New Evidence for the Linder Hypothesis and the two Extensive Margins of Trade

Claudia Bernasconi*
February 2009

Abstract

I analyze the two extensive margins of international trade flows (product and country-level) by an application of the Linder hypothesis. Linder accounts for non homothetic preferences as he believes individuals with different income levels to consume different bundles of goods. Because he considers domestic consumption and production to be linked, the more similar consumption bundles (or demand structures) across countries are, the more intensive is trade. The relevant determinants for demand are per capita income and the income distribution within a country. Although the extensive margin of trade is the driving force behind this hypothesis it has been neglected in the literature so far. To my knowledge previous studies solely analyzed trade volumes. I carefully translate Linder's theory into an empirical model and show that it explains the numerous zeros at the product-level as well as at the country-level present in international trade data very well. I find strong supportive evidence for the demand sided effects proposed by Linder. First, the more similar per capita incomes are the more diversified are the traded bundles of goods. Second, the more uneven the within income distribution is the higher is the extensive margin of trade. Moreover, I find that countries with more similar demand structures also have a higher probability to exhibit positive trade. Thus, the demand sided effects are at work at *both* extensive margins of trade.

To account for the fact that about 40% of all possible trade relations at the country-level are inexistent I adapt the empirical strategy derived in Helpman et al. (2008). I find a highly significant and positive selection bias.

I use a very comprehensive and rich dataset regarding the cross-section, level of disaggregation and data about within income distribution. Unlike previous studies, this enables me to state that my findings suggest the Linder hypothesis to be a global phenomena. I even find that the relationship between similarity in per capita income and the extensive margin is stronger when high income economies are excluded from the analysis. All results are very robust to numerous alterations.

JEL classification: F10, F14, D12, D31
Keywords: Linder Hypothesis, Extensive Margin of Trade, Non-homothetic Preferences, North-South Trade

[0]University of Zurich, Institute for Empirical Research in Economics, Bluemlisalpstrasse 10, CH - 8006 Zurich, Tel: +41 44 634 37 25, Fax: +41 44 634 37 18, e-mail: bernasconi@iew.uzh.ch

1 Introduction

Bilateral as well as world trade flows evolve at two margins, the extensive and the intensive. Although empirical trade literature has a long history research about the extensive margins of international trade flows has started only a few years ago. In this paper I analyze the two extensive margins of international trade flows by an application of the famous Linder hypothesis. As I will argue in more detail below, Linder's hypothesis (1961) that closer levels of per capita income will translate into more intensive trade is driven by the extensive margin of trade. To my knowledge, all previous empirical studies of Linder's theory neglected the extensive margin of trade – although it is the driving force behind the hypothesis.[1] The *product-level* extensive margin determines the diversification of bilaterally traded good bundles. Why does country i export a very diverse good bundle to country j while it exports only in a handful of product categories to country k? The *country-level* extensive margin reflects that some country pairs trade with each other while others do not exchange a single good (trade versus autarky). The identification of the determinants of the two extensive margins of international trade is important as it focuses on how diversified or intense, in terms of product varieties, international trade relations are. Bilateral trade relations which are based on few industries are distinguished from country pairs which are linked through many different industries. Observations with zero bilateral trade are especially interesting as they shed light on the determinants of the establishment of trade relations.

The product-level extensive margin of a bilateral trade flow captures how many different product varieties are shipped, i.e. the diversity of the bundle of goods, while the corresponding intensive margin represents the total value of shipments in each product variety. The extensive margin should be considered as binary. A country i may export a certain product category to a country j or it may not – a one or a zero. The analysis of the extensive margin of bilateral trade flows has started only a few years ago, both theoretically and empirically. In his seminal work Melitz (2003) proposes an extension of Krugman's (1980) trade model that incorporates heterogeneous firms to analyze the intra-industry effects of international trade. Melitz shows an equilibrium in which only the more productive firms export while some less productive firms continue to produce for only the domestic market. Helpman et al. (2008), Hummels and Klenow (2005) and Baldwin and Harrigan (2007) show that the extensive margin boosts bilateral trade flows substantially. To my knowledge all studies about the extensive margin take a supply-sided approach. I provide *complementary* evidence in showing that the extensive margin of trade is also influenced by important demand sided effects.

The international trade volume can be decomposed into the number of active bilateral trade relations – country-level extensive margin – and the trade volumes in established relationships. Trade data contains many zeros at the country-level

[1] A search on *The American Economic Association's electronic bibliography of economic literature* (EconLit) for example, does not give one single reference to studies about the Linder hypothesis which are related to the extensive margin of bilateral trade flows.

extensive margin, i.e. there are many country pairs which do not trade at all. However, systematic treatment of this bulk of country-level zeros has also just begun to emerge, see Helpman et al. (2008), Kohler and Felbermayr (2004), Silva and Tenreyro (2006) and Hallak (2006). While many previous studies neglected the zeros by restricting the attention to positive trade flows I adapt the empirical strategy derived in Helpman, Melitz and Rubinstein (2008), henceforth called HMR. Just as HMR and Hallak (2006) I find a selection bias, i.e. trade flows which are positive are not representative for all possible trade relations[2]. The fact that country i does export to country j contains valuable information which needs to be utilized in order to obtain consistent estimates. To control for this selection bias I use the 2-step procedure of Heckman (1979) and compare it with the benchmark of a standard OLS model.

Linder (1961) formulated one of few trade theories which take a demand rather than a supply-sided approach.[3] The author states that a country cannot achieve a comparative advantage in the production of a good which is not demanded on the home market. In turn *trade will be most intensive among countries with similar demand structures due to a larger overlap of production and comsumption patterns.* When carefully reading Linder's theory it becomes obvious that this overlap of demand is in terms of product variety. Hence, his hypothesis is solely driven by the product-level extensive margin of trade. Linder considers the per capita income to be the most important factor for the demand structure. A close second comes the income distribution within a country. In sum, Linder predicts countries with similar demand structures to have more goods, in terms of product variety, for which demand overlaps (see Linder (1961) page 103) – i.e. the more similar per capita incomes are the more product varieties are traded. To my knowledge the Linder hypothesis has not been applied yet to the extensive margin of trade, although the latter is the driving force behind the hypothesis. This paper aims at filling this gap in the literature. I carefully translate Linder's theory into an empirical model and show that it explains the numerous zeros (of both types) present in international trade data very well. The more similar demand structures are, the fewer zeros occur – at both extensive margins of international trade.
Early empirical tests of the Linder hypothesis concerning trade volumes were rather mixed while more recent studies which use broader datasets as well as richer empirical strategies provide supportive evidence (e.g. Choi (2002), McPherson et al. (2000), Fillat-Castejon and Serrano-Sanz (2004), Thursby and Thursby (1987)). Hallak (2008) identifies the effect of product-quality operating at the demand side through the relationship between per capita income and aggregate demand for quality.[4] His empirical results provide also supportive evidence for the hypothesis.

Linder's verbal theory accounts for non-homothetic preferences. When focusing on

[2] zero and positive trade flows

[3] It is important to understand the demand oriented aspects of trade as for many decades and industries it is rather the case that supply exceeds demand than the other way round.

[4] He emphasizes an aggregation bias and therefore proposes to estimate the Linder hypothesis on a sectoral level in order to account for inter-sectoral determinants.

this aspect the whole hypothesis can be explained very shortly. The author believes that individuals with different income levels tend to consume different bundles of goods. The more similar consumption bundles are the more trade there is potentially *as* consumption and domestic production are linked. Consequently he considers per capita income as well as within income distribution to be crucial for the demand structure. Compelling evidence for non-homothetic preferences associated with international trade is provided by Hunter and Markusen (1988), Hunter (1991), Fieler (2007), Dalgin et al. (2008) and Francois and Kaplan (1996). In contrast, the standard model for empirical trade – the gravity equation – focuses on aggregate GDP, hence homothetic preferences. In line with results from various gravity models that have been estimated I find that aggregate GDP levels are positively related to both the product-level as well as the country-level extensive margin of bilateral trade flows. By decomposing the logarithm of aggregate GDP into the logarithm of GDP per capita and population size I show that those two components impact trade patterns differently.[5] It is average income that is the relevant measure for trade flows. Moreover I find that a more dispersed income distribution increases trade flows at the extensive margin. In sum, my results provide evidence for non-homothetic preferences as I find that per capita income and the income distribution within a country are relevant determinants for both extensive margins of bilateral trade flows (and hence for consumption).

Among others HMR, Fieler (2007), Markusen (1986) and Melitz (2007) point out that trade patterns are different for so called North-North (NN), South-South (SS), North-South (NS) and South-North (SN) relations. This finding has two dimensions. On one hand it means that GDP levels of the exporter and importer are relevant for the resulting trade flow. As mentioned above my results show that per capita GDP levels are relevant determinants of trade patterns. On the other hand it should matter to what income group the the exporter and importer belong. I find that trade flows of two high income countries (NN) differ significantly from trade flows of a country pair in which either both or at least one country is a low or middle income country.[6] Linder himself stated that his theory is not applicable to developing countries. My results strongly reject this hypothesis as I find that the Linder hypothesis is especially applicable to poor countries. Or put differently the so called Linder term is largest[7] if North-North country pairs are excluded from the regression. Unlike previous studies I am really able to analyze the Linder hypothesis' explanatory power for trade patterns of poorer countries. To my knowledge the maximum number of countries used in an empirical test of the Linder hypothesis is 64, in Hallak (2008). This means that there has not been a *comprehensive* study of the Linder hypothesis which (i) provides evidence on the applicability to poorer countries, hence analyzes if the Linder hypothesis is a *global* phenomena or

[5] If homothetic preferences were at work the coefficients of per capita GDP and population size would need to have the same sign, standard error and they would need to be of the same amount.

[6] I use the classification of the World Bank which divides countries into low, middle and high income economies. In the last group there are 35 non-OECD and 30 OECD countries.

[7] in terms of the marginal effect on the extensive margin of trade when increasing the per capita income ratio by one standard deviation

only valid, if at all, for the richer part of the world and (ii) was able to explore if the explanatory power of the Linder hypothesis differs with respect to the income groups of the respective trading partners. This paper aims at filling those gaps in the literature by using a rich dataset, both in the cross section as well as the time series dimension. However, there are a number of studies of the Linder hypothesis which focus on developing countries. Arnon and Weinblatt (1998) find supportive evidence for the Linder hypothesis and also state that the positive relation between the similarity of per capita income and bilateral trade volume is even stronger between low income than high income countries. Additional evidence is shown by McPherson (2001) for five East African developing countries and by Chow (1999) for the Four Tigers in East Asia. Evenett and Venables (2002) show that developing countries could increase their exports substantially by an expansion of the product-level extensive margin.

