LIS
Working Paper Series

No. 658

Revisiting the Sectoral Linder Hypothesis: Aggregation Bias or Fixed Costs?

Hendrik W. Kruse

January 2016

CROSS-NATIONAL DATA CENTER in Luxembourg

Luxembourg Income Study (LIS), asbl

Revisiting the Sectoral Linder Hypothesis: Aggregation Bias or Fixed Costs?*

Hendrik W. Kruse†
University of Göttingen

January 21, 2016

Abstract

This paper reassesses and revisits the Sectoral Linder Hypothesis due to Hallak (2010), according to which similar tastes for quality lead to more intensive trade between similar countries. First, it will be shown that allowing for strictly non-homothetic preferences reduces confoundedness and improves results. Moreovoer, the country/firm level extensive margin is taken into account. This approach allows controlling for unobserved firm level heterogeneity and selection bias (Helpman et al. 2008). The advantage in terms of interpretation is that differences in coefficients at the two margins can be linked to fixed cost effects. The attempt is to show that the Linder effect is confounded with fixed (opportunity) costs of trade thereby leading to downward biased results. There is some evidence that this effect is exacerbated at the aggregate, intersectoral level. Fixed (opportunity) costs seem to be higher in sectors where similar countries trade a lot. The evidence reinforces the sectoral Linder hypothesis, and suggests that the patterns might prevail at the more aggregate levels. Other robustness checks suggest that results are not confined to products that are vertically differentiated.

Keywords: International trade, income similarity, Linder hypothesis, strictly non-homothetic preferences, extensive and intensive margin of trade.

JEL classification: F14, D31

1 Introduction

The Linder hypothesis has triggered a bulk of empirical literature in recent decades.[1] In his most radical deviation from supply-side explanations of international trade Staffan Burenstam Linder (1961) claimed that similarity of demand rather than differences in factor endowment determine the extent to which countries could potentially benefit from bilateral trade. While early empirical test were mixed, recent evidence is increasingly vindicating Linder's hypothesis (cf. e.g. Choi 2002).

Many authors argue that demand shifts towards luxury goods as income increases.[2] Others have emphasized the role of quality differentiation.[3] In fact, Hallak (2010) has claimed that the Linder's theory holds good at the sector level only. Rich countries demand higher quality goods, which gives

*This paper is based on my Master Thesis. I thank my supervisor Inmaculada Martínez-Zarzoso for her excellent support, and advice. Further, I'd like to thank Lutz Depenbusch, Marco Kruse, Janis Wohlfahrt, Tobias Lietz, Ana Abeliansky, Marcello Perez, Teodora Borota, Tristan Kohl, Malte Ehrich and Lennart Kaplan for fruitful discussion, and comments. I would also like to thank Juan Carlos Hallak for providing his do-files and data, and the LIS team for their friendly and helpful user support.

†Email: hkruse@gwdg.de

[1]Deardorff (1984, pp. 504–506) provides a review of early empirical tests.

[2]Cf. eg. Francois and Kaplan 1996, Dalgin et al. 2008, Fieler 2011.

[3]Cf. eg. Bils and Klenow 2001, Choi et al. 2009, Flam and Helpman 1987, Fajgelbaum et al. 2011.

them an advantage in their very production. Of the money they devote to imports most will be spent for imports of a suitable quality, to the effect, that similar countries trade more. This gives rise to the sectoral Linder hypothesis, i.e. the hypothesis that Linder's theory holds at the sector level. One main finding in Hallak (2010) is that the hypothesis fails at the aggregate level due to aggregation bias. A potential reason could be due to intersectoral specialization according to factor proportions in the classical Heckscher-Ohlin sense (cf. Eppinger and Felbermayr 2015).[4]

In this paper, I will revisit and reassess Hallak's (2010) hypothesis. A cross-section of countries will be analysed for 2004. As in Hallak (2010) the sector specific regression analysis will be applied albeit with some methodological changes. The Helpman et al. (2008) methodology will be applied to additionally control for the bias due to the omission of unobserved firm heterogeneity and zero trade flows. I deviate from Hallak in allowing an influence of similiarity at the extensive margin, i.e. in the selection equation. Heteroskedasticity is controlled for using Feasible Generalized Least Squares (FGLS) (Martínez-Zarzoso 2013). The performance of the traditional Linder term—the similarity in per capita GDP—will be compared to an Overlap Index as introduced by Bernasconi (2013) that allows for strict non-homotheticity. This index, however, will be calculated using micro data as in Choi et al. (2009). It will show that the effect of similarity is much stronger at the intensive margin. Also, allowing for strict non-homotheticity greatly improves results at both margins. Furthermore, this paper provides evidence as to when and potentially why aggregation effects might emerge. Controlling for the country level extensive margin allows to control for fixed (opportunity) costs that might play a role, because similarity probably reduces costs of servicing another market by other means (i.e. FDI). Allowing for strict non-homotheticity makes this potential correlation less likely still, and additionaly weakens the confounding effect of Heckscher-Ohlin type specialization. It will show that both adjustments in fact make aggregation effects largely disappear.

This issue is of some importance as aggregation effects would be biases only in so far as the quality Linder hypothesis is concerned. It would still be correct to refute the "aggregate Linder hypothesis", as Hallak does. This would surely limit the relevance of Linder's reasoning, albeit not necessarily rendering it completely unimportant.

The Linder hypothesis is usually applied only to differentiated goods. That is because differentiated goods are most likely to show the relevant characteristics, e.g. quality differentiation. While this may seem as a limitation, the study of differentiated goods can yield important insights. If "what you export matters"—as Hausmann et al. (2007) assert in the title of their paper—determinants of trade in differentiated products could teach us something about the difficulty of developing countries to gain market power in innovative markets. Be that as it may, in the sample employed here it seems that the hypothesis in fact applies for other types of goods, too. In accordance with other robustness checks this seems to suggest that quality is not the only important dimension of specialization that gives rise to Linder effects.

The remainder of this paper is organized as follows. Section 2 provides the motivating theoretical remarks, and introduces the concept of aggregation bias. In section 3 the empirical implementation is discussed. First, it will be described how the Overlap Index was constructed. Then, briefly, the Helpman et al. (2008) method will be outlined. Finally, the treatment of heteroskedasticity will be discussed. In section 4 the data sources are described, and some descriptive statistics are presented. Section 5 has the results, including evidence concerning aggregation bias and robustness checks. Section 6 summarizes and concludes.

2 Motivation

In his classic essay Linder (1961) hypothesized that trade is more intensive between similar countries. He dismissed factor proportions as an explanation for trade flows mainly on the grounds that their exploitation requires international entrepreneurial mobility (cf. ibid, pp. 92-92). The starting point

[4] Other papers focus merely on non-homothetic preferences, without studying the production side. These papers include Hunter (1991), and Reimer and Hertel (2010).

of his argument is that for an entrepreneur to start production of a good, or an inventor to come up with a new idea requires knowledge of the state of need. Arguably, an entrepreneur is most familiar with society's needs in his home country. "[E]xport is the end, not the beginning of the typical market expansion path" (ibid, p.88). In turn, domestic demand for a manufacture determines the comparative advantage of the country, or in Linder's words "*the production functions of goods demanded at home are the relatively most advantageous ones*" (ibid, p.90, italics in the original).[5] That means, of course, that the largest market for a country's goods is to be found in a country that demands relatively high quantities of these kinds of good, i.e. a country with a similar demand structure.

This marks a departure from the convenient assumption of homothetic preferences to be found in many theories of international trade. With homothetic preferences the utility maximizing expenditure shares of different goods are constant irrespective of the household's income. Since the pattern of demand does not change across countries under such circumstances there is no space for demand similarity.

One reason for non-homothetic preferences might be the presence of luxury goods (cf. Francois and Kaplan 1996, Mitra and Trindade 2005, Dalgin et al. 2008). In this paper, however, the focus is upon quality differentiation as the source of non-homotheticity in accordance with Hallak (2010). The fundamental idea is straightforward. With growing income households spend an ever increasing share of it on the quality of the varieties they purchase, while the increment in quantity declines. This focus on quality driven demand similarities is what constitutes the sectoral Linder hypothesis. This does not imply, that at the aggregate level the respective influence will be found. Hallak invokes Simpson's paradox, due to Simpson (1951, p. 241), to explain why an aggregation bias in that sense might occur. In the present context income similarities are likely to be correlated with patterns of trade between sectors, to the effect that on the aggregate level similarities might actually hamper trade. Eppinger and Felbermayr (2015) put it in different words: including the Linder term at the aggregate level might lead to confusion of demand side and Heckscher-Ohlin effects, or comparative advantage due to factor endowment. This interpretation implies, however, that while Linder can explain certain sector specific phenomena, by and large specialization is of a Heckscher-Ohlin type, and thus strongly limits the relevance of demand similarity.

There might, however, be a different reason why the Linder effect does often not show up at the aggregate level. In their seminal paper Helpman et al. (2008, p. 470) show, that the omission of firm-level heterogeneity in gravity equations, affecting the so called extensive margin of trade, i.e. the number of firms exporting to the partner country, introduces a serious bias that dwarfs the bias due to zero trade flows. Unobserved heterogeneity, so they argue, gives rise to asymmetric trade flows that traditional methods cannot account for.[6] In particular, if the distribution of productivity varies across countries, a different share of the firms will find it profitable to export to the other country. That is, even if the zero profit productivity level was symmetric, the trade flows would not usually be (cf. Helpman et al. 2008, p. 451). Given that trade flows are in fact often times asymmetric or even unidirectional (cf. Helpman et al. 2008, p. 447) this will impose a bias upon any coefficient estimate, the corresponding variable of which is related to the extensive margin of trade. The true estimate would then actually be different depending on the direction of the trade flow, which leads to an estimate that captures the effect in neither direction correctly.

[5] Schott (2004) provides an alternative explanation for the supply side effect applicable at the disaggregate sector level. He claims that higher quality goods are more capital and skill intensive, which generally coincides with richer countries' factor abundance, to the effect that countries have an advantage in the production of those qualities that are demanded at home, but not because of the demand. This line of argument is superior in two respects. First, it does not rely on economies of scale, which is an advantage because evidence in that realm is scant (cf. Feenstra 2004, pp. 141-144). Secondly, it provides an intuitive theoretical reason, why the effect of similarity might be different across sectors. Simply, the extent to which skill and capital intensity increases with quality might differ from sector to sector.

[6] Distinguishing exporter and importer specific fixed effects rather than country fixed effects may in fact already lead to the prediction of asymmetries. However, these are monadic variables, while in Helpman et al. a dyadic variable is responsible for asymmetry, which provides an arguably more differentiated picture. However, to be fair, note that it is because in the first stage exporter and importer fixed effects are distinguished, i.e. monadic variables, that leads to an asymmetric prediction as to the probability of trade, and thus to a dyadic asymmetric variable in the second stage. Also the inverse Mills' ratio is only asymmetric because of the distinction of exporter and importer fixed effects.

What is more, coefficients might have different effects at the two margins, and demand similarity might be a case in point. It might well be the case that demand similarity reduces fixed costs (in particular training costs and search costs) to service another country's market by means of FDI or licensing rather than export.[7] This would mean that the less productive firms engage in export and the more productive firms engage in FDI, and thus imply a negative relation of similarity and exporter productivity, which would yield downward biased results. And even if there is no productivity effect, fixed costs will have an effect on the probability of trade. The differentiation of extensive and intensive margin, and subsequent control for selection might thus be sufficient to seperate the fixed cost effect of a variable from its direct effect on trade flows.

The fixed cost story is consistent with Helpman et al. (2004), where demand is canceled out in the proximity-concentration trade-off. An alternative explanation is offered by Fajgelbaum et al. (2015). In their model there is a Linder effect for FDI that under certain circumstances (if trade in intermediate inputs does not kick in) could lead to a decline in Linder-type trade.

