Revisiting the Revisited: An Alternative Test of the Monopolistic Competition Model of International Trade

Isao Kamata

University of Michigan

This version: August 25, 2006

Abstract

A monopolistic competition model of international trade, such as the one presented by Helpman and Krugman (1985), implies that the volume of trade between countries, relative to the total income of the countries, will increase as the countries become more similar in income size. A few studies have empirically tested this implication for aggregate trade and GDP. However, international trade also involves goods that are non-differentiated and produced by multiple countries (i.e., specialization is incomplete), which are “out of the scope” of the monopolistic competition model. To test the model for aggregate trade may thus introduce noise and reduce the accuracy of the test.

This paper proposes an alternative test of the monopolistic competition model that focuses on the trade of differentiated products, which the model exactly aims to describe. The equation for the volume of trade is re-examined and re-derived for the trade in differentiated sectors, without the restrictive assumptions that have been imposed to derive the aggregate trade equation, such as perfect production specialization in all sectors. The derived equation implies that the volume of trade of differentiated products can be also predicted by the similarity between trading countries in their income size, but this income similarity must be adjusted by a factor indicating the extent to which the two countries are symmetric in production structure, or how similar the production share of differentiated products in the total economy is between the two countries. This alternative sectoral volume-of-trade equation is empirically tested using data on trade and production at the industry level that are classified into product categories based on an available study for the classification of traded goods. The results show that, regardless of country groups, the change in the volume of bilateral trade in differentiated sectors seems to be proportional to the change in production structure-adjusted size similarity between two trading countries. This supports the implication of the monopolistic competition model.

Keywords: Monopolistic competition and international trade; New Trade Theory; Gravity

JEL Classification: D43, F1, F12

* I am particularly grateful to Alan Deardorff, Juan Carlos Hallak, and Gary Saxonhouse for extensive advice and discussions. I also thank participants in the International Economics Seminar and the Applied Microeconomics Workshop at the University of Michigan, and participants in seminars at Hitotsubashi University, Kansai University, Keio University, Kobe University, and Yokohama National University, for their valuable comments and feedback on an earlier version of this work. I am solely responsible for all remaining errors.
Introduction

New Trade Theory is characterized by a model of international trade with monopolistic competition among the varieties of products in an industry. This theory was originally motivated by the fact that a large part of international trade is intra-industry rather than inter-industry,\(^1\)\(^2\) a characteristic that neo-classical trade theory such as the Heckscher-Ohlin (H-O) Model or the Ricardian Model cannot explain. The monopolistic competition models of international trade, first presented in the works of Krugman (1979, 1980) and Helpman (1981), have been widely employed and applied in numerous studies of international trade.

This type of model has implications for the volume of trade; in particular, the volume of trade among a group of countries, as a share of the total income of the country group, will be larger as the sizes of incomes of individual countries in the group are more similar to each other. In other words, if two regions have the same total income and consist of the same number of countries, the region in which countries are more equal in income size will trade more within that region.

Although this theoretical implication is clear-cut and has an empirically testable form, only a few studies have directly examined this implication empirically. Helpman (1987) employed time-series data on 14 OECD countries and graphically showed the positive relationship between the volume of trade among the 14 countries as a fraction of their total GDP and the similarity in their respective GDPs. Hummels and Levinsohn (1995) performed more formal empirical tests using panel data on bilateral trade flows between pairs of the same 14 OECD countries, as well as those of another 14 non-OECD countries. They expected that although the data on trade between OECD countries, which seemed to be more intra-industry

---

1 The significance of intra-industry trade has been reported by, for example, Grubel and Lloyd (1975).
2 On the other hand, it is debated in the literature whether such intra-industry trade, or “trade overlap,” observed in the data is a matter of the aggregation of sectors or commodities. See Finger (1975).
trade of differentiated products, would fit to the monopolistic competition model, it would not be the case for trade between non-OECD countries, which did not seem to be characterized by differentiated products within the same industry. Their results, however, showed that similarity between two trading countries in GDP well explains the volume of trade between the two countries, both for the OECD group and for the non-OECD group, which left a puzzle. Recently Debaere (2005) re-examined the work by Hummels and Levinsohn, and claimed that their empirical approach may not be able to properly estimate the sole impact of income similarity on bilateral trade, and thus led them to the puzzling result. He modified the equation explaining the volume of trade and estimated it using updated data for the same OECD and non-OECD samples, and he showed that, with the volume of trade adjusted for the total GDP of two trading countries, a positive association of bilateral trade volume to income similarity between trading countries is significant only for OECD countries. He thus concluded that, as initially expected even by Hummels and Levinsohn, the monopolistic competition model is relevant to explain trade among OECD countries but not to explain trade among non-OECD countries.3