I use a highly disaggregated trade dataset which is kindly provided by CEPII[8]. The dataset contains annual unidirectional tradeflows in 5'017 product cateories and is classified at the 6-digit level in Harmonized System (HS). The cross section contains 151 countries and the time series spans from 1995 until 2004.

Overall, with this paper I aim to capture more of the patterns of bilateral trade flows in order to provide information about appropriate assumptions for formal trade models. The rest of the paper is organized as follows. In section 2 I outline Linder's theory and show why the Linder hypothesis, which relates the closeness of per capita incomes and trade volume, is driven by the extensive margin of trade. Section 3 covers the data. The empirical strategy and the selection model of Heckman are described in the subsequent section. In section 5 I discuss the results and then conclude in the last section.

2 The underlying trade theory of Staffan Linder

The basic proposition of Linder (1961) is that domestic production and consumption is a necessary precondition for a good to be a potential export product. Hence, the range of exportable products is determined by internal demand. The main argument for the former is unfamiliarity with foreign markets as compared with the domestic market. The author argues with the following process. An entrepreneur will set up a business if he is *aware* of profit opportunities. The latter would tend to arise from *domestic* needs. [9] A successful firm would grow by producing for the domestic market. After some time the entrepreneur would become aware of profit opportunities in foreign countries and would potentially enter the export market to further expand. *Exporting is the end, not the beginning, of a typical market expansion path.* Moreover, products which are based on invention are likely to solve some problem

[8] see Gaulier and Zignago (2008)

[9] Thus Switzerland would not produce coast guard ships as the latter are not demanded in Switzerland. In turn Switzerland would not export coast guard ships.

which has been acute in the environment of the inventor. As such products suit the needs of the home market they would, in a first phase, only be sold domestically.[10] "In all, what our arguments amount to is the proposition that production functions are not identical in all countries, but that *the production functions of goods demanded at home are the relatively most advantageous ones.*" (Linder 1961, p. 90)

"It is self-evident that *internal demand determines which products may be imported.* We thus find that *the range of potential exports is identical to, or included in, the range of potential imports.*" (Linder 1961, p. 91)

"The more similar the demand structures of two countries, the more intensive, potentially, is the trade between these two countries." (Linder 1961, p. 94) Linder assumes the per capita income to be the most important determinant of the demand structure. As illustrated in Figure 1 he presumes a positive relationship between per capita income and the degree of quality of a product[11] demanded. The demand structure of a country is represented by a *range* of products which are demanded at home and hence are potential exports. Thus, with increasing similarity in average income the number of goods for which demand overlaps, and thus in which trade can be conducted, raises.[12] *This statement clearly indicates that the prediction Linder made for trade volumes is actually driven by the number of goods, i.e. the product-level extensive margin.* Surprisingly, this important aspect – or even the driving force behind the hypothesis – has been neglected in previous tests of Linder's theory.
Furthermore, the author assumes the skewness of the income distribution within a country to widen the range of potential exports and imports.[13] The reason is that high-income earners in a poor country may demand the same goods as low-income earners in a rich country.

Above I outlined what Linder described as *potential* trade. He introduced the notions trade-creating and trade-braking forces to link potential and actual trade. Already in 1961 Linder emphasized that the "strong elements of monopolistic competition should be incorporated into trade theory as a trade-creating force. (...) The almost unlimited scope for product differentiation – real or advertized – could, in combination with the seemingly unrestricted buyer idiosyncrasies, make possible flourishing trade in what is virtually the same commodity." (Linder 1961 p. 102) Thus, Linder's theory is an early forerunner to the 'new' trade theory of the past three decades. It is important to notice that Linder formulated his thesis *only for manufactures.*
He argued that the "distance factor" would push actual trade below potential trade. In this regard he mentioned spatial distance, transport costs, tariffs, cultural and

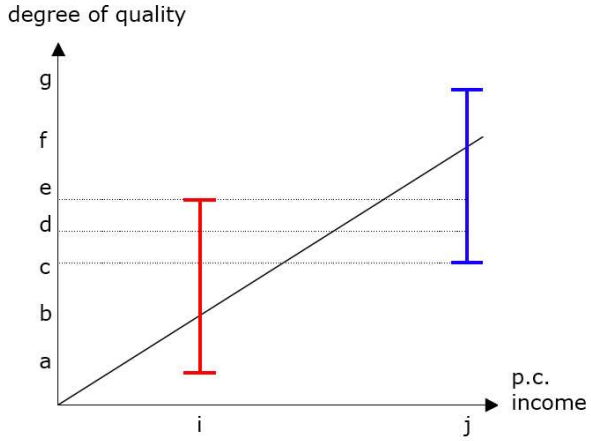
[10]Exceptions to the proposition are likely to occur when the obstacles to become aware of the foreign demand are low, the product does not need inventive effort and little product development work is needed.

[11]i.e. the *type* of a consumer good

[12]Thus, differences in per capita income are a potential obstacle to trade as goods in the production of which one country has a comparative advantage are not demanded in the other country.

[13]Thus, a greater overlapping of demands between countries with different per capita incomes than would be the case if incomes were more evenly distributed.

Figure 1: Overlap of demand structures



political affinities, language and the awareness of market opportunities. The author considered the distance factor to be especially distorting for underdeveloped coutries. He argued that in such countries "many entrepreneurs never raised their trade horizon very much above the local village market. Between such countries there could be hardly any foreign trade, no matter how similar the demand structures may be. Trade may instead be dominated by more aggressively marketed imports from industrialized countries and, to some extent, consist of products less suitable than those obtainable elsewhere." (Linder 1961, p. 108) Hence, Linder predicted SS and SN trade to be much lower than expected from similarity in demand structures.
Note that the trade theory of Linder is dynamic. Growing countries experience increasing per capita incomes and hence changing demand structures and an altered range of potential and actual exports. The following example is align with the theory. If Japan imported cars and exported bicycles, it might a dozen years later export cars and import bicycles.

From the above explanations it becomes clear that Linder's theory accounts for non-homothetic preferences. He does not only examine aggregate demand but analyzes subgroups of production – or varieties. The author states that "the more we divide total production into subgroups, the greater will be the variations in income elasticity. (...) small differences in income levels may produce substantial differences in the structure of demand." (Linder 1961, p. 95) In contrast, in a model with many differentiated product varieties and homothetic preferences (e.g. CES preferences) an increase in average income does not affect the structure of demand. Agents would consume more of every variety. Thus, consumption volumes would change but the structure of demand would not.
While there is still no widely accepted formal theory representing the Linder hypothesis some models have been proposed in recent years. Foellmi et al. (2008) and Ramezzana (2000) suggest models with non-homothetic preferences and indivisible goods. Higher per capita income makes consumers demand larger product variety.

Combined with increasing returns to scale on the production side this translates into a positive relationship between the similarity in per capita income and the product-level extensive margin of trade. However these are 2-country-models and hence do not provide a structural equation which could be estimated. Hallak (2008) introduces a formal derivation of the Linder hypothesis which focuses on the quality dimension of goods. As I do not concentrate on the quality dimension and would need a multi-country model I carefully translate the original trade theory of Staffan Linder into an empirical model.

Linder predicts the number of potential export goods to increase with the similarity in per capita income. This is the relationship on which I focus. Hence, the dependent variable is the extensive margin of the unidirectional trade flow ij[14] and the main explaining variable is an index for the similarity in average income – the so called Linder term. I determine the latter as the ratio of the per capita GDP of the poorer country to the per capita GDP of the richer country, $\in(0,1]$.[15] The skewness of the income distribution in country i and j are other determinants of the demand structures. A more skewed income distribution widens the range of potential exports and imports. This is approximated by the Gini coefficients of country i and j. Both Gini coefficients are expected to affect the extensive margin of the trade flow ij positively, however the impact of the Gini of country j should be larger.[16] The distance factor is accounted for by the inclusion of a set of variables which capture barriers and enhancers to trade, e.g. spatial distance, common language, affiliation of a common free trade agreement. I use the dataset of HMR. As Linder stated that it is not surprising that large countries trade more than small countries I control for the economic masses of the two trading partners. In order to shed more light on the different mechanisms of per capita GDP and population, I control for the per capita GDP level of i and j and the population size of i and j, rather than for aggregate GDP levels. In sum, Linder's trade theory suggests to include three types of variables: (i) an index for per capita income similarity and a proxy for within income distribution in order to approximate similarities in demand structures, (ii) controls for spatial distance, transport costs, cultural and political affinities and language to capture the "distance factor" and (iii) aggregate GDP levels to control for the fact that larger countries unsurprisingly trade more than smaller ones. As Linder formulated his thesis only for manufactures I only include product categories which include differentiated products according to Rauch's classification (conservative).
Additionally I make the following two extensions to the model. First, the trade theory of Foellmi et al. (2008), which is a formal derivation of the Linder hypothesis, motivates that more similar population sizes lead to more trade at the extensive margin. I include the ratio of the population size of the smaller country to the population size of the bigger country, $\in(0,1]$. Second, to test whether Linder was right

[14] I provide a detailed definition of the product-level extensive margin in the subsequent section.

[15] Many different Linder terms have been proposed so far. Most of them are related to the absolute difference in average income. My results are qualitatively the same when I use a traditional Linder term instead.

[16] Only a little demand in country j is needed to make a variety imported. However, such a small demand in country i would not be enough to make an entrepreneur in i produce the variety.

with the prediction that his theory is not applicable to developing countries and to test if the explanatory power of the Linder hypothesis is bigger when both trading partners are high income economies I split the sample into NN and SS/SN/NS.[17]

3 The Data

I use a highly disaggregated tradedataset which is kindly provided by CEPII.[18] The original data source is 6-digit UN COMTRADE data, classified in Harmonized System (HS). The dataset contains annual unidirectional tradeflows in 5'017 product categories. As I analyze one-way trade flows (unidirectional) every country pair is represented twice: once for exports from i to j and once for exports from j to i. The income data is taken from Penn World Tables 6.2 and data about barriers and enhancers to trade is kindly provided by Elhanan Helpman, Marc Melitz and Yona Rubinstein[19]. After matching those three data sources the cross-section covers 151 countries for the period 1995 until 2004. As Linder formulated his theory only for manufactures I include solely those HS codes belonging to James Rauch's differentiated products, 4-digit SITCs, in his conservative classification (see Rauch (1999)). The excluded products are those Rauch classified as reference priced or traded on organized exchanges. The mapping is successfull for 95% of the 5'017 HS 6-digit codes.

In order to capture more of the demand structure I use Gini coefficients to approximate the income distribution within countries. The most comprehensive dataset was constructed by David Dollar and Aart Kraay (2002). I follow their methodology to make Gini coefficients which reflect inequality in consumption, net and gross income comparable. I.e. I run a regression with dummies for the sources of inequality and use those coefficients to transform all Gini coefficients such that they reflect inequality in net income. Additionally, I linearly extrapolate Gini coefficients by country. The conditions to fill in a missing observation are (i) there need to be two known datapoints, one before the gap and one after the gap and (ii) the whole gap between two observed datapoints must not be larger than six years. These data manipulation is reasonable as Gini coefficients change slowly. In addition, virtually every empirical study employing inequality data does such adjustments as otherwise too few datapoints would be available for the analysis. In this paper the number of observation is cut into four when inequality data is included in regressions.