The relevance of demand structures requires non-homothetic preferences. Non-homotheticity implies that the income distribution matters. In its weakest form, only the first ordinary moment—per-capita income—is taken to exhibit an influence. This would be the case if up to a certain level of income only a specific set of goods is consumed. Above that level the marginal increment in the expenditure for a good would have to be constant, while the expenditure share varies with income. Strict non-homotheticity allows for the marginal increment, i.e. the share of an additional dollar spent on the commodity at hand, to vary with income. In order to account for that higher order moments of the income distribution have to be considered. A measure of demand similarity allowing for strict non-homotheticity is less likely to be confounded with fixed costs or Heckscher-Ohlin effects, since at least part of within country income differences are due to capital income (as discussed at length in Piketty's (2014) epoche making book). Insofar as similarity is due to such factors it is less likely to be correlated with similarities in broad *country-level* characteristics like Capital-Labour Ratios (Heckscher-Ohlin) or FDI fixed costs.

Note that based on this argument the present approach might also indicate whether a Linder effect for FDI or fixed costs are important at the extensive margin. A Linder effect for FDI should affect results using both measures, while the confusion with fixed costs effects will be worse for the traditional term.

Then, four testable hypotheses emerge:

Hypothesis 1 *The effect of demand similarity is stronger at the intensive margin.*

Hypothesis 2 *The effect of demand similarity is stronger at both margins if strictly non-homothetic preferences are allowed.*

Hypothesis 3 *Aggregation Bias is less of an issue if the unobserved heterogeneity is controlled for.*

Hypothesis 4 *Aggregation Bias is less of an issue if strict non-homotheticity is allowed for.*

Note that Linder's (1961) hypothesis does not apply to all goods in the same way. He concedes that trade in primary products is not likely to fit his ideas. Instead, most scholars focus on differentiated products. Hallak for instance introduces the Linder effect in a Dixit-Stiglitz "love of variety" model. In order for Hallak's own hypothesis to hold those have to be differentiable with respect to quality, too.

While this might seem as a limitation, the study of differentiated goods is interesting in its own right. According to Hausmann et al. (2007) it matters for growth prospect what kind of goods are being exported. On these grounds it might seem important to understand the determinants of the potentially attractive market of differentiated goods. However, as we shall see, in fact a Linder effect can be found—in the given sample—for other types of goods, too. This, of course, casts some doubt upon quality as a transmission channel.

[7] Evidence for a turn towards investment in the context of the proximity-concentration trade-off has been provided by Brainard (1997). Markusen and Venables (2000) in fact argue that FDI is more likely between similar countries because they often have similar cost structures, as well.

3 Empirical Implementation

Demand similarity

Two measures of demand similarity will be considered. First, in accordance with Hallak (2010), one could concentrate on the similarity or difference of per capita incomes. The traditional Linder term is thus defined as:

$$Linder_{od} = \frac{1}{2} \left(\ln y_o - \ln y_d\right)^2 \tag{1}$$

i.e. the squared difference in log-transformed per capita GDP, where y_o and y_d are the origin and destination country's per capita GDPs respectively. Hallak shows, that equivalently the product of per capita GDPs can be used as a similarity measure:

$$\ln y_o \ln y_d = -\frac{1}{2} \left(\ln y_o - \ln y_d\right)^2 + \frac{1}{2}(\ln y_o)^2 + \frac{1}{2}(\ln y_d)^2 \tag{2}$$

In a cross country regression the latter two terms will be captured by country fixed effects, thereby rendering the measures equivalent up to the sign. From the above it should be obvious that this way of approaching the problem does not allow for strictly non-homothetic preferences.

Choi et al. (2009) propose an index that incorporates all information obtainable from the income distribution. In particular, they calculate the estimated cumulative distance between the income distributions of two countries:

$$Diss_{od} = \frac{1}{2} \int |r_o(y) - r_d(y)| dy \tag{3}$$

where $r(y)$ denotes the estimated density of income for the respective country. One could argue that an additional advantage of the index is that it is less likely to be correlated with differences in factor endowment. Insofar as differences in factor prices are important and show in income differences, the measure is still contaminated but not to the same extent as lower order moments of the distribution. Moreover, as Markusen (2013, p. 260) points out, even with a linear income expansion path if there are households that do not buy the good with high income elasticity (or that buy the lowest quality available and the highest quantity possible) the distribution of income matters. Under certain conditions this can severely affect the estimated impact of income similarity.

In order to see how to incorporate such a measure into a gravity model consider a typical gravity equation á la Anderson and van Wincoop (2003):

$$imp_{od} = \frac{Y_o Y_d}{Y_w} \left(\frac{\tau_{od}}{P_d \Pi_o}\right)^{1-\sigma} \tag{4}$$

P_d and Π_o—the notorious multilateral trade resistance terms—can be ignored as they too will be subsumed under importer and exporter fixed effects. Y_o, Y_d and Y_w are origin, destination and world GDP respectively—the former two serving as proxies for market size and productive capacity of the trading partners—τ_{od} are trade costs, and σ is the elasticity of substitution. One interpretation of Linder's argument is that market size, and production capacity vary for each country, depending on the trading partner. The importing country's GDP matters only insofar for effective market size as there is an overlap of demand—i.e of the income distribution—and the exporting country's GDP matters only insofar for effective production capacity as there is an overlap of demand. Thus, in (4) Y_o and Y_d can be replaced by $S_{od}Y_o$ and $S_{od}Y_d$,[8] where:

$$S_{od} = 1 - Diss_{od} = \int \min\{r_o(y), r_d(y)\} dy \tag{5}$$

[8] Note that while technically a similar adjustment has to be made for y_w this is of no practical relevance, as this term will in any case be captured by importer and exporter fixed effects.

which corresponds to one of the measures used in Bernasconi (2013). I will follow her in calling it the Overlap Index. In accordance with Hallak (2010) the following gravity type equation emerges:

$$imp_{od} = Dist_{od}{}^{\beta_D} e^{\boldsymbol{\beta}_{\mathbf{I}}\mathbf{I}_{od}} S_{od}^{2\beta_{DIS}} e^{\varphi_o + \psi_d + u_{od}} \tag{6}$$

where imp_{od} denotes imports of country d from country o. Trade costs are specified as $\tau_{od} = Dist_{od}{}^{\beta_D} e^{\boldsymbol{\beta}_{\mathbf{I}}\mathbf{I}_{od}}$. $Dist_{od}$ measures bilateral distance, and $\mathbf{I}_{od}$ is a vector of binary trade cost determinants to be specified below. φ_o and ψ_d are exporter and importer fixed effects absorbing country size (GDP), multilateral resistance to trade and other unobserved country level characteristics.

Taking logs of equation (6) yields a linear model as employed in standard applications of the gravity equation:

$$\ln imp_{od} = \varphi_o + \psi_d + \beta_D \ln Dist_{od} + \boldsymbol{\beta}_{\mathbf{I}}\mathbf{I}_{od} + 2\beta_{DIS} \ln S_{od} + u_{od} \tag{7}$$

Substituting $\ln S_{od}$ for $\ln y_o \ln y_d$, yields the specification used in Hallak (2010). It is worth noting, that this approach implies that in a correct specification S_{od}— albeit a share—has to be included in logs. In both cases the expectation would be $\beta_{DIS} > 0$ if demand similarity played positive role.

Unobserved Heterogeneity

In order to control for unobserved heterogeneity the approach by Helpman et al. (2008) will be used. Helpman et al. argue, that the productivity of exporting firms is a function of the estimated probability to observe positive trade. On that ground, they build a two stage model on the basis of Heckman's (1979) selection model. At the first stage a probit estimation is used, modelling the possibility of a positive trade flow. For identification purposes an exclusion restriction is specified, namely that the common religion variable exhibits an influence at the first stage, but not at the second stage. Note that in a selection model a valid exclusion restriction is not always necessary for identification. Depending on the distribution of the predicted values, the assumption of bivariate normal error terms is often enough. Heckman (1979) shows that under specific distributional assumption the expected value of the error term at the second stage is a function of the inverse Mills ratio obtained from the first stage $E(u_{od}|imp_{od} > 0) = \beta_{MR} \frac{\phi(x_{1od}\hat{\boldsymbol{\beta}}_1^*)}{\Phi(x_{1od}\hat{\boldsymbol{\beta}}_1^*)}$, where x_{1od} is the vector of explanatory variables, and $\hat{\boldsymbol{\beta}}_1^*$ is the coefficient vector from the Probit estimation. The inclusion of this term controls for zero trade flows. More importantly, Helpman et al. model the unobserved heterogeneity as a function of the predicted values from the probit regression. In particular they include the following term based on the estimation of the first stage regression:

$$\hat{\bar{z}}_{od}^* \equiv E(z_{od}^*|z_{od} > 0) = \hat{z}_{od}^* + \frac{\phi(x_{1od}\hat{\boldsymbol{\beta}}_1^*)}{\Phi(x_{1od}\hat{\boldsymbol{\beta}}_1^*)} \tag{8}$$

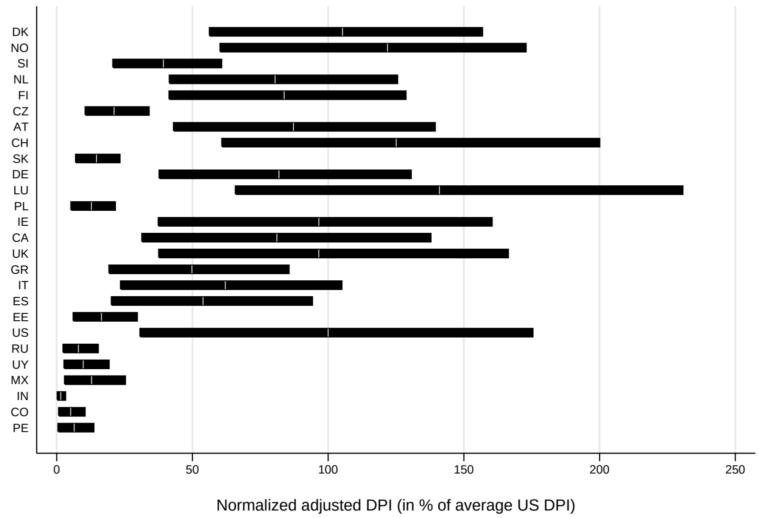
where z_{od} is the latent variable behind the Probit estimation, and $z_{od}^* = \frac{z_{od}}{\sigma_z}$, σ_z being the standard error of z_{od}. Accordingly, $\hat{z}_{od}^*$ is derived from the Probit estimation with $\hat{z}_{od}^* = \Phi^{-1}(\hat{\rho}_{od})$ (cf. Helpman et al. 2008, p. 456), where $\hat{\rho}_{od}$ is the estimated probability of trade from the first stage regression. In accordance with Helpman et al.'s second specification a third order polynomial of $\hat{\bar{z}}_{od}^*$ is included alongside the Inverse Mill's Ratio in equation (7).

Heteroskedasticity

Heteroskedasticity is yet another issue at the second stage to account for. It is well known that heteroskedasticity of the error term will lead to inefficiencies. Santos Silva and Tenreyro (2006), however, show that in constant elasticity models—such as the gravity equation—the presence of heteroskedasticity leads to biased estimates of the respective elasticities.

As a solution Santos Silva and Tenreyro (2006) propose using the Poisson Pseudo-Maximum Likelihood estimator (PPML). The author's show that PPML is always consistent, irrespecitve of the

Figure 1: Income Dispersion Across Countries in 2004



Note: Adjusted Income per adult-equivalent was calculated, rather than simple household income

correct underlying distribution. Implicitly, PPML rests on the assumption of a Constant Variance-to-Mean Ration. I tested this against the more flexible—and potentially more efficient (Martínez-Zarzoso 2013)—estimation using Feasible Generalized Least Squares (FGLS) using Park-test as proposed by Manning and Mullahy (2001, pp. 471–472).[9] It appears that FGLS produces a variance structure that is more consistent with its assumptions than PPML. On this ground the latter was deemed inefficient and we proceed with the results of FGLS estimation.[10]

Multilateral resistance terms are controlled for at the extensive and intensive margin by the inclusion of importer and exporter fixed effects (Feenstra 2004, Head and Mayer 2014).

4 Data

Trade data are taken directly from UNCOMTRADE[11] at the SITC Rev.2 four-digit level. The trade flows were retrieved as reported by the importer if available. In case trade flows were reported by the exporter, but not by the importer, values were taken from the export data. As is well known, reported exports can differ markedly from imports reported by the partner country, and the difference varies greatly. Comparison of import and export values on the subsample for which both were available yielded a median difference of 0.8% (sample) and 4% (broad sample). However, given large differences between median and mean, this hardly appears representative. Absent a reasonable estimate, no correction was made to account for the difference between f.o.b. and c.i.f. values.[12]

As common in the field, I use Rauch's (1999) liberal classification to distinguish three broad categories of goods[13]: Homogenous goods are goods that are traded on organized exchanges. Differentiated

[9] Also see Head and Mayer (2014) for details.