These studies have attempted to test the monopolistic competition model in the context of aggregate trade, which includes all types of goods traded. However, not all goods internationally traded are differentiated products, and the trade of those non-differentiated goods may be driven by other mechanisms than what the monopolistic competition model describes. In fact, to expand the tested implication—the volume of trade will increase as trading countries become more equal in income size—to the level of aggregate trade, they needed to assume that all industries are internally differentiated in terms of product varieties, or alternatively that perfect specialization of production takes place in every sector, which is very restrictive and thus may not be realistic.

---

3 Appendix A reviews the work by Hummels & Levinsohn (1995) and by Debaere (2005).
Moreover, the monopolistic competition model suggests that the volume of trade will be perfectly proportional to size similarity among countries (i.e., the estimate for the coefficient equals one), and thus a strict empirical test should examine whether the relationship is proportional. However, the preceding studies do not include empirical test for this proportionality but only assessed the sign and significance of the coefficient estimates for income similarity.

In this paper, I propose an alternative empirical approach to testing the implication of the monopolistic competition model for the volume of trade. The key is to focus on the trade of differentiated products. I re-examine the model and derive the equation for the volume of bilateral trade of differentiated products. The trade equation for differentiated products can be derived without imposing the restrictive assumption mentioned above, but it does not depend on simple income similarity any more. The alternative equation says that the volume of bilateral trade of differentiated products as a share of the domestic production of these products in two trading countries will be proportional to income size similarity between the two countries adjusted by the symmetry of production structure between the countries. In other words, the volume of bilateral trade in differentiated sectors will be predicted by size similarity and symmetry in the share of differentiated sectors in the domestic production between two countries.

This implication must be tested with the data on trade and production in the sectors of differentiated products. In addition to the data on bilateral trade and GDP that the previous studies have also used, I employ data on manufacturing production for various countries from the United Nations, and I utilize the information on product classification by Rauch (1999) to classify the “differentiated sectors.” This enables me to avoid involving “noises” in aggregate trade. Furthermore, I test the model for strict proportionality of the relation between the volume
of trade and (production-adjusted) size similarity. That is, I perform tests of the hypothesis that estimated coefficients equal one. I also apply non-linear estimation methods, in addition to the benchmark log-linear estimation, to handle zero-trade observations in the data.

The results of the benchmark estimation, as well as of the Tobit estimation, show that (i) the empirical relationship between the volume of trade in the differentiated sectors and production structure-adjusted size similarity between countries is significant, regardless of country groups; and (ii) the relationship is inferred to be proportional; i.e., the change in the volume of bilateral trade in differentiated sectors will be proportional to the change in production structure-adjusted size similarity between two trading countries. These results seem to support the implication of the monopolistic competition model more strongly than the test with aggregate trade data.

This study offers some insight for a series of empirical studies on the gravity equation, to which the monopolistic competition model provides a theoretical background. Most studies have estimated the gravity equation for aggregate trade. For example, Feenstra, Markusen and Rose (2001), Evenett and Keller (2002), and Haveman and Hummels (2004) use the gravity equation for aggregate trade to test which theory of international trade is the most likely to explain the actual trade flows, following Deardorff (1998) pointing out that multiple trade theories can derive the gravity equation. The point of Feenstra et al. is the existence of a home-market effect that may distinguish the monopolistic competition model from others, while Evenett & Keller and Haveman & Hummels focus on the elasticity of national income with respect to the volume of trade, which will be smaller than unity if specialization in production is incomplete. However, aggregate trade involves the trade of various products, some of which the monopolistic

---

4 In fact, the result of my benchmark estimation for the case of aggregate trade is basically consistent with the finding of Evenett & Keller, as discussed later in this paper.
competition model fits well, but others may be characterized by product homogeneity and incomplete specialization; thus all trade should not be explained by a single model in a unified manner.\(^5\) On the other hand, Harrigan (1994), as well as Jensen (2000), have estimated the gravity equation at the sectoral level using data on trade and production in manufacturing industries.\(^6\) They, however, do not necessarily consider differences in product characteristics (differentiated versus homogeneous) across manufacturing industries, which this paper pays careful attention to.\(^7\)

The remainder of this paper is organized as follows. The next section derives the equation explaining the volume of differentiated-sector trade, and discusses its implication in comparison with the aggregate trade equation used in the existing literature. The section presenting the empirical methodologies and specifications follows. The data employed for the empirical analysis are described in section four. The results of the analysis are presented in the fifth section, and the concluding section discusses the ‘puzzles’ emerging from the empirical results.