Distinction of zeros and missing values
The yearly trade matrix, at the country-level, can have three different types of observations in the N·(N-1) countries cells: a positive trade value, a zero if country i does not export anything to country j in a given year and a missing value if both countries do not report their trade statistic to the UN, hence we cannot observe this

[17] North: high income economies (nonOECD and OECD); South: low and middle income economies

[18] For a detailed description of the data and where to download it see Gaulier and Zignago (2008).

[19] At www.economics.harvard.edu/faculty/helpman/Data_Sets_Helpman they kindly provide the dataset used in "Estimating Trade Flows: Trading Partners and Trading Volumes", 2007.

datapoint. International trade datasets, which all trace back to UN COMTRADE, only include positive values. Hence, the researcher needs to figure out himself which of the initially empty cells are zeros and which are missing values. While systematic treatment of the bulk of country-level zeros has begun to emerge, the distinction of zeros and missing values has not attracted a lot of attention in the existing literature.[20] Some studies ignored this important distinction while others made the distinction incautiously. For example Kohler and Felbermayr (2004) treat all non-positive trade values as zeros. They state that according to Gleditsch (2002) 80% of all non-positive trade values are zero. The former statement is for example true for the CEPII data which covers 1995 until 2005. Yet it is far from general. I find for the dataset of Robert Feenstra[21] that the proportion of zeros of the non-positive trade values (i) is on average 60% and (ii) varies substantially over time. Fieler who uses Feenstra's data applies the list of 72 reporting countries corresponding to the period 1984–2000 to the period 1962–1983. But the data for the first period originates from a different source. I find that her procedure of distinguishing zeros from missing values results in a very different country-level extensive margin than when carefully utilizing all information contained in Feenstra's dataset.[22]
As UN COMTRADE indicates for every observation from which country the information is taken I can distinguish properly between zeros and non-observable datapoints. With the assumption that if a country does report in a given year, then it reports for all product categories and partner countries I can determine a list of reporting countries for every year. All trade flows between two non-reporting countries cannot be observed, hence are not a zero but a missing value. Out of 226'500 possible observations[23], more than 69'000 are zero and around 12'000 are non-observable.

Definition of the product-level extensive margin
Besides the country-level extensive margin, the product-level extensive margin of country i's exports to country j is the second measure of interest. It captures how many of the domestically produced varieties country i does export to country j. The simplest approach to construct this variable is to count the product categories in which exports ij are positive. The set of domestically produced varieties is not observed. In general it is possible to match export and production data, however not for a large cross-section. I use the definition of a product-level zero of Baldwin and Harrigan (2007). For exports ij, a zero occurs when country i exports an HS6 code to at least one country k but not to country j.[24] Thus, the measure for the extensive margin reflects how many varieties which i exports overall are exported to j. I normalize the number of exported varieties from i to j by the total quantity of product categories in which i exports to any country k.[25] Additionally, I wheight

[20] Amurgo-Pacheco and Pierola (2008) point out the importance of a careful distinction between zeros and missing values but do not deal with this issue in their study.

[21] World Trade Flows: 1962 – 2000, for a description see Feenstra et al. (2005).

[22] Feenstra uses whenever possible the information of the importer. Additionally, it is known if the trade value is reported by the importer or exporter.

[23] The trade matrix contains 22'650 (=151·150) cells per year, the time series is 10 years.

[24] As Hummels and Klenow (2005) I assume a strictly positive relationship between the number of produced and the number of exported varieties.

[25] This reflects something like an export intensity regarding the extensive margin. I do the nor-

the product categories with total export volumes of country i in variety v. This is similar to the methodology in Feenstra (1994). The classification of goods is somewhat arbitrary as it depends on tariff structures. With total export volumes I can assign each category its overall importance for the exports of country i.

The product-level extensive margin is defined as follows:

$$EM_{ijt} = \frac{\sum_{v \in V_{ijt}} p_{ikvt} \cdot x_{ikvt}}{\sum_{v \in V_{ikt}} p_{ikvt} \cdot x_{ikvt}}, \in (0, 1]$$

where V_{ijt} is the set of product categories in which country i exports to country j in year t, V_{ikt} is the set of product categories in which country i exports to any destination in year t[26], x_{ikvt} is the quantity which country i exports to country k in variety v in year t and p_{ikvt} is the corresponding price.

4 The empirical model

As OLS provides the best linear estimator for the conditional expectation without any distributional assumptions it is used as a benchmark model. However, as a standard OLS regression does not accomodate the bulk of country-level zeros the favored empirical model is the 2-step procedure of Heckman (1979) – the so called self-selection model.
About 40% of all possible trade relations at the country-level are inexistent. While many previous studies neglected the zeros by restricting the attention to positive trade flows I adapt the empirical strategy derived in HMR. The authors argue that "by disregarding countries that do not trade with each other, these studies give up important information contained in the data, and they produce biased estimates as a result". The first stage of the Heckman model consists of a Probit equation that estimates the probability that country i exports to j as a function of observable variables. The fitted values of the Probit equation are then used in the second stage to estimate the coefficients which are assumed to determine the product-level extensive margin of the bilateral trade flow ij. The selection model deals with an omitted variable bias and additionally provides information about the country-level extensive margin.

As described in Winkelmann and Boes (2006, page 213) the desicion whether to apply a Heckman (1979) or a Tobit[27] model is often a matter of interpretation of the underlying economic problem. If one considers exports as a *sequential decision* – first does country i have an interest to export to country j and second if the answer to the first question is yes in how many categories (or how much) does country i export to country j – a Heckman procedure with non-random sample selection is

malization as not all countries are able to produce in every category, e.g. due to absent resources.

[26] Country k is a reference country which I define to be the rest of the world.

[27] see Tobin (1956)

appropriate.[28] However, if an export value of zero is considered as a *natural choice* the corner solution model of Tobit is suitable. In the latter a zero is considered not to be fundamentally different from a positive value, i.e. the positive trade flows are a random sample.
For three reasons which I further outline below I prefer the Heckman model: (i) the censoring value is unknown, (ii) the censoring level is likely to vary across country-pairs[29] and (iii) the Heckman model is a more general version of the Tobit model as it allows the sign of coefficient n to be different in the first and second stage (with a Tobit model they have to be the same).
Even more importantly, the theory of Linder is somehow based on a sequential decision. He emphasizes that exporting is at the end of a market expansion path. An entrepreneur needs to survive a trial and error process at the home market in order to adjust the product such that it meets consumers demand. Exporting right from the beginning "would incur additional costs which might be destructive for his[30] effort to achieve the necessary comparative advantage" (Linder 1961, p. 90). This argumentation is somehow similar to the fixed export costs implemented in HMR. HMR derive the selection model of Heckman from their theoretical model.

Consider the following for the decision whether or not country i exports to country j. Country i will only export to country j if this is profitable. This statement is very general as one can think of country i alltogether.[31] As customary in economics one can consider this *overall export decision* of country i to country j as a cost benefit analysis. Country i exports to country j only if overall benefits of exporting will exceed overall costs. Hence, for country i there is a threshold value of overall costs of exporting to country j at which it is just profitable to export.[32] Of course, country i will only export to country j if the benefits of doing so exceed the costs. This threshold is unknown and likely to vary across destinations j, i.e. likely to be different for every trade flow ij.

Heckman formulated such a problem with two equations:

(1) behavioral equation: $y_{1i}^* = x_{1i}'\beta_1 + u_{1i}$
(2) selection equation: $y_{2i}^* = x_{2i}'\beta_2 + u_{2i}$

where x_{1i} and x_{2i} are vectors of regressors and β_1 and β_2 vectors of parameters. Equation (1) is the behavioral equation of interest and equation (2) is selection equation which determines if observation i has zero or positive value.

(3) $y_{1i} = y_{1i}^*$ if $y_{2i}^* > 0$ or

[28] The sequential decision translates directly into the two-step procedure for the Heckman model, see below.
[29] (i) and (ii) describe an incidental censoring process to which the Heckman model corresponds. The censoring level or threshold itself is a random variable.
[30] the entrepreneur
[31] One can also think of it as the sum of all the individual decisions of the consumers and firms.
[32] This can be considered as a zero profit level for the whole economy i.

$$(4) \quad y_{1i} = 0 \qquad \text{if } y^*_{2i} \leq 0$$

The above described threshold is related to y^*_{2i}. If overall benefits of country i to export to country j exceed the threshold value for the costs of trade flow ij y^*_{2i} is positive, otherwise negative.

If there is a selection problem – i.e. if observations with positive trade flows are not representative for the whole population (zero and positive trade flows) – the error terms u_{1i} and u_{2i} are correlated[33], hence a standard OLS specification leads to inconsistent estimators. The two-step estimation procedure of the selection model of Heckman is as follows: In the first stage a probit model is estimated.[34] The dependent variable is an indicator which equals one if positive trade occurs and zero otherwise – i.e. the country-level extensive margin. The regressors of the selection equation explain the probability that the unidirectional trade flow from country i to country j is positive. The second stage is a standard OLS model with the product-level extensive margin of trade as the dependent variable which however only includes observations with positive trade, i.e. the selected sample. It can be shown that including the *Inverse Mills Ratio (IMR)*[35], evaluated at the fitted values of the probit model, as an additional regressor in the behavioral equation of interest controls for the correlation of the two error terms. Thus standard regression procedures which only take into account positive trade flows suffer from omitted variable bias. When the IMR is not included in the behavioral equation the estimated coefficients for the vector β_1 are biased. If the vectors x_{1i} and x_{2i} include the same set of regressors the identification of the second stage estimates depends solely on the functional form of the IMR. As this is rather restrictive I prefer to make use of an exclusion variable which is a regressor of the selection equation but does not enter the behavioral equation. An exclusion variable needs to be correlated with the dependent variable of the first stage but uncorrelated with the residuals of the behavioral equation when the latter has been estimated with a reliable exclusion restriction. The first condition is fulfilled, however the second I cannot check. HMR have data about regulation costs which are assumed to represent fixed exporting costs and not to affect variable trade costs[36], hence this is assumed to be a valid exclusion variable. They show that their religion variable is a valid exclusion restriction too. The constellation of religion affiliation is assumed to affect if country i exports to country j but not the product-level extensive margin of this potential unidirectional trade flow.[37]

[33] This means that the behavioral equation of interest (1) and the selection equation (2) are not independent.

[34] The first stage corresponds to the selection equation and the second stage to the behavioral equation of interest.