[10] Results for Park type regressions, PPML and other estimation techniques are available upon request.

[11] http://comtrade.un.org/db/

[12] Note that UN COMTRADE does not report data for Taiwan. It was assumed that the COMTRADE partner designated "Other Asia, nes" largely coincides with Taiwan, in accordance with the UN International Trade Statistics knowledge base, see: http://unstats.un.org/unsd/tradekb/Knowledgebase/Taiwan-Province-of-China-Trade-data.

[13] http://www.macalester.edu/research/economics/PAGE/HAVEMAN/Trade.Resources/Verbal.Desc/SITC/sitc_r2.

and referenced priced goods are not, but for the latter at least there are reference prices to give traders some orientation[14]. The former, on the other hand, are differentiated to an extent that precludes formulation of such prices.

Per-capita GDP—used to calculate the traditional Linder term—was taken from the CEPII gravity data set.[15] The income distribution measure was constructed employing the micro household income data kindly provided by the *Luxembourg Income Study (LIS) Database,* http://www.lisdatacenter.org (multiple countries; October 2015),.[16] Wave VI was used, however, only for those countries that reported about 2004.[17] The construction of the index was conducted exactly as explained in Choi et al. (2009). Adjusted disposable income was calculated as disposable household income per adult equivalent using a square root equivalence scale (cf. Choi et al. 2009, p. 274). It was converted to US-Dollar using current exchange rates taken from the Penn World Table data base version 6.1.[18] Data availability constrained the sample size for the Overlap Index. Only 27 countries reported micro data in 2004 [19]. Countries part of this narrow sample are listed and distinguished in table A.1 in the appendix.

In order to make sure sample selection does not drive results, a second broader sample was used for the traditional Linder term. The broader sample of countries for 1995 and 2004 was constructed using the same criteria as in Hallak (2010). Hallak drops countries with a population smaller than Singapore's in 1995.[20] On top of that, countries with a total import volume of less than $2 billion in differentiated goods were dropped as in Hallak (2010). Even though the criteria are based on 1995 values the same set of countries was used for the broad sample analysis in 2004. Note that Hong Kong had to be dropped because it was not contained in the Correlates of War database. That leaves us with 62 countries, listed in the appendix in table A.1. While Hallak (2010, cf. p. 460) drops small sectors, I will not follow him in this respect for the 2004 data, but conduct robustness checks in that direction. Table A.2 in the appendix lists all sectors according to their classification. Small sectors are labelled accordingly.

Trade cost determinants are taken from the CEPII data set. They include geographical great-circle distance between most populated cities or regions, and the common dummy variables for contiguity, common official language, maintenance of a regional trade agreement, and colonial past ($\mathbf{I}_{od}$).[21] Note, that the indicator for a colonial past has been modified in accordance with Hallak (2010, p. 460), so as to ignore colonial relationships that ended before 1922. Independence dates used in this context were also taken from the CEPII data set, as well as measures of aggregate GDP. The common religion variable was calculated based on the definition of Helpman et al. (2008, p. 480), but including more religions.[22] It is based on the data from the Correlates of War World Religion data set version 1.1 (Maoz and Henderson 2013). Since data are only collected every five years, I used 2005 data for the 2004 estimation, assuming that not too much change will take place in such a short period of time.

---

prn

[14]Table A.2 in the appendix lists all sectors for the categories at the 3-digit level. Tables A.5, A.6 and A.7 provide product descriptions and information on subsectors.

[15]http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=8

[16]http://www.lisdatacenter.org/data-access/

[17]The difference to Bernasconi (2013) here is that I use survey data, whereas she relies on decile information from the UNU-WIDER WIID dataset.

[18]https://pwt.sas.upenn.edu/php_site/pwt61_form.php. Later versions of PWT only report the exchange rate of the US-Dollar vis-à-vis the Euro for countries that were later to become Eurozone members. Hence, PWT 6.1 is the most up-to-date version that could be used. Heston et al. (2002).

[19]All of them but Iceland will be included in the sample.

[20]There have been some data revisions, apparently, leading to a slightly different sample than in Hallak (2010). This does not, however, significantly affect results. The results in Hallak (2010) can be obtained using either sample.

[21]For details see Mayer and Zignago 2011 and Head et al. 2010, p. 12f

[22]$religion_{od} = \sum_i (\textit{\% of Religion i in o} * \textit{\% of Religion i in d})$, where the following religions where considered: Protestant Christian, Catholic Christian, Sunni Islam, Shia Islam, Judaism, Buddhism, Hinduism, Bahaï Faith, Taoism, Confucianism and Animist religions.

Figure 2: Aggregate data with Linder term

(a) Narrow Sample (b) Narrow Sample excluding India

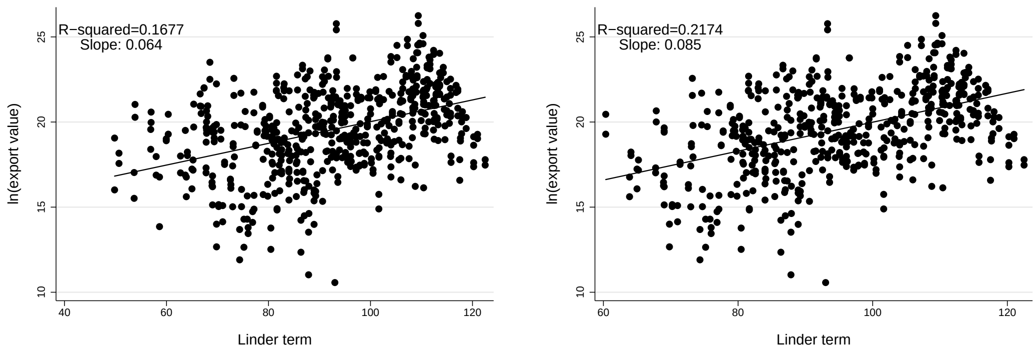


Figure 3: Aggregate data with Overlap Index

(a) Narrow Sample (b) Narrow Sample excluding India

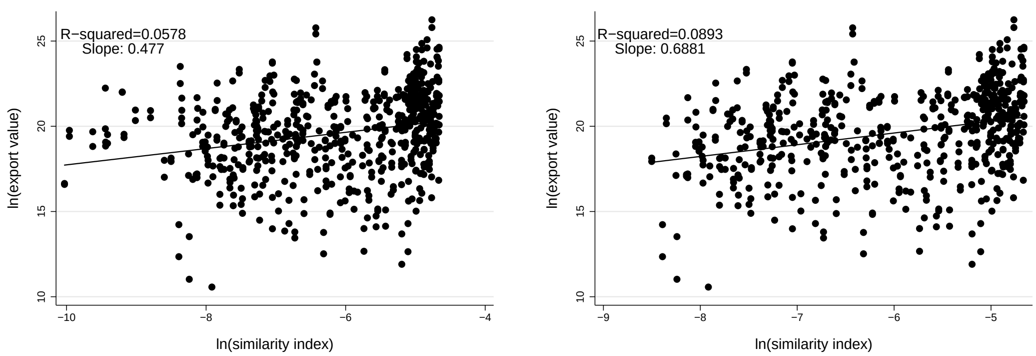


Figure 4: Correlation for Country Pairs Excluded in narrow sample

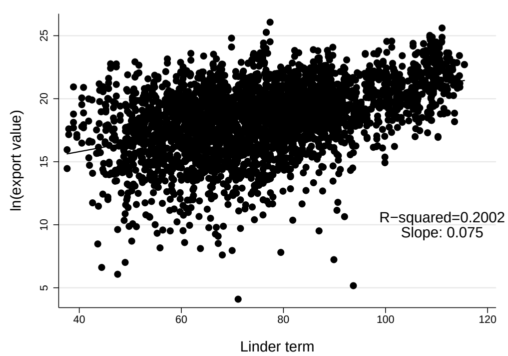


Table 1: Summary statistics

Countries	Gini	p90/p10	p10	p50	p90	mean	min	max	Variance	Skewness	Kurtosis	obs
AT	0.269	3.232	13837	25024	44724	27994	-90440	224491	249000000	3	24	5147
CA	0.318	4.379	10097	22590	44214	26031	-16910	494750	338000000	5	79	27820
CH	0.268	3.282	19553	36398	64168	40137	-146264	433198	519000000	4	50	3270
CO	0.506	12.614	264	982	3325	1651	0	208461	6537757	15	831	8994
CZ	0.266	3.208	3396	5894	10896	6783	78	118296	16900000	6	95	4351
DE	0.278	3.451	12137	22979	41879	26273	-34851	5337549	1750000000	106	13322	11294
DK	0.228	2.782	18084	32189	50314	33789	-2102895	1844869	355000000	3	1640	83349
EE	0.347	4.799	1979	4368	9498	5286	-13591	45288	14800000	3	17	4155
ES	0.316	4.652	6493	15217	30204	17302	-1926	260593	124000000	3	40	12996
FI	0.257	3.100	13311	24136	41259	26889	-16816	1680792	638000000	32	1661	11229
GR	0.327	4.420	6208	13521	27440	15983	-14925	146443	121000000	3	21	5568
IE	0.317	4.274	12036	26701	51446	30992	0	587091	755000000	11	189	6080
IN	0.491	10.880	94	287	1020	477	-937	61561	581065	23	1294	41546
IT	0.340	4.447	7576	16601	33686	19934	-29803	898932	369000000	13	399	8012
LU	0.269	3.498	21153	40227	73999	45246	-42963	599592	661000000	4	60	3622
MX	0.457	8.456	957	2758	8091	4122	0	1169591	43700000	54	7687	22595
NL	0.266	3.018	13346	23185	40273	25819	-109088	491701	312000000	8	169	9356
NO	0.256	2.869	19337	34579	55468	39110	-64173	11400000	7200000000	71	7184	13131
PE	0.519	24.549	178	1350	4363	2075	-10208	78293	9645940	9	143	18904
PL	0.315	4.038	1710	3523	6904	4099	-149729	136246	12100000	0	424	32214
RU	0.388	6.263	779	2017	4881	2566	-756	35365	4426662	3	20	3136
SI	0.231	2.921	6666	11795	19471	12610	0	66012	30600000	2	9	3725
SK	0.269	3.294	2262	4237	7450	4712	-11926	55707	8147473	5	66	5147
UK	0.344	4.413	12092	24964	53364	30983	-199063	1166608	1010000000	13	325	27753
US	0.364	5.699	9875	26523	56275	32093	-14273	963104	1040000000	9	160	76447
UY	0.424	6.905	893	2190	6165	3136	0	101916	12300000	8	129	18392

Note:
Gini: Gini Coefficient obtained directly from LIS Inequality and Poverty Key Figures: http://www.lisdatacenter.org/lis-ikf-webapp/app/search-ikf-figures
p90/p10: Decile Ratio for Adjusted Disposable Household Income, all other variables refer to the Adjsuted Disposable Household Income, calculated using the LISSY interface: http://www.lisdatacenter.org/data-access/

Table 2: Frequencies and Coefficients

	Negative		Positive		Signif.	Coefficients	
	All	Significant	All	Significant	Ratio	Median	Swamy
			A.1 Traditional Linder term — Narrow Sample				
Probit	92 (60.13%)	28 (18.30%)	61 (39.87%)	16 (10.46%)	0.571	-0.021	-0.023
FGLS (2^{nd} st.)	62 (42.18%)	28 (19.05%)	85 (57.82%)	45 (30.61%)	1.607	0.055	0.059
			A.2 Overlap Index — Narrow Sample				
Probit	66 (43.42%)	16 (10.53%)	86 (56.58%)	29 (19.08%)	1.813	0.018	0.030
FGLS (2^{nd} st.)	56 (38.36%)	26 (17.81%)	90 (61.64%)	56 (38.36%)	2.154	0.127	0.093
			B.1 Traditional Linder term — Broad Sample				
Probit	87 (55.77%)	31 (19.87%)	69 (44.23%)	30 (19.23%)	0.968	-0.005	-0.000
FGLS (2^{nd} st.)	65 (41.67%)	55 (35.26%)	91 (58.33%)	77 (49.36%)	1.400	0.017	0.016

Note:
Total amount and percentage of coefficients in parentheses reported in columns 2-5.
The significance ratio is defined as the ratio of the number of significantly negative to significantly positive sectors.
The numbers of sectors vary across specifications because for the Maximum Likelihood based methods (Probit) sectors in which no convergence of the coefficients was achievable were omitted, or because of insufficient observations at the second stage.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Descriptive Statistics

Figure 1 plots income dispersion in 2004—calculated using the LIS data—taking into consideration only incomes above the first decile and below the ninth decile. They are illustrated in ascending order of their decile ratios. Note that adjusted income as defined above is illustrated rather than household disposable income. Table 1 summarizes key income related statistics for the narrow 2004 sample.

