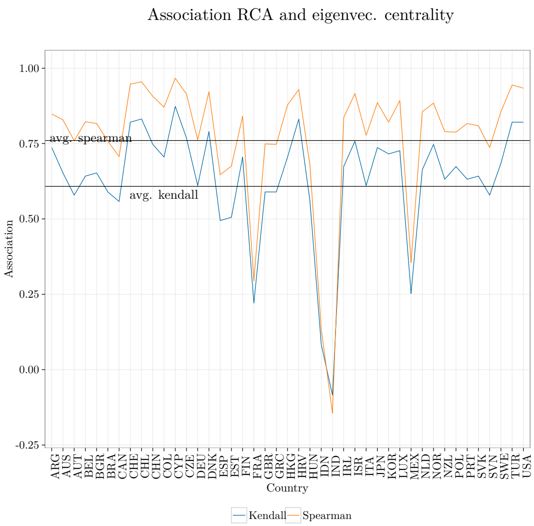
In the introduction I outlined the similarity of out eigenvector centrality and the RCA. Given the similarity, if structural RCA and network centrality show comparable results, the later may be preferred as a simpler measure. Additionally, the network shock propagation literature showed two aspects about network centrality, Firstly, a industry with higher network centrality in the input-output network contributes relative more to the value-added. Moreover, it showed that network centrality is related to microshocks.

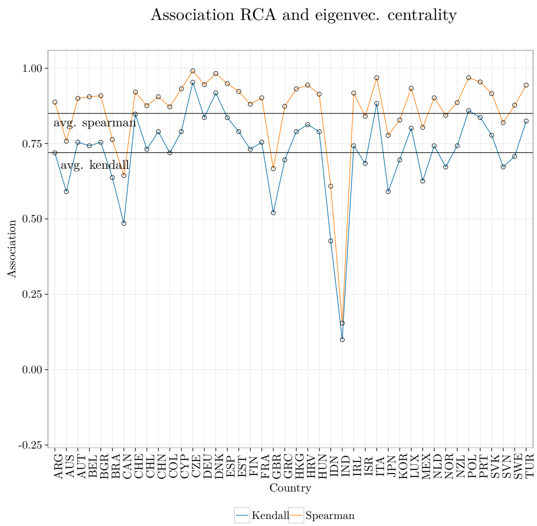
In the following I first describe Tthe left graph, which shows the association between network centrality and RCA according to the first normalization. First interpreting the left graph I see that for most countries the association of the rankings are quite high. This can be seen as the average of Spearman */rho* is 0.76 and Kendall's */tau* is 0.58. Both China and the USA have a high similarity between the rankings. On the other hand Finland and Mexico show relative high dissimilarity between the rankings. I obtain very strong results for India and Indonesia. For India is that Spearmans */rho* and Kendall's */tau* are negative. For Indonesia, the results show a low yet positive coefficients for both measures.

1. A column stochastic matrix is a matrix, where each column sum is equal to one.

The right graph shows overall a higher association between RCA and network centrality. This can be seen from the fact that the average of Spearman /*rho* is 0.85 and Kendall's /*tau* 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 sign of the association is positive. Both Finland and Mexico show a increased similarity in the graph. Further, it is noteworthy that for Canada the strenght of both association 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 conection 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 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.

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.

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.

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.*

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