[35] IMR $= \phi(\hat{I})/\Phi(\hat{I})$, where ϕ is the density function of the standard normal distribution, Φ the cumulative density function of the standard normal distribution and $\hat{I}$ is the predicted value of the of the dependent variable in the first stage.

[36] The theoretical model of HMR suggests that trade barriers that affect fixed trade costs but do not affect variable trade costs satisfy the conditions for an exclusion variable.

[37] This religion variable is not used in other studies, HMR constructed it especially to gain a plausible exclusion variable for their two-stage procedure.

The usual t-statistic of the coefficient of the IMR is a test for the nullhypothesis that there is no selection problem, i.e. that the selection and the behavioral equation are independent or that the selected sample is random. If the coefficient is significant there is a selection problem and a standard OLS specification would yield inconsistent estimates. The sign of the coefficient indicates if the selection is positive or negative. The selection model of Heckman needs the assumption of a bivariate normal distribution for the random disturbances.

Due to the numerous country-level zeros the selection model of Heckman is used. While the country-level zeros determine the dependent variable in the first stage, the product-level zeros determine the dependent variable in the second stage (the product-level extensive margin of trade flow ij).

5 Discussion of Results

Five types of variables are used to explain the two extensive margins of trade: (i) the relative per capita income (Linder term) and Gini coefficients of i and j to approximate the similarity in demand structures, (ii) per capita GDP level as well as population size of both trading partners to control for the fact that bigger[38] countries trade more, (iii) a set of variables which captures barriers and enhancers to trade (spatial distance, membership in WTO, freetrade agreement, common language etc.), (iv) the ratio of population sizes[39] and (v) a dummy for NN trade relations to test whether the explanatory power of the Linder hypothesis is bigger for high income countries and to test whether Linder was right with the prediction that his hypothesis is only applicable to developed countries.[40]
Overall most coefficients are highly significant and the empirical model derived from Linder's trade theory seems to fit the data very well.

In general I find that most variables that impact the amount of the product-level extensive margin also affect the probability that country i exports to country j, i.e. the country-level extensive margin. In almost all cases, the impact goes in the same direction. The test for a selection bias is the usual t-statistic of the Inverse Mills Ratio in the behavioral equation, which is always highly significant and positive. Hence, after controlling for similarity in demand structures, levels of income and barriers and enhancers to trade, the product-level extensive margin of trade relations which exhibit positive exports is significantly higher than of all trade relation (which may exhibit zero or positive exports). Alike the results of HMR, Hallak (2006) as well as Silva and Tenreyro (2006) I find a highly significant and positive selection bias.

[38] in terms of aggregate GDP level

[39] This is based on the theory of Foellmi et al. (2008) which constitutes a formal derivation of the Linder hypothesis.

[40] I split the sample into NN and SS/SN/NS trade, hence I interact the dummy for *NN* with every coefficient.

The Linder term – which captures similarity in per capita income – is positive and statistically highly significant in all specifications (see Table 3). This finding does not rely on the "new" definition of the Linder term. The results are qualitatively the same when I use a traditional Linder term.[41] This provides strong supportive evidence for the Linder hypothesis regarding the product-level extensive margin of trade. Closeness in per capita income does not only increase the product-level extensive margin of trade flow ij but also the probability that country i does export to country j – the country-level extensive margin. The inclusion of the squared Linder term indicates that the relationship between the extensive margin of trade and income similarity is rather concave than linear. The marginal effect is positive for 75% of the observations. Also the logarithm of the Linder term suggests a non-linear but strongly positive relationship between income similarity and the product-level extensive margin of trade.
The second determinant of the demand structure – the income distribution within a country – is more difficult to analyze due to lack of data. If Gini coefficients are included one is left with less than a fourth of the observations. Table 4 illustrates that only the Gini coefficient of the exporter is positive and significantly different from zero. However, if income levels are dropped both the exporter's and importer's Gini coefficient have a significantly positive impact on the extensive margin. Although the positive relationship between within income inequality and the extensive margin of trade is only obtained by dropping aggregate income levels as controls, this result is very interesting as it provides further evidence that trade patterns are also determined by demand-sided mechanisms. There are only very few studies which incorporate within income distribution to analyze the determination of trade patterns (e.g. Dalgin et al. (2008), Francois and Kaplan (1996)). Moreover, the findings for Gini coefficient of i and j are true for both extensive margins. I cannot confirm Linder's prediction that within inequality should increase the diversity of imported bundles of goods stronger than of exported budles. This could be due to supply-sided forces at work or due to lack of data.

To analyze the effects of aggregate GDP levels of both trading partners on the two extensive margins of trade in a first step a traditional gravity equation is suitable. I find that the "gravity forces" are at work for both extensive margins of trade (see Table 5). The gravity model explains higher trade by larger economic masses of i and j and lower distance. The decomposition of aggregate GDP into per capita GDP and population shows that those two components shed more light on the mechanisms of the extensive margin than simply the level of aggregate GDP. The richer[42] the trading partners, the higher the two extensive margins. However, population size tells a different story. A small country, in terms of population, does import a significantly more diverse bundle of goods. This effect can be explained with the supply side. Small countries have much smaller resources and cannot produce "enough" varieties domestically, that is why they need to import more product varieties. As far as varieties need to be invented, fixcosts can prevent small countries with low

[41] alternative Linder term: absolute difference in per capita GDP, normalized by the sum of per capita GDP of i and j

[42] in terms of per capita GDP

resources from inventing new (or many) products. However, the population size of the destination country does not influence the probability that i exhibits positive exports to j. The results suggest bigger countries, in terms of population size, to export a more diverse bundle of good as well as to have a higher probability of exporting. These results show that the aggregate GDP does not tell the whole story as the two components per capita GDP and population size affect trade differently. Moreover, they indicate that the covariates not always affect the two extensive margins similarly. The above described differing effects of poulation and average income levels are very robust to many different specifications.

In all regressions I control for a set of variables which approximate trade costs. It is a robust result in the literature that more distant countries and country pairs which are jointly not a member of WTO trade less while a common official language, a free trade agreement, a currency union, common membership in the WTO, a common legal origin and colonial ties increase trade. My results show that those effects do also apply to both extensive margins of trade.

The theory of Foellmi et al. (2008) motivates that more similar population sizes lead to more trade at the product-level extensive margin. I find strong supportive evidence for this effect (see Table 6). It is very robust and at work at both extensive margins. Inclusion of this index for population size similarity alters the other coefficients quantitatively only marginally.

In order to analyze if the Linder hypothesis (i) is only applicable to developed countries as Linder himself predicted and (ii) if the explanatory power of similarity in per capita income for the extensive margin of trade flows is bigger when both trading partners are high income economies I split the sample into NN and SS/NS/SN (see table 7 and 8). The previous results actually already reject (i). The dataset includes 151 countries – 46 are high income economies and 105 are low (42) or middle (61) income economies. As the previous results (which include all 151 countries) show a strong and highly significant Linder type relationship for trade flows they suggests that the Linder hypothesis is a global phenomena, hence not only at work in the rich part of the world. Table 3 shows the "baseline results" for all 151 countries, in table 7 the same regressions are estimated for NN country pairs separately and in table 8 NN country pairs are excluded from the sample. In all three tables I report beta coefficients. A 1 standard deviation increase in the per capita GDP ratio increases the product-level extensive margin by 0.06 when NN trade flows are excluded from the analysis. The coefficient is halved, namely 0.03, when the regression is run for NN country pairs separately. Both coefficients are highly significant. These findings are contrary to what Linder himself predicted and what many would maybe expect. Note that the Linder term as well as per capita GDP levels are insignificant when only high income countries are considered. Overall these results suggest that the Linder hypothesis is especially applicable to poor countries and a global phenomena. Linder himself stated that this is not the case as in developing countries "many entrepreneurs never raised their trade horizon very much above the local village market". Though in the last 40 to 50 years patterns of trade changed

a lot and developing countries do not only export primary goods but have started to produce and export finished goods of various types. My results are in lign with those of Arnon and Weinblatt (1998), i.e. that the positive relationship between similarity in per capita income and intensity of trade is even stronger in the poorer part of the world.

I analyzed the above posed questions additionally with the Feenstra trade dataset.[43] This alternative trade datasource is organized at the 4-digit level in the Standard International Trade Classification. 138 countries can be matched with income data and information about barriers to trade. The time series lasts from 1962 until 2000. The application of the Feenstra trade data provides qualitatively the same results as if the CEPII data is used. This makes my results very robust as it indicates that the findings do not rely on (i) the product classification or the corresponding level of disaggregation[44] or (ii) the chosen dataset. Moreover, I tested if the results rely on the fact that I wheight categories with total exports. This is not the case as the findings are very similar if varieties are not wheighted at all and hence are simply counted. Furthermore, the results are qualitatively the same if the dependent variable of the second stage is calculated as the logarithm of the product-level extensive margin.

In general I find that the product-level extensive margin is driven partly by the same determinants as trade volumes. To the extent that volumes and the extensive margin are affected similarly this means that determinants, at least partly, affect trade volumes only *indirectly*, namely via the product-level extensive margin of trade. And vice versa it is plausible to hypothesize that my findings would also apply, to some extent, to bilateral trade volumes.

6 Conclusion

An Essay on Trade and Transformation written by Staffan Linder is an early forerunner to the 'new' trade theory of the past three decades. Numerous studies tested Linder's trade theory empirically. While most early tests could not confirm the hypothesis, more recent studies provide supportive evidence for it. However, the driving force behind the Linder hypothesis – namely the product-level extensive margin of trade – has been neglected in the literature so far. Linder (1961) predicted the intensity of trade to increase with the similarity in demand structures. He assumed countries with similar demand patterns to develop similar industries. Increasing similarity in demand structures would raise the number of goods for which demand overlaps, and thus in which trade can be conducted. Nations with similar demand patterns would trade with each other in similar but differentiated goods. This clearly indicates that the prediction Linder made for trade volumes is actually driven by

[43] See Feenstra et al. (2005) for a description.

[44] CEPII data is organized at the 6-digit level in the Harmonized System and Feenstra's data at the 4-digit level in the Standard International Trade Classification.

the number of product varieties in which trade takes place, i.e. the product-level extensive margin. Put differently, the author believes that individuals with different income levels tend to consume different bundles of goods. The more similar consumption bundles are the more trade there is potentially if consumption and domestic production are linked. Hence non homothetic preferences are accounted for. Per capita income is the most important determinant of demand and as a close second comes the income distribution within a country. I find strong supportive evidence that the diversification of bilaterally traded bundles of goods enlarges with increasing similarity in per capita income.[45] Furthermore, I find that countries with a more uneven income distribution export in significantly more product varieties. If GDP levels are dropped from the specification, the skewness of the income distribution of both trading partners increases the extensive margin of trade. Linder explains such an effect with broader demands and hence a larger overlap with the demand structure of the trading partner. *In sum, I find strong supportive evidence that the two characteristics for similarity in demand structures – between per capita income distribution and within income distribution – are positively correlated with the product-level extensive margin, i.e. my results clearly confirm the Linder hypothesis and provide strong evidence for non homothetic preferences.* Or put differenty, my findings suggest that for the determination of trade patterns important demand side effects are at work. Moreover, I find that countries with more similar demand structures do not only trade more along the product-level extensive margin but have also a higher probability to exhibit positive trade. Thus, the demand sided effects are at work at *both* extensive margins of trade.