Figures 2a and 3a show simple correlation between log-transformed export values and the log-transformed overlap index and the Linder term at the aggregate level respectively for the narrow sample. Both measures have a strongly positive relation with exports. A univariate linear model fit to the data for both indices seems to suggest that the Linder term has more explanatory power—the respective model's R^2 amounts to about 17% compared to a meager 6% for the overlap Index. However, note that the Linder term conveys information on per-capita GDP which will be controlled for later on. What is more, India could be an outlier. After all it is the poorest country in the sample, and plotting the Linder term against the Overlap Index in fact reveals that the overlap indices for India are always somewhat remote from the rest of the sample. However, dropping India from the sample increases the R^2 to merely 9%. Still, crude correlation does not imply causation—as is well known—and authors like Kennedy and McHugh (1980) have argued that the correlation is largely due to geographical concentration of similar countries. I.e., the correlations might simply—to a different extent—capture the effect of geographical distance.

Figure 4 reproduces figure 2a using only such country pairs that were excluded in the narrow sample. Apparently, the correlation is quite similar to the narrow sample.

5 Results & Discussion

Table 2 shows benchmark results for regressions for differentiated products according to Rauch's (1999) classification, disaggregated at the three digit level. Differentiated products are used because it seems to be commonplace in the literature that for such goods the theory is most likely to hold (cf. Hallak 2010, Bernasconi 2013). Moreover, Bastos and Silva (2010) argue that differentiated products, according to the Rauch (1999) classification coincide with vertically differentiated products as required by theory.

Results are reported for the two measures of similarity at both margins using the narrow sample. Additionally, results for the broader sample using the per capita term are reported, in order to unveil potential peculiarities of the sample at hand. The first four columns report number and share of sectors in which coefficients turned out to be negative, significantly negative, positive, and significantly positive. A coefficient is designated significant if it is significant at the 10% level. The significance ratio—reported in column 5—is defined as the ratio of the number of sectors in which the coefficient is significantly positive and significantly negative, i.e. a ratio above 1 indicates that on average the Linder effect is found. In addition to the median coefficient in column 6, in column 7 the weighted average coefficient are reported, where the inverse of the standard errors were used as weights in accordance with Swamy (1970). In general, the Swamy coefficient will be considered more sensible as it downweighs sectors with a large error variance. Panel A.1 reports results for the Linder term in the narrow sample, A.2 has results for the Overlap Index, and panel B reports results for the broad sample using the per capita measure only.

Apparently, the small sample generally underestimates significance, whereas qualititively results are similar, by and large. Median and Swamy coefficients indicate absence of the Linder effects at the extensive margin in both cases. Also, in both cases the percentage of coefficents with negative sign is more than half. Results are farther from Linder in the narrow sample, however. At the intensive margin on the other hand, results seem more favourable using the narrow sample. Only the percentage of significantly positve coefficients is lower than in the broad sample. In both cases evidence is stronger at the intensive margin, while at the extensive margin evidence is at best mixed. It seems like the Linder effect is not working through the extensive margin at all, when looking at the traditional Linder term.

When looking at the more sophisticated measure things change. The significance ratio and percentage of significantly positive coefficients increases, median and Swamy coefficients turn positive at the extensive margin. Still evidence seems weaker by all standards compared with the intensive margin.

Thus, it seems the first two hypotheses can be answered in the affirmative. For all specifications the effect is stronger at the intensive margin—indicating an adverse fixed costs effect—and the strict non-homotheticity measure outperforms the traditional Linder term by all standards.

Control variables are reported in table 3. All variables appear to have the expected signs. Again, the small sample underestimates significance for most variables. However, the additional controls from the Helpman et al. model are significant reasonably often. The exclusion is significant only in about 18% to 19% of the cases. However, recall that an exclusion is not always necessary in selection models. Also, given the percentage of significant coefficients for the other variables, it will be hard to find a variable that would significantly improve on that.

Aggregation Bias

Next, we will turn to the issue of aggregation bias. Table 4 presents results at all levels of aggregation. In addition to the specifications used in table 2 a Heckman (1979) selection model is fit (the difference to the Helpman et al. (HMR) version being that only the inverse Mill's ratio is included), as well as a simple FGLS without correction for zero trade flows or firm level heterogeneity. Note that the two-stage model could not be fit at the aggregate level (0-digit) for lack of variation in the trade indicator variable. The aggregation effect is clearly present if zero trade flows are not controlled for irrespective of similarity measure or sample. Interestingly, it seems to suffice to control for zero trade flows (using the Heckman specification) in order to make the aggregation effect largely disappear. Instead, for the traditional Linder term it seems like the bias has migrated to the extensive margin when looking at the coefficients. This is consistent with the hypothesis that fixed costs are driving the results. It is likely that fixed costs vary more across sectors, as quality differentiation might be achievable by rather simple means, i.e. using lower quality input materials. The effect of lower training and search costs

Table 3: 2-stage FGLS control variables

	Negative		Positive		Signif.	Coefficients	
	All	Significant	All	Significant	Ratio	Median	Swamy
			A. Traditional Linder term				
1st stage							
ldist	151 (98.69%)	136 (88.89%)	2 (1.31%)	0 (0.00%)	∞	-1.103	-1.120
rel	59 (38.56%)	10 (6.54%)	94 (61.44%)	27 (17.65%)	2.700	0.343	0.279
lang	23 (16.55%)	1 (0.72%)	116 (83.45%)	64 (46.04%)	64.000	0.596	0.868
contig	23 (27.38%)	7 (8.33%)	61 (72.62%)	28 (33.33%)	4.000	0.188	0.355
rta	51 (33.55%)	12 (7.89%)	101 (66.45%)	31 (20.39%)	2.583	0.284	0.271
col	8 (50.00%)	3 (18.75%)	8 (50.00%)	1 (6.25%)	0.333	0.006	-0.069
2nd stage							
ldist	100 (68.03%)	66 (44.90%)	47 (31.97%)	26 (17.69%)	0.394	-1.065	-1.049
lang	59 (44.36%)	27 (20.30%)	74 (55.64%)	33 (24.81%)	1.222	0.219	0.160
contig	29 (34.52%)	12 (14.29%)	55 (65.48%)	36 (42.86%)	3.000	0.302	0.647
rta	52 (35.62%)	24 (16.44%)	94 (64.38%)	48 (32.88%)	2.000	0.563	0.422
col	6 (40.00%)	1 (6.67%)	9 (60.00%)	0 (0.00%)	∞	0.155	0.590
ivm	60 (40.82%)	19 (12.93%)	87 (59.18%)	50 (34.01%)	2.632	1.059	1.085
zst	51 (34.69%)	17 (11.56%)	96 (65.31%)	47 (31.97%)	2.765	1.879	2.105
zst2	84 (57.14%)	41 (27.89%)	63 (42.86%)	17 (11.56%)	0.415	-0.360	-0.371
zst3	66 (44.90%)	19 (12.93%)	81 (55.10%)	37 (25.17%)	1.947	0.027	0.029
			B. Overlap Index				
1st stage							
ldist	151 (99.34%)	135 (88.82%)	1 (0.66%)	1 (0.66%)	0.007	-1.101	-1.111
rel	55 (36.18%)	9 (5.92%)	97 (63.82%)	29 (19.08%)	3.222	0.388	0.314
lang	24 (17.39%)	1 (0.72%)	114 (82.61%)	58 (42.03%)	58.000	0.583	0.863
contig	24 (28.57%)	6 (7.14%)	60 (71.43%)	26 (30.95%)	4.333	0.206	0.350
rta	56 (37.33%)	18 (12.00%)	95 (63.33%)	28 (18.67%)	1.556	0.953	0.181
col	8 (50.00%)	3 (18.75%)	8 (50.00%)	2 (12.50%)	0.667	0.005	-0.092
2nd stage							
ldist	103 (70.55%)	68 (46.58%)	43 (29.45%)	22 (15.07%)	0.324	-1.124	-1.094
lang	62 (47.33%)	28 (21.37%)	69 (52.67%)	36 (27.48%)	1.286	0.076	0.145
contig	30 (36.14%)	12 (14.46%)	53 (63.86%)	36 (43.37%)	3.000	0.309	0.580
rta	54 (37.50%)	23 (15.97%)	90 (62.50%)	44 (30.56%)	1.913	0.364	0.413
col	7 (43.75%)	1 (6.25%)	9 (56.25%)	1 (6.25%)	1.000	0.094	0.404
ivm	50 (34.25%)	16 (10.96%)	96 (65.75%)	51 (34.93%)	3.188	1.150	1.182
zst	50 (34.25%)	14 (9.59%)	96 (65.75%)	43 (29.45%)	3.071	1.683	2.122
zst2	83 (56.85%)	37 (25.34%)	63 (43.15%)	15 (10.27%)	0.405	-0.392	-0.370
zst3	62 (42.47%)	16 (10.96%)	84 (57.53%)	29 (19.86%)	1.813	0.024	0.028

Note:
Total amount and percentage of coefficients in parentheses reported in columns 2-5.
The significance ratio is defined as the ratio of the number of significantly negative to significantly positive sectors.
The numbers of sectors vary across specifications because for the Maximum Likelihood based methods (Probit) sectors in which no convergence of the coefficients was achievable were omitted, or because of insufficient observations at the second stage.

Table 4: Aggregation Bias

	Negative		Positive		Signif.	Coefficients	
	All	Significant	All	Significant	Ratio	Median	Swamy
			A.1 Traditional Linder term — Narrow Sample				
Probit							
3-digit	92 (60.53%)	28 (18.42%)	60 (39.47%)	16 (10.53%)	0.571	-0.021	-0.024
2-digit	31 (59.62%)	9 (17.31%)	21 (40.38%)	7 (13.46%)	0.778	-0.023	-0.026
1-digit	5 (71.43%)	1 (14.29%)	2 (28.57%)	1 (14.29%)	1.000	-0.046	-0.040
FGLS (HMR)							
3-digit	61 (41.78%)	28 (19.18%)	85 (58.22%)	45 (30.82%)	1.607	0.056	0.059
2-digit	24 (51.06%)	9 (19.15%)	23 (48.94%)	12 (25.53%)	1.333	-0.004	0.018
1-digit	2 (33.33%)	0 (0.00%)	4 (66.67%)	1 (16.67%)	∞	0.080	0.095
FGLS (Heckman)							
3-digit	62 (41.89%)	29 (19.59%)	86 (58.11%)	48 (32.43%)	1.655	0.051	0.038
2-digit	23 (47.92%)	12 (25.00%)	25 (52.08%)	10 (20.83%)	0.833	0.006	0.009
1-digit	1 (16.67%)	0 (0.00%)	5 (83.33%)	3 (50.00%)	∞	0.181	0.168
FGLS ($y > 0$)							
3-digit	75 (48.08%)	47 (30.13%)	81 (51.92%)	55 (35.26%)	1.170	0.005	0.012
2-digit	33 (60.00%)	22 (40.00%)	22 (40.00%)	15 (27.27%)	0.682	-0.023	-0.024
1-digit	5 (50.00%)	5 (50.00%)	5 (50.00%)	4 (40.00%)	0.800	0.019	-0.039
0-digit	1	1	0	0			-0.070***
			A.2 Overlap Index — Narrow Sample				
Probit							
3-digit	66 (43.71%)	16 (10.60%)	85 (56.29%)	29 (19.21%)	1.813	0.018	0.031
2-digit	25 (49.02%)	8 (15.69%)	26 (50.98%)	8 (15.69%)	1.000	0.000	0.018
1-digit	3 (42.86%)	0 (0.00%)	4 (57.14%)	1 (14.29%)	∞	0.022	0.075
FGLS (HMR)							
3-digit	56 (38.62%)	26 (17.93%)	89 (61.38%)	56 (38.62%)	2.154	0.128	0.093
2-digit	21 (43.75%)	8 (16.67%)	27 (56.25%)	15 (31.25%)	1.875	0.089	-0.009
1-digit	1 (16.67%)	0 (0.00%)	5 (83.33%)	4 (66.67%)	∞	0.351	0.325
FGLS (Heckman)							
3-digit	45 (30.61%)	20 (13.61%)	102 (69.39%)	63 (42.86%)	3.150	0.119	0.112
2-digit	21 (42.86%)	5 (10.20%)	28 (57.14%)	15 (30.61%)	3.000	0.043	0.088
1-digit	1 (16.67%)	0 (0.00%)	5 (83.33%)	4 (66.67%)	∞	0.370	0.322
FGLS ($y > 0$)							
3-digit	49 (31.41%)	27 (17.31%)	107 (68.59%)	83 (53.21%)	3.074	0.081	0.074
2-digit	24 (43.64%)	15 (27.27%)	31 (56.36%)	22 (40.00%)	1.467	0.048	0.014
1-digit	5 (50.00%)	3 (30.00%)	5 (50.00%)	4 (40.00%)	1.333	0.022	-0.007
0-digit	1	1	0	0			-0.053***

Continued on next page

Table 4: Aggregation Bias (Continued)

	Negative		Positive		Signif.	Coefficients	
	All	Significant	All	Significant	Ratio	Median	Swamy
			B. Traditional Linder term — Broad Sample				
Probit							
3-digit	87 (56.13%)	31 (20.00%)	68 (43.87%)	29 (18.71%)	0.935	-0.006	-0.001
2-digit	37 (67.27%)	17 (30.91%)	18 (32.73%)	5 (9.09%)	0.294	-0.021	-0.019
1-digit	5 (71.43%)	1 (14.29%)	2 (28.57%)	1 (14.29%)	1.000	-0.046	-0.040
FGLS (HMR)							
3-digit	64 (41.29%)	54 (34.84%)	91 (58.71%)	77 (49.68%)	1.426	0.018	0.016
2-digit	16 (29.09%)	12 (21.82%)	39 (70.91%)	33 (60.00%)	2.750	0.032	0.055
1-digit	2 (33.33%)	0 (0.00%)	4 (66.67%)	1 (16.67%)	∞	0.080	0.095
FGLS (Heckman)							
3-digit	64 (41.03%)	55 (35.26%)	92 (58.97%)	76 (48.72%)	1.382	0.011	0.013
2-digit	23 (41.82%)	18 (32.73%)	32 (58.18%)	28 (50.91%)	1.556	0.011	0.014
1-digit	1 (16.67%)	0 (0.00%)	5 (83.33%)	3 (50.00%)	∞	0.181	0.168
FGLS ($y > 0$)							
3-digit	62 (39.74%)	50 (32.05%)	94 (60.26%)	79 (50.64%)	1.580	0.014	0.011
2-digit	24 (43.64%)	21 (38.18%)	31 (56.36%)	26 (47.27%)	1.238	0.007	0.002
1-digit	5 (50.00%)	5 (50.00%)	5 (50.00%)	4 (40.00%)	0.800	0.019	-0.039
0-digit	1	1	0	0			-0.072***

Note:
Total amount and percentage of coefficients in parentheses reported in columns 2-5.
The significance ratio is defined as the ratio of the number of significantly negative to significantly positive sectors.
The numbers of sectors vary across specifications because for the Maximum Likelihood based methods (Probit) sectors in which no convergence of the coefficients was achievable were omitted, or because of insufficient observations at the second stage.

for similar countries' MNEs (Multinational Enterprises) could make FDI relatively more attractive across sectors and thus increase the fixed opportunity costs of trade. However, note that neither the significance ratio nor the percentage of significantly positive coefficients show this pattern. Nothing of the sort shows up for the narrow sample and the Overlap Index. One reason could be, as mentioned before, that overlap is less likely to be linked to fixed costs than similarity in per capita income.

Also note that moving from 3-digits to 2-digits in the small sample there is strange deterioration in the results when looking at the median or—in case of the Overlap Index—at the Swamy coefficient. Nothing of that sort, however, happens in the broad sample. Quite the contrary, after zero trade flows are controlled for results at the intensive margin seem to even get better with aggregation. One might thus consider the movements in the narrow sample as irregularities.

Robustness Checks

Robustness checks are provided in table 5. In Panels A.1 and B.1 small sectors were taken out of consideration, i.e. sectors the corresponding 2-digit sector of which comprised trade worth less than \$5 billion, and 3-digit sectors in which total trade amounts to less than \$2 billion both among the broad sample countries in 2004. Additionally, sectors belonging to one-digit category 9, i.e. "commodities and transactions not classified elsewhere in the SITC", were dropped. These criteria correspond to the ones used in Hallak (2010).[23] Panels A.2 and B.2 use Bernasconi's (2013) screening procedure

[23]The sectors that were dropped in this exercise are designated accordingly in table A.2 in the appendix

which deletes trade flows smaller than \$2,000, since small trade flows are assumed to be more prone to measurement error.

Panels A.1, B.1, A.2 and B.2 strongly reinforce some of the results obtained earlier. The Overlap Index fares better by all standards with the negligible exception of the significance ratio at the second stage in panel A.1. Results look much better now at the extensive margin in both cases, however, the percentage of significantly positive coefficients is still much lower than at the intensive margin.

In panels A.3 and B.3 India, identified as a potential outlier earlier is dropped. This slightly improves results without qualitatively changing the picture.

As mentioned above in early studies it was argued that income similarity was confounded with distance. This can be ruled out as using alternative measures of distance—CEPII's distance between capitals (panels A.4 and B.4) and population weighted distance (panel A.5 and B.5)—has virtually no effect on the results obtained.

Finally, for the traditional Linder term a test for the importance of the supply side of Linder's hypothesis suggests itself. In panel A.6 exporter GDP per capita is replaced by an index of export sophistication due to Hausmann et al. (2007). The index has the same scale as GDP per capita [24]. Let x_{ik} denote exports of product k by country i, and X_i total exports of country i than export sophistication is defined as:

$$EXPY_i = \sum_j \sum_l \frac{x_{il}}{X_i} \frac{x_{il}/X_j}{\sum_j x_{jl}/X_j} y_j \tag{9}$$

It is straightforward to see that this is a weighted average of world GDP per capita. The EXPY index was calculated at the SITC2 level. As can be seen in figure A.1 in the appendix the peak of the distribution of EXPY is to the right of the peak of GDP per capita. Also, the distribution is less dispersed. Both indicates that in the range of middle income and low income countries, the export portfolio is relatively sophisticated in comparison to their GDP, while it is less so for richer countries—on average, that is. As it measures sophistication *across* sectors EXPY cannot account for quality differentiation or within sector horizontal differentiation. Should this redefined Linder term yield better results it thus indicates that it is not the quality side that matters most. At the same time it sheds light on the question to what extent emerging economies can gain market power in differentiated products by upgrading their export portfolio.

While there still is a positive effect at the intensive margin, it is much smaller now. This indicates that the effect is not just a matter of inter sectoral upgrading, and reinforces the hypothesis that quality differentiation is important with respect to bilateral trade patterns.

However, it could still be that horizontal differentiation is more important at the sector level. In an attempt to see what is driving the results I split the set of differentiated goods sectors according to a simple rule. If the number of 4-digit and 5-digit subsectors was above average for a 3-digit sector, it was assumed that horizontal differentiation was more prevalent than vertical (quality) differentiation. Sectors in which horizontal differentiation was more prevalent were thus distinguished from sectors in which vertical differentiation was more prevalent. Evidently, the hypothesis does not hold for all sectors. But if the hypothesis does not hold for sectors with a high degree of hoizontal differentiation, this does not put into question the validity of the quality Linder hypothesis. Results are reported in table A.3 in the appendix. It seems that results are qualitatively similar. The Swamy coefficients seem to suggest that at the intensive margin results are stronger for vertically differentiated goods. This is in line with the percentage of significant coefficients. Pretty much every other criterion is inconclusive.[25]

[24] Figure A.1 in the appendix shows the estimated probability density functions of EXPY and GDP per capita estimated using STATA's *kdensity* command

[25] In another experiment, I seperated sectors that included the terms "parts" or "nes" in their product descruption. Both terms seem to indicate horizontal differentiation (albeit not ruling out vertical differentiation). Again, no structural differences between the new samples were visible. Table A.5 in the appendix lists all differentiated good sectors, providing information on subsectors at lower levels of aggregation. Additionally, it has been highlighted if the sector description contains the words "parts" or "nes". The table suggests that a fair amount of sectors could comprise horizontally differentiated goods.

Table 5: Robustness checks

	Negative		Positive		Signif.	Coefficients	
	All	Significant	All	Significant	Ratio	Median	Swamy
			A. Traditional Linder term				
1.: Small Sectors Excluded							
Probit	68 (59.13%)	16 (13.91%)	47 (40.87%)	12 (10.43%)	0.750	-0.021	-0.016
FGLS (2^{nd} st.)	44 (40.37%)	22 (20.18%)	65 (59.63%)	30 (27.52%)	1.364	0.055	0.054
2.: Trade Flows < USD 2,000 Excluded							
Probit	77 (54.23%)	27 (19.01%)	65 (45.77%)	20 (14.08%)	0.741	-0.015	-0.010
FGLS (2^{nd} st.)	48 (36.09%)	22 (16.54%)	85 (63.91%)	40 (30.08%)	1.818	0.075	0.057
3.: India Excluded							
Probit	80 (52.98%)	29 (19.21%)	71 (47.02%)	25 (16.56%)	0.862	-0.012	-0.011
FGLS (2^{nd} st.)	52 (35.86%)	23 (15.86%)	93 (64.14%)	60 (41.38%)	2.609	0.216	0.160
4.: Distance between capitals							
Probit	90 (58.82%)	28 (18.30%)	63 (41.18%)	16 (10.46%)	0.571	-0.021	-0.023
FGLS (2^{nd} st.)	55 (37.41%)	27 (18.37%)	92 (62.59%)	47 (31.97%)	1.741	0.077	0.065
5.: Weighted distance							
Probit	92 (59.74%)	30 (19.48%)	62 (40.26%)	16 (10.39%)	0.533	-0.028	0.001
FGLS (2^{nd} st.)	60 (40.82%)	28 (19.05%)	87 (59.18%)	45 (30.61%)	1.607	0.059	0.044
6.: EXPY as exporter supply							
Probit	116 (75.32%)	57 (37.01%)	38 (24.68%)	6 (3.90%)	0.105	-0.449	-0.561
FGLS (2^{nd} st.)	71 (48.63%)	36 (24.66%)	75 (51.37%)	39 (26.71%)	1.083	0.136	0.022
			B. Overlap Index				
1.: Small Sectors Excluded							
Probit	45 (39.47%)	10 (8.77%)	69 (60.53%)	21 (18.42%)	2.100	0.028	0.046
FGLS (2^{nd} st.)	45 (41.28%)	22 (20.18%)	64 (58.72%)	40 (36.70%)	1.818	0.098	0.080
2.: Trade Flows < USD 2,000 Excluded							
Probit	54 (37.76%)	11 (7.69%)	89 (62.24%)	34 (23.78%)	3.091	0.049	0.065
FGLS (2^{nd} st.)	48 (36.36%)	22 (16.67%)	84 (63.64%)	49 (37.12%)	2.227	0.148	0.097
3.: India Excluded							
Probit	65 (43.05%)	16 (10.60%)	86 (56.95%)	29 (19.21%)	1.813	0.041	0.055
FGLS (2^{nd} st.)	55 (38.46%)	18 (12.59%)	88 (61.54%)	56 (39.16%)	3.111	0.157	0.138
4.: Distance between capitals							
Probit	66 (43.42%)	15 (9.87%)	86 (56.58%)	31 (20.39%)	2.067	0.019	0.033
FGLS (2^{nd} st.)	55 (37.67%)	25 (17.12%)	91 (62.33%)	54 (36.99%)	2.160	0.145	0.099
5.: Weighted distance							
Probit	73 (47.71%)	19 (12.42%)	80 (52.29%)	27 (17.65%)	1.421	0.005	0.013
FGLS (2^{nd} st.)	57 (39.04%)	27 (18.49%)	89 (60.96%)	50 (34.25%)	1.852	0.099	0.076

Note:
Total amount and percentage of coefficients in parentheses reported in columns 2-5.
The significance ratio is defined as the ratio of the number of significantly negative to significantly positive sectors.
The numbers of sectors vary across specifications because for the Maximum Likelihood based methods (Probit) sectors in which no convergence of the coefficients was achievable were omitted, or because of insufficient observations at the second stage.