About 40% of all possible trade relations at the country-level are inexistent. While many previous studies neglected the zeros by restricting the attention to positive trade flows I adapt the empirical strategy derived in HMR. The authors argue that "by disregarding countries that do not trade with each other, these studies give up important information contained in the data, and they produce biased estimates as a result" and derive an empirical strategy which incorporates the self-selection model of Heckman. The first stage of the Heckman model consists of a Probit equation that estimates the probability that country i exports to j – the country-level extensive margin. The fitted values of the Probit equation are then used in the second stage to estimate the coefficients which are assumed to determine the product-level extensive margin of the bilateral trade flow ij. I find a statistically significant and positive selection bias. There are some effects which impact the country-level and the product-level extensive margin differently. Overall my findings illustrate the importance of the country-level extensive margin. Hence, to understand the dynamics of international trade flows a researcher needs to consider both the country-level as well as the product-level extensive margin of trade.

Furthermore, I find that the "gravity forces" are also at work for the determination of the extensive margin of bilateral trade flows. I.e. that trade increases with larger

[45] It is possible that earlier results which focused on trade values are mixed because they neglected the product-level extensive margin of trade. Helpman et al. (2008) find that estimators for bilateral trade volumes are biased if the extensive margin is not controlled for.

economic masses of the trading partners and lower distance. The decomposition of aggregate GDP into per capita GDP and population size shows that those two components affect the extensive margin differently. Richer countries, in terms of average GDP, trade more. However, bigger countries, in terms of population, import a significantly less diverse bundle of goods. Small countries have fewer resources and cannot produce "enough" varieties domestically. Hence, they need to import more product varieties. However, the population size of the destination country j does not influence the probability that i exhibits positive exports to j. The results suggest bigger countries, in terms of population size, to export a more diverse bundle of good as well as to have a higher probability of exporting. In sum, those findings provide strong evidence for non homothetic preferences. It is the per capita and not the aggregate GDP that is the relevant determinant for trade flows. These results illustrate that it is important (i) to analyze the two components of aggregate GDP – per capita GDP and population size – as they have differing impacts on trade and (ii) to distinguish between the country-level and the product-level extensive margin. In general I find that the product-level extensive margin is driven partly by the same determinants as trade volumes. To the extent that volumes and the extensive margin are affected similarly this means that those determinants, at least partly, affect trade volumes only indirectly, namely via the product-level extensive margin of trade. And vice versa it is plausible to hypothesize that my findings would also apply, to some extent, to bilateral trade volumes.

I use a very comprehensive and rich dataset. It covers unidirectional bilateral trade flows between 151 countries on the 6-digit level from 1995 until 2004. To my knowledge the maximum number of countries used in an empirical test of the Linder hypothesis is 64, in Hallak (2008). As previous studies restricted their analysis to the richer part of the world it was not possible to study if the Linder effect is a general phenomena or only, if at all, valid for the richer part of the world. This paper is a *comprehensive* study. With a cross-section of 151 countries I find a highly significant Linder effect, hence strong evidence that the latter is a global phenomena. I even find that the relationship between similarity in per capita income and the extensive margin is stronger when high income economies are excluded from the analysis. Furthermore the results provide strong evidence against Linders prediction that his hypothesis is only valid for developed countries.
Overall, my findings are robust to (i) the usage of an alternative trade data set, (ii) several specifications of the regression equation, (iii) alternative definitions of variables and (iv) the application of a standard OLS regression instead of the selection model which relies on distributional assumptions.

The partitioning of bilateral trade values into the extensive and intensive margin needs to be determined. HMR as well as Hummels and Klenow (2005) show that the extensive margin captures a substantial part of bilateral trade volumes. However, those studies take a supply sided approach. Such a partitioning together with a demand oriented approach would shed more light on that issue. Another important dimension is the quality of traded products, e.g. studied by Juan Carlos Hallak but otherwise it attracted few attention in the literature.

References

Amurgo-Pacheco, A. & Pierola, M. D. (2008). Patterns of Export Diversification in Developing Countries: Intensive and Extensive margins. Policy Research Working Paper Series 4473, The World Bank.

Arnon, A. & Weinblatt, J. (1998). Linder's Hypothesis Revisited: Income Similarity Effects for Low Income Countries. *Applied Economics Letters*, *5*(10), 607–11.

Baldwin, R. & Harrigan, J. (2007). Zeros, Quality and Space: Trade Theory and Trade Evidence, NBER Working Paper Nr. 13214.

Choi, C. (2002). Linder Hypothesis Revisited. *Applied Economics Letters*, *9*(9), 601–05.

Chow, P., Kellman, M., & Shachmurove, Y. (1999). A Test of the Linder Hypothesis in Pacific NIC Trade 1965-1990. *Applied Economics*, *31*(2), 175–82.

Dalgin, M., Trindade, V., & Mitra, D. (2008). Inequality, Nonhomothetic Preferences, and Trade: A Gravity Approach. *Southern Economic Journal*, *74*(3), 747–74.

Dollar, D. & Kraay, A. (2002). Growth Is Good for the Poor. *Journal of Economic Growth*, *7*(3), 195–225.

Evenett, S. & Venables, A. (2002). Export Growth in Developing Countries: Market Entry and Bilateral Trade Flows, London School of Economics. Mimeo.

Feenstra, R. (1994). New Product Varieties and the Measurement of International Prices. *American Economic Review*, *84*(1), 157–77.

Feenstra, R., Lipsey, R., Haiyan, D., Ma, A., & Mo, H. (2005). World Trade Flows: 1962-2000, NBER Working Paper Nr. 11040.

Fieler, A. C. (2007). Non-Homotheticity and Bilateral Trade: Evidence and a Quantitative Explanation, New York University. Mimeo.

Fillat-Castejón, C. & Serrano-sanz, J. M. (2004). Linder Revisited: Trade and Development in the Spanish Economy. *International Review of Applied Economics*, *18*(3), 323–348.

Foellmi, R., Hepenstrick, C., & Zweimueller, J. (2008). Income Effects in the Theory of Monopolistic Competition and International Trade, University of Zurich. Mimeo.

Francois, J. & Kaplan, S. (1996). Aggregate Demand Shifts, Income Distribution, and the Linder Hypothesis. *Review of Economics and Statistics*, *78*(2), 244–250.

Gaulier, G. & Zignago, S. (2008). BACI: A World Database of International Trade Analysis at the Product-level, CEPII Working Paper.

Gleditsch, K. S. (2002). Expanded Trade and GDP Data. *Journal of Conflict Resolution*, *46*(5), 712–724.

Glick, R. & Rose, A. (2002). Does a Currency Union Affect Trade? The Time Series Evidence. *European Economic Review of Economics and Statistics*, *46*, 1125–1151.

Hallak, J. C. (2006). Product quality and the direction of trade. *Journal of International Economics*, *68*(1), 238–265.

Hallak, J. C. (2008). A Product-Quality View of the Linder Hypothesis. *Forthcoming in The Review of Economics and Statistics*.

Heckman, J. (1979). Sample Selection as a Specification Error. *Econometrica*, *47*, 153–161.

Helpman, E., Melitz, M., & Rubinstein, Y. (2008). Estimating Trade Flows: Trading Partners and Trading Volumes. *Quarterly Journal of Economics*, *123*, 441–487.

Hummels, D. & Klenow, P. (2005). The Variety and Quality of a Nation's Exports. *American Economic Review*, *95*, 704–23.

Hunter, L. (1991). The Contribution of Nonhomothetic Preferences to Trade. *Journal of International Economics*, *30*(4), 345–358.

Hunter, L. & Markusen, J. (1988). Per Capita Income as a Basis for Trade, in Robert Feenstra, Empirical Methods for International Trade, Cambridge: MIT Press.

Kohler, W. & Felbermayr, G. (2004). Exploring the Intensive and Extensive Margins of World Trade. *CESifo Working Paper Series No. 1276*, *September*.

Krugman, P. (1980). Scale Economies, Product Differentiation, and the Pattern of Trade. *American Economic Review*, *70*, 950–59.

Linder, S. (1961). *An Essay on Trade and Transformation*. Uppsala: Almqvist and Wiksells.

Markusen, J. (1986). Explaning the Volume of Trade: An Eclectic Approach. *American Economic Review*, *76*, 1002–1011.

McPherson, M. A., Redfearn, M. R., & Tieslau, M. (2001). International Trade and Developing Countries: An Empirical Investigation of the Linder Hypothesis. *Applied Economics*, *33*(5), 649–57.

Mcpherson, M. A., Redfearn, M. R., & Tieslau, M. A. (2000). A Re-Examination Of The Linder Hypothesis: A Random-Effects Tobit Approach. *International Economic Journal*, *14*(3), 123–136.

Melitz, J. (2007). North, South and distance in the gravity model. *European Economic Review*, *51*(4), 971–991.

Melitz, M. (2003). The Impact of Trade on Intra-industry Reallocations and Aggregate Industry Productivity. *Econometrica*, *71*, 1695–1725.

Ramezzana, P. (2000). Per Capita Income Demand for Variety, and International Trade: Linder Reconsidered. CEP Discussion Papers 0460, Centre for Economic Performance, LSE.

Rauch, J. E. (1999). Networks Versus Markets in International Trade. *Journal of International Economics*, *48*(1), 7–35.

Rose, A. (2000). One Money, One Market: The Effect of Common Currencies on Trade. *Economic Policy*, *75*(30), 7–33.

Rose, A. (2004). Do We Really Know that the WTO Increases Trade? *American Economic Review*, *94*, 98–114.

Silva, S. & Tenreyro, S. (2006). The Log of Gravity. *Review of Economics and Statistics*, *88*, 641–58.

Thursby, J. & Thursby, M. (1987). Bilateral Trade Flows, the Linder Hypothesis, and Exchange Risk. *Review of Economics and Statistics*, *69*, 488–95.

Tobin, J. (1956). Estimation of relationships for limited dependent variables. Cowles Foundation Discussion Papers 3R, Cowles Foundation, Yale University.

Winkelmann, R. & Boes, S. (2006). *Analysis of Microdata*. Berlin: Springer-Verlag.