On top of that, the analysis was run for homogenous and reference priced goods as well. Results are reported in table A.4 in the appendix. While results are notably weaker at the extensive margin, the picture is less clear at the intensive margin. Results for reference priced goods using the overlap measure are actually stronger by all standards than those reported in table 2.

These findings do certainly not warrant full ascription of Linder type effects to quality differentiation. Still, they do indicate that quality could be an important channel among others.

6 Conclusion

It was shown that the Linder-effect operates mainly at the intensive margin. Results at the extensive margin are much weaker and in some cases negative. This indicates that the Linder effect could be confounded with fixed cost effects that have not been properly controlled for otherwise. A potential reason is FDI. It might be easier to find and train employees and meet regulations in a country that produces similar things, thereby lowering the fixed costs of FDI. This in turn means that a firm will be relatively less likely to service the foreign market using exports. This is consistent with indications of aggregation bias at the extensive margin in the broad sample. However, in order to prove this assertion more research is necessary, explicitly modelling the relation of FDI and trade at both margins.

The effect at the extensive margin seems to create aggregation bias found in earlier studies (Hallak 2010). Those results were easily reproduced when not controlling for zero trade flows. It has to be conceded, however, that results concerning agreggation bias are much less clear when looking at the narrow sample. Productivity effects, on the other hand, seem to play less of a role—other than initially suspected.

In general results seem to be stronger when allowing for strict non-homotheticity and the measure seems to be less confounded with potential fixed cost effects.

It seems, then, that similarity does have a positive effect on trade or interconnection in general (if FDIs were to be taken into account). This potentially hampers the possibility in particular of poor and emerging countries to gain market power in richer countries. It was shown that upgrading the export portfolio (increasing the EXPY score) cannot fully make up for differences in demand. While there still was a positive effect of similarity when using the EXPY index instead of exporter GDP per capita, it was much smaller.

As for other product groups, results mainly seem to get worse at the extensive margin, consistent with a fixed costs story. Also, when splitting differentiated goods into different groups according to the extent of horizontal differentiation results at the intensive margin come in supportive of the Linder hypothesis for both samples. It seems, then, that while quality surely might be an important driver of Linder effects, it is probably not the only one. An alternative and promising candidate could be innovativeness. I.e. it could be that countries with similar demand structures produce and consume goods of a similar degree of novelty, and innovativeness. It is also possible that innovations in one country are more likely to be consumed in a country at a similar development stage. These conjectures, however, will have to be tested in future research.

References

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

Bastos, P. and Silva, J. 2010. 'Identifying vertically differentiated products'. *Economics Letters* 106, 32–34.

Bernasconi, Claudia. 2013. 'Similarity of Income Distributions and the Extensive and Intensive Margin of Bilateral Trade Flows'. *Working Paper* (115). University of Zurich.

Bils, M. and Klenow, P.J. 2001. 'Quantifying Quality Growth'. *American Economic Review* 91(4), 1006–1030.

Brainard, S. Lael. 1997. 'An Empirical Assessment of the Proximity-Concentration Trade-off Between Multinational Sales and Trade'. *American Economic Review* 87(4), 520–544.

Choi, C. 2002. 'The Linder hypothesis revisited'. *Applied Economics Letters* 9, 601–605.

Choi, Y.C., Hummels, D. and Xiang, C. 2009. 'Explaining import quality: the role of the income distribution'. *Journal of International Economics* 77, 265–275.

Dalgin, M., Trindade, V. and Mitra, D. 2008. 'Inequality, Non-homothetic Preferences, and Trade: A Gravity Approach'. *Southern Economic Journal* 74(3), 747–774.

Deardorff, A. 1984. Testing Trade Theories and Predicting Trade Flows. *in* Jones and Kenen., eds, 'Handbook of International Economics Vol. 1'. Amsterdam: North-Holland.

Eppinger, Peter S. and Felbermayr, Gabriel J. 2015. 'Bilateral Trade and similarity of income distributions: The role of second moments'. *Economics Letters* 126, 159–172.

Fajgelbaum, P., Grossmann, Gene M. and Helpman, E. 2011. 'Income Distribution, Product Quality, and International Trade'. *Journal of Political Economy* 119(4), 712–765.

Fajgelbaum, P., Grossmann, Gene M. and Helpman, E. 2015. 'A Linder Hypothesis for Foreign Direct Investment'. *Review of Economic Studies* 82.

Feenstra, R. C. 2004. *Advanced International Trade: Theory and Evidence.* Princeton University Press. Princeton, New Jersey.

Fieler, A. C. 2011. 'Nonhomotheticity and Bilateral Trade: Evidence and a Quantitative Explanation'. *Econometrica* 79(4), 1069–1101.

Flam, H. and Helpman, E. 1987. 'Vertical Product Differentiation and North-South Trade'. *The American Economic Review* 77(5), 810–822.

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

Hallak, Juan C. 2010. 'A Product-Quality View of the Linder Hypothesis'. *The Review of Economics and Statistics* 92(3), 453–466.

Hausmann, Ricardo, Hwang, Jason and Rodrik, Dani. 2007. 'What you export matters'. *Journal of Economic Growth* 12, 1–25.

Head, K. and Mayer, T. 2014. Gravity Equations: Workhorse,Toolkit, and Cookbook. *in* G. Gopinath, E. Helpman and K. Rogoff., eds, 'Handbook of International Economics'. Vol. 4. pp. 131–195.

Head, K., Mayer, T. and Ries, J. 2010. 'The erosion of colonial trade linkages after independence'. *Journal of International Economics* 81, 1–14.

Heckman, J.J. 1979. 'Sample Selection Bias as a Specification Error'. *Econometrica* 47(1), 153–161.

Helpman, E., Melitz, M. and Rubinstein, Y. 2008. 'Estimating Trade Flows: Trading Volumes and Trading Partners'. *The Quarterly Journal of Economics* 123(2), 441–487.

Helpman, Elhanan, Melitz, Marc J. and Yeaple, Stehpen R. 2004. 'Export Versus FDI with Heterogenous Firms'. *American Economic Review* 94(1), 300–316.

Heston, A., Summers, R. and Aten, B. 2002. ‘Penn World Table Version 6.1’. *Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania* .

Hunter, L. 1991. ‘The contribution of nonhomothetic preferences to trade’. *Journal of International Economics* 30, 345–358.

Kennedy, T.E. and McHugh, R. 1980. ‘An Intertemporal Test and Rejection of the Linder Hypothesis’. *Southern Economic Journal* 46(3), 898–904.

Linder, Staffan B. 1961. *An Essay on International Trade and Transformation*. Almquist & Wiksell. Stockholm.

Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (multiple countries; October 2015), Luxembourg: LIS.

Manning, W.G. and Mullahy, J. 2001. ‘Estimating Log Models: To Transform or Not to Transform?’. *Journal of Health Economics* 20(4), 461–494.

Maoz, Z. and Henderson, E.A. 2013. ‘The World Religion Dataset, 1945-2010: Logic, Estimates, and trends’. *International Interactions* 39(3), 265–291.

Markusen, James R. 2013. ‘Putting per-capita income back into trade theory’. *Journal of International Economics* 90, 255–265.

Markusen, James R. and Venables, Anthony J. 2000. ‘The theory of endowment, intra-industry and multi-national trade’. *Journal of International Economics* 52, 209–234.

Martínez-Zarzoso, I. 2013. ‘The log of gravity revisited’. *Applied Economics* 45, 311–327.

Mayer, T. and Zignago, S. 2011. ‘Notes on CEPII’s distances measures: The GeoDist database’. *CEPII Working Paper 2011-25* .

Mitra, D. and Trindade, V. 2005. ‘Inequality and trade’. *Canadian Journal of Economics* 38(4), 1253–1271.

Piketty, Thomas. 2014. *Capital in the 21st century*. Harvard University Press. Cambridge.

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

Reimer, Jeffrey J. and Hertel, Thomas W. 2010. ‘Nonhomothetic Preferences and International Trade’. *Review of International Economics* 18(2), 408–425.

Santos Silva, J. M. C. and Tenreyro, S. 2006. ‘The log of gravity’. *The Review of Economics and Statistics* 88(4), 641–658.

Schott, P. K. 2004. ‘Across-product versus within-product Specialization in International Trade’. *Quarterly Journal of Economics* 199(2), 647–678.

Simpson, E. H. 1951. ‘The Interpretation of Interaction in Contingency Tables’. *Journal of the Royal Statistical Society. Series B (Methodological)* 13(2), 238–241.

Swamy, P. A. V. B. 1970. ‘Efficient Inference in a Random Coefficient Regression Model’. *Econometrica* 38(2), 311–323.

Appendix

Table A.1: List of Countries

Algeria^{H}	Greece^{H,L}	Poland^{H,L}
Argentina^{H}	Hungary^{H}	Portugal^{H}
Australia^{H}	India^{H,L}	Romania^{H}
Austria^{H,L}	Indonesia^{H}	Russia^{L}
Bangladesh^{H}	Iran^{H}	Saudi Arabia^{H}
Belgium^{H}	Ireland^{H,L}	Singapore^{H}
Brazil^{H}	Israel^{H}	Slovakia^{H,L}
Bulgaria^{H}	Italy^{H,L}	Slovenia^{L}
Canada^{H,L}	Japan^{H}	South Africa^{H}
Chile^{H}	Luxembourg^{L}	South Korea^{H}
China^{H}	Lybia^{H}	Spain^{H,L}
Colombia^{H,L}	Malaysia^{H}	Sweden^{H}
Croatia^{H}	Mexico^{H,L}	Switzerland^{H,L}
Czech Republic^{H,L}	Morocco^{H}	Taiwan^{H}
Denmark^{H,L}	Netherlands^{H,L}	Thailand^{H}
Dominican Republik^{H}	New Zealand^{H}	Tunisia^{H}
Ecuador^{H}	Nigeria^{H}	Turkey^{H}
Egypt^{H}	Norway^{H,L}	Ukraine^{H}
Estonia^{L}	Pakistan^{H}	Uruguay^{L}
Finland^{H,L}	Paraguay^{H}	USA^{H,L}
France^{H}	Peru^{H,L}	Venezuela^{H}
Germany^{H,L}	Philippines^{H}	Vietnam^{H}
Great Britain^{H,L}		

Note: Superscripts H: Broad Sample; L: Narrow Sample

Table A.2: Sectors according to Rauch's (1999) classification, SITC Rev. 2

Differentiated Goods

001^{S}	211^{S}	278^{S}	572^{S}	635	662	696	726	759	782	844	884
011^{S}	212^{S}	291	583	641^{S}	663	697	727	761	783	845	885
034	223^{S}	292	584^{S}	642	664	699	728	762	784	846	892
048	233^{S}	322^{S}	591	651	665	711^{S}	736	763	785	847	893
056	244^{S}	323^{S}	592^{S}	652	666	713	737	764	786	848	894
057^{S}	245^{S}	335	598	653	667^{S}	714	741	771	791^{S}	851	895
058^{S}	248	431^{S}	611	654	672^{S}	716	742	772	792	871	896
061^{S}	267^{S}	524	613^{S}	655	673	718	743	773	793	872	897
071	268^{S}	533	621	656	678	721	744	774	812	873	898
073	269^{S}	541	625	657	679	722	745	775	821	874	899
098	271^{S}	551^{S}	628	658	693	723	749	776	831	881	931^{S}
111	273^{S}	553	633^{S}	659	694	724	751	778	842	882	941^{S}
121^{S}	277^{S}	554	634^{S}	661	695	725	752	781	843	883^{S}	951^{S}