7 Appendix

7.1 Definition of variables

The dependent variable

- *Extensive Margin*: Trade data come from UN COMTRADE, classified in Harmonized System (Version 1992), 6-digit. 5'017 different product categories (HS6 codes) are included. The product-level extensive margin is defined as

$$EM_{ijt} = \frac{\sum_{v \in V_{ijt}} p_{ikvt} \cdot x_{ikvt}}{\sum_{v \in V_{ikt}} p_{ikvt} \cdot x_{ikvt}}, \in (0, 1].$$

  Where V_{ijt} is the set of product categories in which country i exports to country j in year t, V_{ikt} is the set of product categories in which country i exports to any destination in year t[46], x_{ikvt} is the quantity which country i exports to country k in variety v in year t and p_{ikvt} is the corresponding price.

Proxies for demand structure

Gross Domestic Product (GDP) and per capita GDP data come from Penn World Tables 6.2.; GDP data is real – i.e. constant prices – and PPP converted. Aggregated GDP is in millions.

- *Relative per capita income*:

$$\text{Rel. p.c. GDP} = \frac{\text{p.c. GDP of poorer country in a given country pair } ij}{\text{p.c. GDP of richer country in a given country pair } ij}, \in (0, 1)$$

- *Traditional Linder term*:

$$\text{Traditional Linder term} = \frac{|\text{p.c. GDP } i \text{ - p.c. GDP } j|}{\text{p.c. GDP } i \text{ + p.c. GDP } j}, \in (0, 1)$$

- *Gini coefficient*: Gini coefficients come from WIID (Version 2c), Dollar Kraay and Milanovic.

- *Population*: Population data come from Penn World Tables 6.2; in thousands.

$$\text{Rel. population} = \frac{\text{Population of smaller country in a given country pair } ij}{\text{Population of bigger country in a given country pair } ij}, \in (0, 1)$$

Proxies for "Distance factor"

- *Distance*: the distance (in km) between importers j and exporters i capitals (in logs). Source: HMR.

- *WTO (both)*: A binary variable that equals one if both countries belong to the GATT/WTO, and zero otherwise. Source: HMR and WTO.

[46] Country k is a reference country which I define to be the rest of the world.

- *WTO (none)*: A binary variable that equals one if both countries do not belong to the GATT/WTO, and zero otherwise. Source: HMR and WTO.
- *Free Trade Agreement (FTA)*: A binary variable that equals one if exporting country j and importing country i belong to a common regional trade agreement, and zero otherwise. Source: HMR and WTO.[47]
- *Currency Union (CU)*: A binary variable that equals one if importing country j and exporting country i use the same currency or if within the country pair money was interchangeable at a 1:1 exchange rate for an extended period of time (see Rose (2000),(2004) as well as Glick und Rose (2002)), and zero otherwise. Source: HMR.
- *Common Border*: A binary variable that equals one if importer j and exporter i are neighbors that meet a common physical boundary, and zero otherwise. Source: HMR.
- *Legal System*: A binary variable that equals one if the importing country j and exporting country i share the same legal origin, and zero otherwise. Source: HMR.
- *Common Language*: A binary variable that equals one if importer j and exporter i have at least one common official language, and zero otherwise. Source: HMR.
- *Colonial Ties*: A binary variable that equals one if importing country j ever colonized exporting country i or vice versa, and zero otherwise. Source: HMR.
- *Island*: A binary variable that equals one if both importer j and exporter i are islands, and zero otherwise. Source: CIA World Factbook.
- *Landlocked*: A binary variable that equals one if both importing country j and exporting country i have no coastline or direct access to sea, and zero otherwise. Source: CEPII.[48]
- *Religion:* A wheighted average which reflects population shares of protestants, catholics and muslims, $\in (0,1)$. Source: HMR.

$$\begin{aligned}\text{Religion} = \quad & (\%\text{ Protestants in country } i \cdot \%\text{ Protestants in country } j) \\ & +(\%\text{ Catholics in country } i \cdot \%\text{ Catholics in country } j) \\ & +(\%\text{ Muslims in country } i \cdot \%\text{ Muslims in country } j)\end{aligned}$$

Country Groups

Countries are divided up into two groups – North and South or high income and developing countries. The classification comes from the World Bank – low income

[47] www.wto.org/english/tratop e/region_e/region_e.htm
[48] http://www.cepii.fr/anglaisgraph/bdd/distances.htm

economies, middle income economies and high income economies.[49] Low and middle income economies are combined to the group "developing countries" or "South".

- *North-North (NN)*: A binary variable that equals one if both importer j and exporter i are a high-income country.
- *South-South (SS)*: A binary variable that equals one if both importer j and exporter i are a developing country (low or middle income country).
- *North-South (NS)*: A binary variable that equals one if the importer j is a developing country and the exporter i is a high-income country.
- *South-North (SN)*: A binary variable that equals one if the importer j is a high income country and the exporter i is a developing country.

Selection Bias

- *IMR*: The Inverse Mills Ratio ist the ratio of the density function and cumulative density function of a distribution. The standard normal distribution is used:

$$IMR = \frac{\phi(z)}{\Phi(z)}$$

Table 1: Summary Statistics

	observations	mean	median	standard deviation	min	max
ALL OBSERVATIONS						
Relative per capita GDP	177'151	0.363	0.278	0.281	0.001	0.999
Extensive Margin	177'151	0.144	0.013	0.249	0	0.999
Number of bilaterally exported HS6 codes	177'151	195.275	7	461.440	0	2'876
Number of multilaterally exported HS6 codes	177'151	1'795.494	2'464	1'197.620	0	2'897
OBSERVATIONS WHICH EXIBIT POSITIVE EXPORTS						
Relative per capita GDP	126'595	0.367	0.281	0.285	0.001	0.999
Extensive Margin	126'595	0.201	0.059	0.274	0.000	0.999
Number of bilaterally exported HS6 codes	126'595	273.258	31	525.976	1	2'876
Number of multilaterally exported HS6 codes	126'595	2'512.528	2'688	453.323	286	2'897

6-digit UN COMTRADE data, classified in Harmonized System (Version 1992), includes 5'017 different product categories (HS6 codes). xxx thereof are differentiated products according to James Rauch's classification.

[49] http://go.worldbank.org/D7SN0B8YU0

Table 2: List of Countries

Name	NBER/ UN Code	Name	NBER/ UN Code	Name	NBER/ UN Code
Afghanistan	450040	Monaco		French Polynesia,	
Albania	580080	Gabon	142660	Vanuata	
Algeria	130120	Gambia	162700	New Zealand	715540
Angola	160240	Germany	532800	Nicaragua	345580
Argentina	330320	Ghana	162880	Niger	165620
Australia	710360	Greece	533000	Nigeria	165660
Austria	550400	Greenland*	223040	Norway	555780
Bahamas	350440	Guatemala	343200	Oman	445120
Bahrain	440480	Guinea	163240	Pakistan	455860
Bangladesh	450500	Guinea-Bissau,	166240	Panama	365900
Barbados	350520	Cape Verde		Papua New Guin.	725980
Belgium,	530560	Guyana	363280	Paraguay	336000
Luxembourg		Haiti	353320	Peru	336040
Belize	360840	Honduras	343400	Philippines	456080
Benin	162040	Hong Kong	453440	Poland	586160
Bermuda	220600	Hungary	583480	Portugal	536200
Bhutan	450640	Iceland	553520	Qatar	446340
Bolivia	330680	India	453560	Romania	586420
Brazil	330760	Indonesia, Macau	453600	Rwanda	166460
Brunei	450960	Iran	443640	Saudi Arabia	446820
Bulgaria	581000	Iraq	443680	Senegal	166860
Burkina Faso	168540	Ireland	533720	Seychelles	166900
Burundi	161080	Israel	413760	Sierra Leone	166940
Cambodia	451160	Italy	533800	Singapore	457020
Cameroon	141200	Jamaica	353880	Somalia	167060
Canada	211240	Japan	413920	Southern African	117100
Cayman Islds*	351360	Jordan	444000	Customs Union	
Central African Rep	141400	Kenya	164040	Spain	537240
Chad	141480	Kiribati, Solomon	722960	Sri Lanka	451440
Chile	331520	Tonga, Tuvalu		Sudan	137360
China	481560	Korea North	484080	Surinam	367400
Colombia	331700	Korea South	454100	Sweden	557520
Comoros	161740	Kuwait	444140	Switzerland	557560
Congo	141780	Laos	454180	Syria	447600
Costa Rica	341880	Lebanon	444220	Tanzania	168340
Cote D'Ivoire	163840	Liberia	164300	Thailand	457640
Cuba	351920	Libya	134340	Togo	167680
Cyprus	441960	Madagascar	164500	Trinidad-Tobago	357800
Czech Republic,	582000	Malawi	164540	Tunisia	137880
Slovakia		Malaysia	454580	Turkey	447920
Denmark, Faroe	532080	Maldives	454620	Turks Caicos Isl*	357960
Islds		Mali	164660	Uganda	168000
Djibouti	162620	Malta	574700	United Kingdom	538260
Dominican Republic	352140	Mauritania	164780	United Arab Emir.	447840
Ecuador	332180	Mauritius	164800	Uruguay	338580
Egypt	138180	Mexico	334840	USA	218400
El Salvador	342220	Mongolia	484960	Venezuela	338620
Equatorial Guinea	162260	Morocco	135040	Vietnam	487040
Ethiopia	162300	Mozambique	165080	Western Sahara*	137320
Fiji	722420	Myanmar (Burma)*	451040	Yemen	448870
Finland	552460	Nepal	455240	Zambia	168940
Former Ussr	688100	Netherlands Antilles	355320	Zimbabwe	167160
Former Yugoslavia	598900	Netherlands	535280		
France, Andorra,	532500	New Caledonia,	725400		

* As PWT 6.2 does not provide GDP data for those countries they are not included in the regressions. Later alternative income data sources will be considered for those countries.

Table 3: Baseline Results

	Benchmark excl. zeros	Benchmark incl. zeros	Selection equation	Behavioral equation	Behavioral equation	Selection equation	Behavioral equation	Selection equation	Behavioral equation
Relative p.c. GDP	0.0830***	0.0978***	0.0468***	0.0567***	0.1811***				
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)				
Relative p.c. GDP squared					-0.1223***				
					(0.000)				
ln(Relative p.c. GDP)						0.0154***	0.0806***		
						(0.000)	(0.000)		
Traditional Linder term								-0.0493***	-0.0652***
								(0.000)	(0.000)
ln(GDP p.c. importer)	0.0424***	0.0473***	0.0114*	0.0301**	0.0187	0.0074	-0.0041	0.0107*	0.0249*
	(0.007)	(0.000)	(0.059)	(0.032)	(0.191)	(0.213)	(0.775)	(0.075)	(0.078)
ln(GDP p.c. exporter)	0.0307*	0.0467***	0.0164***	0.0535***	0.0412**	0.0138**	0.0185	0.0159**	0.0478***
	(0.072)	(0.001)	(0.009)	(0.002)	(0.017)	(0.025)	(0.278)	(0.011)	(0.005)
ln(Population importer)	-0.1486**	-0.2137***	0.0049	-0.2637***	-0.2584***	0.0083	-0.2118***	0.0071	-0.2428***
	(0.011)	(0.000)	(0.790)	(0.000)	(0.000)	(0.646)	(0.000)	(0.697)	(0.000)
ln(Population exporter)	0.5103***	0.5022***	0.0940***	0.5812***	0.5900***	0.0982***	0.6392***	0.0962***	0.6029***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Distance	-0.2928***	-0.2159***	-0.1112***	-0.4109***	-0.4082***	-0.1104***	-0.4062***	-0.1107***	-0.4091***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
WTO (both)	0.0312***	0.0244***	0.0200***	0.0405***	0.0401***	0.0197***	0.0401***	0.0200***	0.0404***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
WTO (none)	-0.0003	0.0037**	-0.0104	-0.0019	-0.0028	-0.0111*	-0.0036*	-0.0105*	-0.0024
	(0.869)	(0.021)	(0.106)	(0.326)	(0.155)	(0.082)	(0.066)	(0.099)	(0.224)
FTA	0.0713***	0.1035***	0.0367***	0.0472***	0.0464***	0.0358***	0.0466***	0.0363***	0.0468***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Currency Union	0.0040**	-0.0070***	0.0909***	0.0269***	0.0257***	0.0900***	0.0238***	0.0904***	0.0257***
	(0.022)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Land border	0.0397***	0.0476***	-0.0124	0.0557***	0.0560***	-0.0134	0.0554***	-0.0126	0.0555***
	(0.000)	(0.000)	(0.296)	(0.000)	(0.000)	(0.258)	(0.000)	(0.286)	(0.000)
Legal	0.0619***	0.0494***	0.0144***	0.0608***	0.0608***	0.0146***	0.0610***	0.0144***	0.0607***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Language	0.0502***	0.0316***	0.0466***	0.1074***	0.1076***	0.0462***	0.1080***	0.0463***	0.1075***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Island	0.0161***	0.0083***	0.0215***	0.0262***	0.0268***	0.0196***	0.0254***	0.0208***	0.0261***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Religion			0.0627***			0.0628***		0.0627***	
			(0.000)			(0.000)		(0.000)	
IMR				0.3907***	0.3858***		0.3829***		0.3882***
				(0.000)	(0.000)		(0.000)		(0.000)
Observations	126595	177151	166507	115951	115951	166507	115951	166507	115951
R^2 adjusted	0.861	0.822		0.873	0.874		0.874		0.874

* p-value < 0.10, ** p-Wert < 0.05, *** p-Wert < 0.01
The product-level extensive margin is the dependent variable in the benchmark specification (standard OLS) and the behavioral equation (2nd stage in selection model). The country-level extensive margin is the dependent variable in the selection equation (1st stage in selection model). Robust p-values in brackets. Exporter, importer and time fixed effects. The selection equation is a Probit model, marginal effects at sample means are reported, for dummy variables the discrete change from 0 to 1 is reported. The behavioral equation is an OLS regression, beta coefficients are reported. Religion is excluded variable in second stage specifications.

Table 4: Within Income Distribution – Gini Coefficients

	Selection equation	Behavioral equation	Selection equation	Behavioral equation
Relative p.c. GDP	0.0146***	0.0544***	0.0174***	0.0597***
	(0.000)	(0.000)	(0.000)	(0.000)
Gini importer	0.0001	0.0001	0.0009***	0.0043***
	(0.219)	(0.669)	(0.000)	(0.000)
Gini exporter	0.0002*	0.0005**	0.0010***	0.0043***
	(0.064)	(0.011)	(0.000)	(0.000)
ln(GDP p.c. importer)	0.0069	0.0043		
	(0.180)	(0.703)		
ln(GDP p.c. exporter)	0.0103**	0.0868***		
	(0.046)	(0.000)		
ln(Population importer)	-0.0181*	-0.1063***		
	(0.060)	(0.000)		
ln(Population exporter)	0.0195**	0.0883***		
	(0.036)	(0.000)		
Distance	-0.0148***	-0.1243***	-0.0125***	-0.1156***
	(0.000)	(0.000)	(0.000)	(0.000)
WTO (both)	0.0016	0.0039	0.0049**	0.0158***
	(0.290)	(0.283)	(0.011)	(0.000)
WTO (none)	-0.0015	-0.0210**	-0.0023	-0.0254***
	(0.622)	(0.017)	(0.492)	(0.004)
FTA	0.0069***	0.0290***	0.0091***	0.0368***
	(0.003)	(0.000)	(0.000)	(0.000)
Currency union	0.0085***	0.0196*	0.0096***	0.0204**
	(0.000)	(0.051)	(0.000)	(0.046)
Land border	0.0048	0.1271***	0.0082***	0.1381***
	(0.107)	(0.000)	(0.000)	(0.000)
Legal	0.0008	0.0356***	0.0008	0.0360***
	(0.227)	(0.000)	(0.287)	(0.000)
Language	0.0078***	0.0452***	0.0096***	0.0566***
	(0.000)	(0.000)	(0.000)	(0.000)
Island	0.0016	0.0306***	0.0015	0.0292***
	(0.381)	(0.000)	(0.470)	(0.000)
Landlock	0.0012	0.0106**	0.0002	0.0113**
	(0.475)	(0.018)	(0.917)	(0.013)
Religion	0.0138***		0.0179***	
	(0.000)		(0.000)	
IMR		0.9276***		0.9612***
		(0.000)		(0.000)
Observations	38009	31676	38009	31676
R^2 adjusted		0.873		0.869

* p-value < 0.10, ** p-Wert < 0.05, *** p-Wert < 0.01

The product-level extensive margin is the dependent variable in the behavioral equation (2nd stage in selection model). The country-level extensive margin is the dependent variable in the selection equation (1st stage in selection model). Robust p-values in brackets. Exporter, importer and time fixed effects. The selection equation is a Probit model, marginal effects at sample means are reported, for dummy variables the discrete change from 0 to 1 is reported. Religion is excluded variable in second stage specifications.

Table 5: Gravity Model applied to the Extensive Margin

	Selection equation	Behavioral equation	Selection equation	Behavioral equation
ln(GDP importer)	0.0445***	0.1673***		
	(0.000)	(0.000)		
ln(GDP exporter)	0.0572***	0.2524***		
	(0.000)	(0.000)		
ln(GDP p.c. importer)			0.0147**	0.0449***
			(0.016)	(0.001)
ln(GDP p.c. exporter)			0.0194***	0.0720***
			(0.002)	(0.000)
ln(Population importer)			0.0028	-0.2736***
			(0.878)	(0.000)
ln(Population exporter)			0.0949***	0.5749***
			(0.000)	(0.000)
Distance	-0.1140***	-0.4214***	-0.1147***	-0.4223***
	(0.000)	(0.000)	(0.000)	(0.000)
WTO (both)	0.0210***	0.0414***	0.0200***	0.0414***
	(0.000)	(0.000)	(0.000)	(0.000)
WTO (none)	-0.0119*	-0.0008	-0.0108*	-0.0006
	(0.069)	(0.703)	(0.098)	(0.764)
FTA	0.0353***	0.0510***	0.0353***	0.0511***
	(0.000)	(0.000)	(0.000)	(0.000)
Currency union	0.0938***	0.0363***	0.0936***	0.0363***
	(0.000)	(0.000)	(0.000)	(0.000)
Land border	-0.0084	0.0582***	-0.0097	0.0578***
	(0.480)	(0.000)	(0.414)	(0.000)
Legal	0.0143***	0.0592***	0.0142***	0.0591***
	(0.000)	(0.000)	(0.000)	(0.000)
Language	0.0489***	0.0961***	0.0487***	0.0956***
	(0.000)	(0.000)	(0.000)	(0.000)
Colonial ties	0.0504***	0.0575***	0.0492**	0.0575***
	(0.009)	(0.000)	(0.011)	(0.000)
Island	0.0246***	0.0248***	0.0244***	0.0248***
	(0.000)	(0.000)	(0.000)	(0.000)
Landlock	0.0317***	0.0157***	0.0317***	0.0157***
	(0.000)	(0.000)	(0.000)	(0.000)
Religion	0.0638***		0.0631***	
	(0.000)		(0.000)	
IMR		0.4016***		0.4014***
		(0.000)		(0.000)
Observations	166507	115951	166507	115951
R^2 adjusted		0.874		0.874

* p-value < 0.10, ** p-Wert < 0.05, *** p-Wert < 0.01

The product-level extensive margin is the dependent variable in the behavioral equation (2nd stage in selection model). The country-level extensive margin is the dependent variable in the selection equation (1st stage in selection model). Robust p-values in brackets. Exporter, importer and time fixed effects. The selection equation is a Probit model, marginal effects at sample means are reported, for dummy variables the discrete change from 0 to 1 is reported. The behavioral equation is an OLS regression, beta coefficients are reported. Religion is excluded variable in the second stage specification.

Table 6: Population Size Ratio

	Selection equation	Behavioral equation	Behavioral equation
Relative p.c. GDP	0.0439***	0.0507***	0.1615***
	(0.000)	(0.000)	(0.000)
Relative p.c. GDP squared			-0.1150***
			(0.000)
ln(Relative population size)	0.0021**	0.0011***	0.0012***
	(0.012)	(0.001)	(0.000)
ln(GDP p.c. importer)	0.0115*	0.0063**	0.0038
	(0.056)	(0.039)	(0.216)
ln(GDP p.c. exporter)	0.0166***	0.0111***	0.0085**
	(0.008)	(0.002)	(0.017)
ln(Population importer)	0.0050	-0.0365***	-0.0357***
	(0.782)	(0.000)	(0.000)
ln(Population exporter)	0.0939***	0.0825***	0.0837***
	(0.000)	(0.000)	(0.000)
Distance	-0.1105***	-0.1244***	-0.1235***
	(0.000)	(0.000)	(0.000)
WTO (both)	0.0198***	0.0223***	0.0220***
	(0.000)	(0.000)	(0.000)
WTO (none)	-0.0103	-0.0028	-0.0042
	(0.107)	(0.379)	(0.187)
FTA	0.0361***	0.0564***	0.0555***
	(0.000)	(0.000)	(0.000)
Currency union	0.0903***	0.0591***	0.0566***
	(0.000)	(0.000)	(0.000)
Land border	-0.0124	0.0852***	0.0856***
	(0.294)	(0.000)	(0.000)
Legal	0.0144***	0.0297***	0.0297***
	(0.000)	(0.000)	(0.000)
Language	0.0461***	0.0507***	0.0508***
	(0.000)	(0.000)	(0.000)
Colonial ties	0.0501***	0.1290***	0.1289***
	(0.007)	(0.000)	(0.000)
Island	0.0181***	0.0273***	0.0279***
	(0.000)	(0.000)	(0.000)
Landlock	0.0208***	0.0210***	0.0212***
	(0.000)	(0.000)	(0.000)
Religion	0.0628***		
	(0.000)		
IMR		0.9829***	0.9702***
		(0.000)	(0.000)
Observations	166507	115951	115951
R^2 adjusted		0.875	0.876

* p-value < 0.10, ** p-Wert < 0.05, *** p-Wert < 0.01

The product-level extensive margin is the dependent variable in the behavioral equation (2nd stage in selection model). The country-level extensive margin is the dependent variable in the selection equation (1st stage in selection model). Robust p-values in brackets. Exporter, importer and time fixed effects. The selection equation is a Probit model, marginal effects at sample means are reported, for dummy variables the discrete change from 0 to 1 is reported. Religion is excluded variable in the second stage specification.