Reference Priced Goods

011	046	062	223	273	335	514	533	592	653	674	686
012	047	072	233	274	341	515	541	598	654	676	693
014	048	074	247	278	351	516	562	611	655	677	699
022	054	081	251	288	411	522	582	634	661	678	776
025	056	091	266	292	431	523	583	641	662	682	778
034	057	112	267	322	511	524	584	642	671	683	
036	058	122	268	323	512	531	585	651	672	684	
037	061	211	271	334	513	532	591	652	673	685	

Homogeneous Products

001	024	043	058	075	222	261	281	333	562	682	687
011	025	044	061	081	232	263	282	411	634	683	688
012	035	045	071	091	246	264	287	423	651	684	689
022	041	054	072	121	247	265	288	424	667	685	961
023	042	057	074	211	251	268	289	522	681	686	971

Note: Superscript S denotes small sector; respective sector was dropped in some analysis

Table A.3: Different degrees of vertical specialization

	Negative		Positive		Signif.	Coefficients	
	All	Significant	All	Significant	Ratio	Median	Swamy
			A. Traditional Linder term				
1.: Vertically Differentiated							
Probit	91 (59.87%)	28 (18.42%)	61 (40.13%)	16 (10.53%)	0.571	-0.021	-0.023
FGLS (2^{nd} st.)	62 (42.47%)	28 (19.18%)	84 (57.53%)	45 (30.82%)	1.607	0.054	0.059
2.: Horizontally differentiated							
Probit	22 (59.46%)	6 (16.22%)	15 (40.54%)	5 (13.51%)	0.833	-0.018	-0.016
FGLS (2^{nd} st.)	13 (37.14%)	7 (20.00%)	22 (62.86%)	8 (22.86%)	1.143	0.081	0.046
			B. Overlap Index				
1.: Vertically Differentiated							
Probit	65 (43.05%)	16 (10.60%)	86 (56.95%)	29 (19.21%)	1.813	0.018	0.031
FGLS (2^{nd} st.)	56 (38.62%)	26 (17.93%)	89 (61.38%)	56 (38.62%)	2.154	0.128	0.092
2.: Horizontally differentiated							
Probit	18 (48.65%)	4 (10.81%)	19 (51.35%)	8 (21.62%)	2.000	0.016	0.047
FGLS (2^{nd} st.)	16 (45.71%)	4 (11.43%)	19 (54.29%)	11 (31.43%)	2.750	0.097	0.054

Note:
Total amount and percentage of coefficients in parentheses reported in columns 2-5. Significant means, significant at least at the 10% level.
The significance ratio is defined as the ratio of the number of significantly negative to significantly positive sectors.
The numbers of sectors vary across specifications because for the Maximum Likelihood based methods (Probit) sectors in which no convergence of the coefficients was achievable were omitted, or because of insufficient observations at the second stage.
Sectors were designated as horizontally differentiated if they had more 4-digit and 5-digit sectors than average.

Table A.4: Different Product Groups

	Negative		Positive		Signif.	Coefficients	
	All	Significant	All	Significant	Ratio	Median	Swamy
			A. Traditional Linder term				
1.: Homogenous Products							
Probit	46 (76.67%)	20 (33.33%)	14 (23.33%)	2 (3.33%)	0.100	-0.068	-0.062
FGLS (2^{nd} st.)	28 (46.67%)	14 (23.33%)	32 (53.33%)	16 (26.67%)	1.143	0.016	-0.012
2.: Reference Priced Products							
Probit	56 (60.87%)	22 (23.91%)	36 (39.13%)	5 (5.43%)	0.227	-0.014	-0.036
FGLS (2^{nd} st.)	45 (48.91%)	20 (21.74%)	47 (51.09%)	23 (25.00%)	1.150	0.018	0.022
			B. Overlap Index				
1.: Homogenous Products							
Probit	40 (66.67%)	12 (20.00%)	20 (33.33%)	3 (5.00%)	0.250	-0.030	-0.027
FGLS (2^{nd} st.)	20 (33.33%)	10 (16.67%)	40 (66.67%)	22 (36.67%)	2.200	0.109	0.087
2.: Reference Priced Products							
Probit	52 (56.52%)	15 (16.30%)	40 (43.48%)	9 (9.78%)	0.600	-0.015	-0.017
FGLS (2^{nd} st.)	28 (30.43%)	14 (15.22%)	64 (69.57%)	40 (43.48%)	2.857	0.120	0.138

Note:
Total amount and percentage of coefficients in parentheses reported in columns 2-5.
The significance ratio is defined as the ratio of the number of significantly negative to significantly positive sectors.
The numbers of sectors vary across specifications because for the Maximum Likelihood based methods (Probit) sectors in which no convergence of the coefficients was achievable were omitted, or because of insufficient observations at the second stage.

Table A.5: Differentiated goods SITC Rev.2 3-digit

SITC	Description	No. of sub-sectors 4-digit	5-digit
001	Live animals chiefly for food	1	0
011	Meat and edible meat offal, fresh, chilled or frozen	1	0
034	Fish, fresh, chilled or frozen	1	0
048	Cereal, flour or starch preparations of fruits or vegetables	3	2
056	Vegetables, roots and tubers, prepared or preserved, **nes**	1	2
057	Fruit and nuts, fresh, dried	1	0
058	Fruit, preserved, and fruits preparations	2	0
061	Sugar and honey	1	0
071	Coffee and coffee substitutes	1	0
073	Chocolate and other preparations containing cocoa, **nes**	1	0
098	Edible products and preparations, **nes**	1	9
111	Non-alcoholic beverages, **nes**	1	2
121	Tobacco unmanufactured; tobacco refuse	1	0
211	Hides and skins, excluding furs, raw	4	0
212	Furskins, raw	1	2
223	Seeds and oleaginous fruit, whole or broken, for other fixed oils	2	0
233	Synthetic rubber, latex, etc; waste, scrap of unhardened rubber	1	2
244	Cork, natural, raw and waste	1	2
245	Fuel wood and wood charcoal	1	2
248	Wood, simply worked, and railway sleepers of wood	3	4
267	Other man-made fibres suitable for spinning, and waste	1	2
268	Wool and other animal hair (excluding tops)	2	4
269	Old clothing and other old textile articles; rags	1	2
271	Fertilizers, crude	1	0
273	Stone, sand and gravel	1	3
277	Natural abrasives, **nes**	2	2
278	Other crude minerals	1	0
291	Crude animal materials, **nes**	2	11
292	Crude vegetable materials, **nes**	6	8
322	Coal, lignite and peat	1	0
323	Briquettes; coke and semi-coke; lignite or peat; retort carbon	1	3
335	Residual petroleum products, **nes** and related materials	1	3
431	Animal and vegetable oils and fats, processed, and waxes	1	2
524	Radioactive and associated material	1	0
533	Pigments, paints, varnishes and related materials	3	8
541	Medicinal and pharmaceutical products	5	14
551	Essential oils, perfume and flavour materials	1	0
553	Perfumery, cosmetics, toilet preparations, etc	1	0
554	Soap, cleansing and polishing preparations	3	0
572	Explosives and pyrotechnic products	3	2
583	Polymerization and copolymerization products	2	0
584	Regenerated cellulose; derivatives of cellulose; vulcanized fibre	2	2
591	Pesticides, disinfectants	3	2
592	Starches, insulin and wheat gluten; albuminoidal substances; glues	1	2
598	Miscellaneous chemical products, **nes**	3	12
611	Leather	4	5

Continued on next page

Table A.5: Differentiated goods SITC Rev.2 3-digit (Continued)

SITC	Description	No. of sub-sectors 4-digit	5-digit
613	Furskins, tanned or dressed; pieces of furskin, tanned or dressed	1	0
621	Materials of rubber	1	6
625	Rubber tires, tire cases, inner and flaps, for wheels of all kinds	5	2
628	Articles of rubber, **nes**	2	2
633	Cork manufactures	1	2
634	Veneers, plywood, "improved" wood and other wood, worked, **nes**	1	3
635	Wood manufactures, **nes**	3	3
641	Paper and paperboard	1	3
642	Paper and paperboard, precut, and articles of paper or paperboard	2	11
651	Textile yarn	2	9
652	Cotton fabrics, woven (not including narrow or special fabrics)	1	5
653	Fabrics, woven, of man-made fibres (not narrow or special fabrics)	4	9
654	Textile fabrics, woven, other than cotton or man-made fibres	5	10
655	Knitted or crocheted fabrics (including tubular, etc, fabrics)	2	4
656	Tulle, lace, embroidery, ribbons, trimmings and other small wares	1	6
657	Special textile fabrics and related products	8	14
658	Made-up articles, wholly or chiefly of textile materials, **nes**	5	17
659	Floor coverings, etc	7	13
661	Lime, cement, and fabricated construction materials	1	3
662	Clay and refractory construction materials	1	5
663	Mineral manufactures, **nes**	4	6
664	Glass	9	6
665	Glassware	2	5
666	Pottery	3	0
667	Pearl, precious and semi-precious stones, unworked or worked	1	0
672	Ingots and other primary forms, of iron or steel	1	5
673	Iron and steel bars, rods, shapes and sections	1	9
678	Tube, pipes and fittings, of iron or steel	2	0
679	Iron, steel casting, forging and stamping, in the rough state, **nes**	2	2
693	Wire products (excluding insulated electrical wire); fencing grills	1	2
694	Nails, screws, nuts, bolts, rivets, etc, of iron, steel or copper	1	3
695	Tools for use in the hand or in machines	3	8
696	Cutlery	1	6
697	Household equipment of base metal, **nes**	4	13
699	Manufactures of base metal, **nes**	6	20
711	Steam boilers and auxiliary plant; and **parts** thereof, **nes**	2	0
713	Internal combustion piston engines, and **parts** thereof, **nes**	4	2
714	Engines and motors, non-electric; **parts**, **nes**; group 714, item 71888	1	2
716	Rotating electric plant and **parts** thereof, **nes**	3	3
718	Other power generating machinery and **parts** thereof, **nes**	1	4
721	Agricultural machinery (excluding tractors) and **parts** thereof, **nes**	4	17
722	Tractors (other than those falling in heading 74411 and 7832)	1	0
723	Civil engineering, contractors' plant and equipment and **parts**, **nes**	2	7
724	Textile and leather machinery, and **parts** thereof, **nes**	6	17
725	Paper and paper manufacture machinery, and **parts** thereof, **nes**	3	4
726	Printing, bookbinding machinery, and **parts** thereof, **nes**	5	10

Continued on next page

Table A.5: Differentiated goods SITC Rev.2 3-digit (Continued)