Table 7: Baseline Results – only for North-North Country Pairs (NN)

	Benchmark excl. zeros	Benchmark incl. zeros	Selection equation	Behavioral equation	Behavioral equation	Selection equation	Behavioral equation	Selection equation	Behavioral equation
Relative p.c. GDP	0.0082*	0.0157***	0.0015	0.0289***	-0.1900***				
	(0.051)	(0.000)	(0.260)	(0.000)	(0.000)				
Relative p.c. GDP squared					0.1932***				
					(0.000)				
ln(Relative p.c. GDP)						0.0004	0.0226***		
						(0.579)	(0.001)		
Traditional Linder term								-0.0016	-0.0266***
								(0.360)	(0.000)
ln(GDP p.c. importer)	0.0257**	0.0005	0.0001	-0.0063	0.0336**	0.0000	-0.0122	-0.0000	-0.0094
	(0.016)	(0.963)	(0.935)	(0.646)	(0.024)	(0.988)	(0.398)	(0.995)	(0.501)
ln(GDP p.c. exporter)	0.0058	0.0103	-0.0012	-0.0049	0.0340*	-0.0012	-0.0107	-0.0012	-0.0080
	(0.709)	(0.508)	(0.307)	(0.785)	(0.069)	(0.337)	(0.560)	(0.292)	(0.655)
ln(Population importer)	0.1194	0.2794	-0.0046	-0.6489***	-0.5302***	-0.0048	-0.6850***	-0.0048	-0.6663***
	(0.521)	(0.154)	(0.403)	(0.001)	(0.008)	(0.383)	(0.001)	(0.390)	(0.001)
ln(Population exporter)	1.4896***	1.5988***	0.0110*	1.4941***	1.6220***	0.0109*	1.4712***	0.0109*	1.4808***
	(0.000)	(0.000)	(0.075)	(0.000)	(0.000)	(0.078)	(0.000)	(0.078)	(0.000)
Distance	-0.2951***	-0.2525***	-0.0058***	-0.4193***	-0.4175***	-0.0058***	-0.4196***	-0.0058***	-0.4195***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
WTO (both)	0.0065	0.0093	0.0000	0.0091	0.0094	0.0000	0.0088	0.0000	0.0090
	(0.530)	(0.342)	(0.996)	(0.507)	(0.491)	(0.989)	(0.522)	(0.997)	(0.513)
WTO (none)	0.0298***	0.0431***	0.0020***	0.0222***	0.0238***	0.0020***	0.0226***	0.0020***	0.0224***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
FTA	0.0364***	0.0536***	-0.0018	0.0479***	0.0490***	-0.0018	0.0478***	-0.0018	0.0478***
	(0.000)	(0.000)	(0.358)	(0.000)	(0.000)	(0.360)	(0.000)	(0.359)	(0.000)
Currency Union	-0.0168***	-0.0104***	-0.0052	-0.0125**	-0.0136***	-0.0041	-0.0122**	-0.0047	-0.0123**
	(0.000)	(0.001)	(0.602)	(0.014)	(0.008)	(0.637)	(0.017)	(0.616)	(0.016)
Land border	-0.0174***	-0.0078**	-0.0075	0.0589***	0.0570***	-0.0079	0.0596***	-0.0076	0.0593***
	(0.000)	(0.036)	(0.276)	(0.000)	(0.000)	(0.270)	(0.000)	(0.275)	(0.000)
Legal	0.0553***	0.0458***	0.0018***	0.0579***	0.0591***	0.0017***	0.0581***	0.0018***	0.0580***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Language	0.0421***	0.0368***	-0.0024**	0.0421***	0.0467***	-0.0024**	0.0413***	-0.0024**	0.0415***
	(0.000)	(0.000)	(0.045)	(0.000)	(0.000)	(0.048)	(0.000)	(0.045)	(0.000)
Island	0.0545***	0.0654***	-0.0051***	0.0482***	0.0468***	-0.0052***	0.0479***	-0.0051***	0.0481***
	(0.000)	(0.000)	(0.010)	(0.000)	(0.000)	(0.009)	(0.000)	(0.010)	(0.000)
Religion			0.0040***			0.0040***		0.0040***	
			(0.002)			(0.002)		(0.002)	
IMR				0.2008***	0.2002***		0.2007***		0.2008***
				(0.000)	(0.000)		(0.000)		(0.000)
Observations	16177	17670	11044	9551	9551	11044	9551	11044	9551
R^2 adjusted	0.961	0.952		0.938	0.938		0.938		0.938

* p-value < 0.10, ** p-Wert < 0.05, *** p-Wert < 0.01
The product-level extensive margin is the dependent variable in the benchmark specification (standard OLS) and the behavioral equation (2nd stage in selection model). The country-level extensive margin is the dependent variable in the selection equation (1st stage in selection model). Robust p-values in brackets. Exporter, importer and time fixed effects. The selection equation is a Probit model, marginal effects at sample means are reported, for dummy variables the discrete change from 0 to 1 is reported. The behavioral equation is an OLS regression, beta coefficients are reported. Religion is excluded variable in second stage specifications.

Table 8: Baseline Results – only for Country Pairs which do *not* include two High Income Countries (SS/SN/NS)

	Benchmark excl. zeros	Benchmark incl. zeros	Selection equation	Behavioral equation	Behavioral equation	Selection equation	Behavioral equation	Selection equation	Behavioral equation
Relative p.c. GDP	0.0587***	0.0686***	0.0820***	0.0587***	0.1841***				
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)				
Relative p.c. GDP squared					-0.1179***				
					(0.000)				
ln(Relative p.c. GDP)						0.0294***	0.1009***		
						(0.000)	(0.000)		
Traditional Linder term								-0.0895***	-0.0742***
								(0.000)	(0.000)
ln(GDP p.c. importer)	0.0631***	0.0699***	0.0156**	0.0454***	0.0332**	0.0075	0.0000	0.0145*	0.0384**
	(0.001)	(0.000)	(0.049)	(0.005)	(0.044)	(0.337)	(0.998)	(0.067)	(0.019)
ln(GDP p.c. exporter)	0.0492**	0.0628***	0.0240***	0.0733***	0.0604***	0.0186**	0.0262	0.0232***	0.0657***
	(0.015)	(0.000)	(0.003)	(0.000)	(0.002)	(0.018)	(0.185)	(0.004)	(0.001)
ln(Population importer)	-0.1401**	-0.2130***	0.0086	-0.2451***	-0.2402***	0.0157	-0.1785***	0.0127	-0.2203***
	(0.042)	(0.001)	(0.714)	(0.000)	(0.000)	(0.497)	(0.008)	(0.587)	(0.001)
ln(Population exporter)	0.5039***	0.5127***	0.1097***	0.5630***	0.5723***	0.1182***	0.6363***	0.1136***	0.5883***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Distance	-0.3109***	-0.2259***	-0.1374***	-0.4532***	-0.4500***	-0.1360***	-0.4471***	-0.1366***	-0.4510***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
WTO (both)	0.0320***	0.0248***	0.0253***	0.0444***	0.0441***	0.0249***	0.0445***	0.0253***	0.0445***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
WTO (none)	-0.0057**	-0.0032	-0.0176**	-0.0040*	-0.0050**	-0.0190**	-0.0060**	-0.0178**	-0.0044*
	(0.018)	(0.105)	(0.034)	(0.093)	(0.035)	(0.021)	(0.011)	(0.031)	(0.063)
FTA	0.0804***	0.1027***	0.0602***	0.0650***	0.0634***	0.0592***	0.0635***	0.0597***	0.0643***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Currency Union	0.0136***	-0.0098***	0.1203***	0.0552***	0.0528***	0.1187***	0.0501***	0.1196***	0.0533***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Land border	0.0588***	0.0641***	-0.0047	0.0673***	0.0677***	-0.0066	0.0671***	-0.0051	0.0670***
	(0.000)	(0.000)	(0.740)	(0.000)	(0.000)	(0.642)	(0.000)	(0.720)	(0.000)
Legal	0.0604***	0.0436***	0.0165***	0.0592***	0.0597***	0.0171***	0.0610***	0.0166***	0.0596***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Language	0.0480***	0.0385***	0.0638***	0.1087***	0.1093***	0.0635***	0.1099***	0.0636***	0.1090***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Colonial ties	0.0679***	0.0685***	0.0632**	0.0618***	0.0616***	0.0624**	0.0618***	0.0634**	0.0617***
	(0.000)	(0.000)	(0.015)	(0.000)	(0.000)	(0.016)	(0.000)	(0.014)	(0.000)
Island	0.0186***	0.0128***	0.0372***	0.0300***	0.0306***	0.0345***	0.0295***	0.0366***	0.0302***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Landlock	-0.0012	0.0020	0.0182***	0.0095***	0.0092***	0.0141**	0.0067***	0.0160**	0.0084***
	(0.452)	(0.114)	(0.004)	(0.000)	(0.000)	(0.028)	(0.000)	(0.012)	(0.000)
Religion			0.0767***			0.0773***		0.0767***	
			(0.000)			(0.000)		(0.000)	
IMR				0.4528***	0.4466***		0.4432***		0.4500***
				(0.000)	(0.000)		(0.000)		(0.000)
Observations	110418	159481	152121	103058	103058	152121	103058	152121	103058
R^2 adjusted	0.817	0.770		0.842	0.843		0.844		0.843

* p-value < 0.10, ** p-Wert < 0.05, *** p-Wert < 0.01
The product-level extensive margin is the dependent variable in the benchmark specification (standard OLS) and the behavioral equation (2nd stage in selection model). The country-level extensive margin is the dependent variable in the selection equation (1st stage in selection model). Robust p-values in brackets. Exporter, importer and time fixed effects. The selection equation is a Probit model, marginal effects at sample means are reported, for dummy variables the discrete change from 0 to 1 is reported. The behavioral equation is an OLS regression, beta coefficients are reported. Religion is excluded variable in second stage specifications.