SITC	Description	No. of sub-sectors 4-digit	5-digit
727	Food-processing machines (non-domestic) and **parts** thereof, **nes**	2	5
728	Other machinery, equipment, for specialized industries; **parts nes**	3	15
736	Metalworking machine-tools, **parts** and accessories thereof, **nes**	4	13
737	Metalworking machinery (other than machine-tools), and **parts**, **nes**	3	6
741	Heating and cooling equipment and **parts** thereof, **nes**	6	4
742	Pumps for liquids; liquid elevators; and **parts** thereof, **nes**	4	2
743	Pumps, compressors; centrifuges; filtering apparatus; etc, **parts**	4	0
744	Mechanical handling equipment, and **parts** thereof, **nes**	2	8
745	Other non-electric machinery, tools and mechanical apparatus, **nes**	2	9
749	Non-electric **parts** and accessories of machinery, **nes**	4	3
751	Office machines	3	10
752	Automatic data processing machines and units thereof	4	0
759	Parts, **nes** of and accessories for machines of headings 751 or 752	1	3
761	Television receivers	2	0
762	Radio-broadcast receivers	3	0
763	Gramophones, dictating machines and other sound recorders	2	4
764	Telecommunication equipment, **nes**; **parts** and accessories, **nes**	5	7
771	Electric power machinery, and **parts** thereof, **nes**	2	5
772	Electrical apparatus for making and breaking electrical circuits	2	0
773	Equipment for distribution of electricity	2	7
774	Electro-medical and radiological equipment	2	0
775	Household type equipment, **nes**	6	17
776	Thermionic, microcircuits, transistors, valves, etc	4	2
778	Electrical machinery and apparatus, **nes**	3	6
781	Passenger motor vehicles (excluding buses)	1	0
782	Lorries and special purposes motor vehicles	2	0
783	Road motor vehicles, **nes**	2	0
784	Motor vehicle **parts** and accessories, **nes**	1	0
785	Cycles, scooters, motorized or not; invalid carriages	3	2
786	Trailers, and other vehicles, not motorized, **nes**	2	5
791	Railway vehicles and associated equipment	1	0
792	Aircraft and associated equipment, and **parts** thereof, **nes**	2	3
793	Ships, boats and floating structures	3	8
812	Sanitary, plumbing, heating, lighting fixtures and fittings, **nes**	3	3
821	Furniture and **parts** thereof	3	7
831	Travel goods, handbags etc, of leather, plastics, textile, others	1	4
842	Men's and boys' outerwear, textile fabrics not knitted or crocheted	5	19
843	Womens, girls, infants outerwear, textile, not knitted or crocheted	6	24
844	Under garments of textile fabrics, not knitted or crocheted	3	9
845	Outerwear knitted or crocheted, not elastic nor rubberized	3	15
846	Under-garments, knitted or crocheted	5	12
847	Clothing accessories, of textile fabrics, **nes**	2	8
848	Articles of apparel, clothing accessories, non-textile, headgear	4	9
851	Footwear	1	5
871	Optical instruments and apparatus	1	5
872	Medical instruments and appliances, **nes**	1	3

Continued on next page

Table A.5: Differentiated goods SITC Rev.2 3-digit (Continued)

SITC	Description	No. of sub-sectors 4-digit	5-digit
873	Meters and counters, **nes**	2	0
874	Measuring, checking, analysis, controlling instruments, **nes**, **parts**	6	13
881	Photographic apparatus and equipment, **nes**	3	8
882	Photographic and cinematographic supplies	1	5
883	Cinematograph film, exposed and developed	1	0
884	Optical goods **nes**	2	4
885	Watches and clocks	2	11
892	Printed matter	3	10
893	Articles, **nes** of plastic materials	4	5
894	Baby carriages, toys, games and sporting goods	4	11
895	Office and stationary supplies, **nes**	3	10
896	Works of art, collectors' pieces and antiques	1	6
897	Gold, silver ware, jewelry and articles of precious materials, **nes**	1	0
898	Musical instruments, **parts** and accessories thereof	3	10
899	Other miscellaneous manufactured articles, **nes**	7	30
931	Special transactions, commodity not classified according to class	1	0
941	Animals, live, **nes**, (including zoo animals, pets, insects, etc)	1	0
951	Armoured fighting vehicles, war firearms, ammunition, **parts**, **nes**	1	6

Note: Highlighted wordings indicate different intermediate inputs or horizontal specialization. Source: Rauch (1999)

Table A.6: Reference priced goods SITC Rev.2 3-digit

SITC	Description	No. of sub-sectors 4-digit	5-digit
011	Meat and edible meat offal, fresh, chilled or frozen	3	2
012	Meat and edible meat offal, in brine, dried, salted or smoked	1	0
014	Meat and edible meat offal, prepared, preserved, nes; fish extracts	3	0
022	Milk and cream	1	0
025	Eggs, birds', and egg yolks, fresh, dried or preserved	1	0
034	Fish, fresh, chilled or frozen	3	0
036	Crustaceans and molluscs, fresh, chilled, frozen, salted, etc	1	0
037	Fish, crustaceans and molluscs, prepared or preserved, nes	2	0
046	Meal and flour of wheat and flour of meslin	1	2
047	Other cereal meals and flour	1	2
048	Cereal, flour or starch preparations of fruits or vegetables	1	2
054	Vegetables, fresh or simply preserved; roots and tubers, nes	4	8
056	Vegetables, roots and tubers, prepared or preserved, nes	2	3
057	Fruit and nuts, fresh, dried	5	15
058	Fruit, preserved, and fruits preparations	2	6
061	Sugar and honey	1	0
062	Sugar confectionery and preparations, non-chocolate	1	2
072	Cocoa	2	2
074	Tea and mate	1	0
081	Feeding stuff for animals (not including unmilled cereals)	4	12
091	Margarine and shortening	1	2
112	Alcoholic beverages	4	6
122	Tobacco, manufactured	2	0
211	Hides and skins, excluding furs, raw	1	2
223	Seeds and oleaginous fruit, whole or broken, for other fixed oils	2	0
233	Synthetic rubber, latex, etc; waste, scrap of unhardened rubber	1	7
247	Other wood in the rough or roughly squared	1	0
251	Pulp and waste paper	5	4
266	Synthetic fibres suitable for spinning	3	12
267	Other man-made fibres suitable for spinning, and waste	1	3
268	Wool and other animal hair (excluding tops)	1	0
271	Fertilizers, crude	1	0
273	Stone, sand and gravel	3	3
274	Sulphur and unroasted iron pyrites	1	0
278	Other crude minerals	5	17
288	Non-ferrous base metal waste and scrap, nes	1	0
292	Crude vegetable materials, nes	1	0
322	Coal, lignite and peat	1	0
323	Briquettes; coke and semi-coke; lignite or peat; retort carbon	1	2
334	Petroleum products, refined	1	2
335	Residual petroleum products, nes and related materials	3	9
341	Gas, natural and manufactured	1	2
351	Electric current	1	0
411	Animal oils and fats	1	3
431	Animal and vegetable oils and fats, processed, and waxes	3	2
511	Hydrocarbons, nes, and derivatives	4	15

Continued on next page

Table A.6: Reference priced goods SITC Rev.2 3-digit (Continued)

SITC	Description	No. of sub-sectors 4-digit	5-digit
512	Alcohols, phenols etc, and their derivatives	3	13
513	Carboxylic acids, and their derivatives	3	9
514	Nitrogen-function compounds	4	7
515	Organo-inorganic and heterocyclic compounds	4	6
516	Other organic chemicals	4	15
522	Inorganic chemical elements, oxides and halogen salts	3	17
523	Other inorganic chemicals; compounds of precious metals	4	26
524	Radioactive and associated material	1	2
531	Synthetic dye, natural indigo, lakes	2	2
532	Dyeing and tanning extracts, and synthetic tanning materials	2	2
533	Pigments, paints, varnishes and related materials	1	0
541	Medicinal and pharmaceutical products	2	4
562	Fertilizers, manufactured	2	11
582	Condensation, polycondensation and polyaddition products	8	16
583	Polymerization and copolymerization products	7	22
584	Regenerated cellulose; derivatives of cellulose; vulcanized fibre	2	5
585	Other artificial resins and plastic materials	2	3
591	Pesticides, disinfectants	1	0
592	Starches, insulin and wheat gluten; albuminoidal substances; glues	1	6
598	Miscellaneous chemical products, nes	1	5
611	Leather	2	0
634	Veneers, plywood, "improved" wood and other wood, worked, nes	2	2
641	Paper and paperboard	7	18
642	Paper and paperboard, precut, and articles of paper or paperboard	1	0
651	Textile yarn	5	25
652	Cotton fabrics, woven (not including narrow or special fabrics)	1	4
653	Fabrics, woven, of man-made fibres (not narrow or special fabrics)	3	7
654	Textile fabrics, woven, other than cotton or man-made fibres	1	0
655	Knitted or crocheted fabrics (including tubular, etc, fabrics)	1	0
661	Lime, cement, and fabricated construction materials	3	3
662	Clay and refractory construction materials	1	3
671	Pig and sponge iron, spiegeleisen, etc, and ferro-alloys	3	6
672	Ingots and other primary forms, of iron or steel	1	4
673	Iron and steel bars, rods, shapes and sections	2	9
674	Universals, plates, and sheets, of iron or steel	6	18
675	Hoop and strip of iron or steel, hot-rolled or cold-rolled	1	4
676	Rails and railway track construction materials, of iron or steel	1	2
677	Iron or steel wire (excluding wire rod), not insulated	1	4
678	Tube, pipes and fittings, of iron or steel	2	0
682	Copper	1	6
683	Nickel	1	4
684	Aluminium	1	6
685	Lead	1	4
686	Zinc	1	4
693	Wire products (excluding insulated electrical wire); fencing grills	2	3
699	Manufactures of base metal, nes	1	5
776	Thermionic, microcircuits, transistors, valves, etc	1	0

Continued on next page

Table A.6: Reference priced goods SITC Rev.2 3-digit (Continued)

SITC	Description	No. of sub-sectors 4-digit	5-digit
778	Electrical machinery and apparatus, nes	2	11

Note: Source: Rauch (1999)

Table A.7: Homogeneous goods SITC Rev.2 3-digit

SITC	Description	No. of sub-sectors 4-digit	5-digit
001	Live animals chiefly for food	4	6
011	Meat and edible meat offal, fresh, chilled or frozen	3	2
012	Meat and edible meat offal, in brine, dried, salted or smoked	1	0
022	Milk and cream	1	4
023	Butter	1	0
024	Cheese and curd	1	0
025	Eggs, birds', and egg yolks, fresh, dried or preserved	1	0
035	Fish, dried, salted or in brine; smoked fish	1	4
041	Wheat and meslin, unmilled	2	0
042	Rice	2	4
043	Barley, unmilled	1	0
044	Maize, unmilled	1	0
045	Cereals, unmilled	3	3
054	Vegetables, fresh or simply preserved; roots and tubers, nes	2	0
057	Fruit and nuts, fresh, dried	2	6
058	Fruit, preserved, and fruits preparations	1	7
061	Sugar and honey	3	0
071	Coffee and coffee substitutes	1	3
072	Cocoa	1	0
074	Tea and mate	1	0
075	Spices	2	7
081	Feeding stuff for animals (not including unmilled cereals)	1	9
091	Margarine and shortening	1	0
121	Tobacco unmanufactured; tobacco refuse	2	4
211	Hides and skins, excluding furs, raw	1	0
222	Seeds and oleaginous fruit, whole or broken, for 'soft' fixed oil	6	0
232	Natural rubber latex; rubber and gums	1	3
246	Pulpwood (including chips and wood waste)	1	3
247	Other wood in the rough or roughly squared	2	4
251	Pulp and waste paper	1	2
261	Silk	2	2
263	Cotton	3	0
264	Jute, other textile bast fibres, nes, raw, processed but not spun	1	0
265	Vegetable textile fibres, excluding cotton, jute, and waste	3	6
268	Wool and other animal hair (excluding tops)	3	0
281	Iron ore and concentrates	3	0
282	Waste and scrap metal of iron or steel	1	3
286	Ores and concentrates of uranium and thorium	1	0
287	Ores and concentrates of base metals, nes	8	10
288	Non-ferrous base metal waste and scrap, nes	1	6
289	Ores and concentrates of precious metals, waste, scrap	1	2
333	Crude petroleum and oils obtained from bituminous minerals	1	0
334	Petroleum products, refined	4	5
411	Animal oils and fats	1	5
423	Fixed vegetable oils, soft, crude refined or purified	4	2
424	Other fixed vegetable oils, fluid or solid, crude, refined	5	0

Continued on next page

Table A.7: Homogeneous goods SITC Rev.2 3-digit (Continued)

SITC	Description	No. of sub-sectors 4-digit	5-digit
522	Inorganic chemical elements, oxides and halogen salts	2	14
562	Fertilizers, manufactured	2	6
634	Veneers, plywood, "improved" wood and other wood, worked, nes	1	0
651	Textile yarn	2	14
667	Pearl, precious and semi-precious stones, unworked or worked	1	3
681	Silver, platinum and other metals of the platinum group	2	7
682	Copper	1	3
683	Nickel	1	0
684	Aluminium	1	0
685	Lead	1	3
686	Zinc	1	0
687	Tin	2	4
688	Uranium depleted in U235, thorium, and alloys, nes; waste and scrap	1	0
689	Miscellaneous non-ferrous base metals, employed in metallurgy	2	7
961	Coin (other than gold coin), not being legal tender	1	0
971	Gold, non-monetary (excluding gold ores and concentrates)	1	3

Note: Source: Rauch (1999)

Figure A.1: Distribution of Per-Capita GDP and Export Sophistication (EXPY)