**Monopolistic Competition Model and Volume of Trade**

In this section, to account for the volume of trade I derive two formulas from the monopolistic competition model of international trade introduced by Helpman and Krugman

---

\(^5\) Feenstra, Markusen & Rose also divide trade into three categories according to Rauch (1999) to estimate their gravity equation, but the explanatory variables are for the aggregate; i.e., GDPs of exporter and importer countries.

\(^6\) Harrigan introduced a variety of the proxies for scale economies in his equation to see whether the home-market effect would be significant, which would indicate the monopolistic competition rather than Armington preference for national varieties. Jensen’s interest was in the size of the estimated elasticity of volume of imports to the importer’s income.

\(^7\) Other empirical work such as Harrigan (2001) and Anderson and van Wincoop (2003) carefully derive a structural gravity-like equation from a generalized monopolistic competition model, but due to the unobservability of variables, their attention is limited to a certain factor such as distance or trade cost. Lai and Zhu (2004), on the other hand, made an extended effort to measure as many variables as possible to estimate their structural and generalized volume-of-trade equation with data.
(1985, Chapters 6-8). This model are characterized as follows: (i) some sectors have a number of product varieties (I hereinafter call these sectors “differentiated sectors”); (ii) each of the product varieties in a differentiated sector is produced monopolistically competitively by a single firm; (iii) consumers throughout the world have identical preferences that are characterized by a two-tier utility function (the upper-tier utility is homothetic, and the sub-utility over product varieties within a sector takes a CES functional form).

Here I consider an equilibrium of frictionless trade so that the price of each good or differentiated product is equal throughout the world. In this trade equilibrium, every product of the differentiated sectors produced in each country will be divided among all consumers worldwide, according to their share of world income. The volume of exports from one country to another is thus expressed as follows:

\[ EX_{i,j} = \sum_{s \in D} e_j p_s Q_{s,i} + \sum_{s \in H} EX_{s,i,j} , \]

where
- \( D \): the group of the differentiated sectors;
- \( H \): the group of homogeneous sectors;
- \( i,j \): scripts for countries \((i \neq j)\);
- \( EX_{s,i,j} \): exports from Country \( i \) to Country \( j \) in Sector \( s \);
- \( Q_{s,i} \): Country \( i \)'s production in Sector \( s \);
- \( X_{s,w} \): the world total production in Sector \( s \);
- \( p_s \): the equilibrium price of Sector \( s \);
- \( e_j \): Country \( j \)'s GDP share in the world \((= Y_j/Y_w)\).

Note that the volume of trade between a specific pair of countries in the sectors of homogeneous products (or “homogeneous sectors”), \( EX_{s,i,j} \) for \( s \in H \), is indeterminate. That is, although a country will export a homogeneous product if the amount of the product it produces domestically is greater than that consumed domestically, how much of its product will be imported by which
country(ies) cannot be determined because the product is homogeneous and its price is equal regardless of the origin of the product in the considered equilibrium.

**Aggregate volume of trade**

The version of the formula for the *aggregate* volume of trade, which has been employed in such studies as Helpman (1987), Hummels and Levinsohn (1995), and Debaere (2005), is derived by further assuming the following:

(A1) Each country in the world is completely specialized in the homogeneous sectors as well; that is, every homogeneous product is produced by no more than one country. Under this assumption, a product produced by a sole producer country (i.e., a sole exporter) will be imported by all other countries according to their share of world income, and thus no indeterminacy will exist in the quantities of bilateral trade. Therefore, the volume of exports from Country $i$ to Country $j$ is expressed as:

$$EX^j_i = \sum_{s \in D, H} e_j p_s Q_{s,j}.$$  

(A2) Products in any sector are tradable, i.e., there exist no non-traded sectors.8 Under this assumption, the aggregate value of a country’s production over the sectors equals its income, or GDP. That is;

$$\sum_{s \in D, H} p_s Q_{s,i} = Y_i;$$

$$EX^j_i = e_j Y_i.$$  

Therefore, following Helpman (1987), the total *aggregate* bilateral trade volume between Countries $i$ and $j$ is expressed as:

---

8 The following weaker assumption can also derive Equation (2) below, which expresses the volume of trade per GDP in terms of the similarity between countries in income: every country has an equal share of non-traded sectors in its GDP.
\[ VT_{ij} = EX_i^j + EX_i^j = e_i Y_i + e_j Y_j = 2Y_i Y_j / Y_w = e_{ij} \cdot 2(Y_i Y_j / Y_w) \]

\[ \Leftrightarrow VT_{ij} / Y_{ij} = e_{ij} [1 - (Y_i / Y_{ij})^2 - (Y_j / Y_{ij})^2] \]

where \( Y_{ij} = Y_i + Y_j \): Country \( i-j \) pair’s total GDP

\[ e_{ij} = Y_{ij} / Y_w: \] Country \( i-j \) pair’s world GDP share.

The term in the square brackets on the right-hand side of the second equation indicates similarity of the total incomes, or sizes of the economy, of the two trading countries. This term becomes greater as the incomes of the two countries become more equal, and takes the maximum value of 0.5 when the two countries are exactly equal in size; i.e., \( Y_i / Y_{ij} = Y_j / Y_{ij} = 1/2 \). Using this index of income “similarity,”

\[ VT_{ij} / Y_{ij} = e_{ij} \cdot \text{sim}_{ij}, \quad (2A) \]

where \( \text{sim}_{ij} = [1 - (Y_i / Y_{ij})^2 - (Y_j / Y_{ij})^2] \).

This Equation (2A) implies that the volume of aggregated bilateral trade, as a share of the total income (GDP) of the two trading countries, will be greater as their respective national incomes are more similar.

\[ Volume \ of \ trade \ in \ the \ differentiated \ sectors \]

The two assumptions A1 and A2 are very restrictive. Since Equation (2A) can be derived only with these restrictive assumptions, its validity would be equally limited. It is, however, possible to derive an alternative formula that can explain the volume of trade in the differentiated sectors without imposing these assumptions. Since countries are completely specialized in production of unique varieties in the differentiated sector, by taking the first term for the differentiated sectors in Equation (1);

\[ \text{Note that } Y_i / Y_{ij} = 1 - Yi / Y_{ij}. \] This index takes its minimum value of zero when two countries are completely dissimilar; i.e., \( Y_i / Y_{ij} = 0 \) and \( Y_j / Y_{ij} = 1 \), vice versa.

\[ \text{Helpman (1987), as well as Hummels & Levinsohn (1995), calls this term the “dispersion” index, while Debaere (2005) names it the “similarity index.” I follow the latter since this index being larger means two countries being more similar in income.} \]
\[ EX_{ij}^{D} = \sum_{s \in D} e_{js}Q_{s,ij} = e_{i}X_{i}^{D}, \]

where \( EX_{ij}^{D} \): exports in differentiated sectors from Country \( i \) to Country \( j \)

\( X_{i}^{D} \): Country \( i \)'s domestic production value in differentiated sectors:

\[ X_{i}^{D} = \sum_{s \in D} P_{s}Q_{s,i}. \]

Therefore, the volume of bilateral trade \textit{in the differentiated sectors} is expressed as:

\[ VT_{ij}^{D} = EX_{i}^{D} + EX_{j}^{D} = e_{i}X_{i}^{D} + e_{j}X_{j}^{D} = e_{i}\{(Y_{j} / Y_{i}) \cdot X_{i}^{D} + (Y_{i} / Y_{j}) \cdot X_{j}^{D}\} \]

\[ \Leftrightarrow VT_{ij}^{D} = e_{ij}\left(1 - \frac{X_{i}^{D}}{X_{j}^{D}} \cdot \frac{Y_{i}}{Y_{j}} - \frac{X_{j}^{D}}{X_{i}^{D}} \cdot \frac{Y_{j}}{Y_{i}}\right) = e_{ij}\left[1 - \left(\frac{X_{i}^{D}}{X_{j}^{D}} / Y_{j}\right) \cdot \left(\frac{Y_{i}}{Y_{j}}\right)^{2} - \left(\frac{X_{j}^{D}}{X_{i}^{D}} / Y_{i}\right) \cdot \left(\frac{Y_{j}}{Y_{i}}\right)^{2}\right], \]

where \( VT_{ij}^{D} \): volume of Country \( i-j \) bilateral trade in differentiated sectors

\( X_{ij}^{D} \): Countries \( i \& j \)'s total domestic production in differentiated sectors

\( (X_{ij}^{D} = X_{i}^{D} + X_{j}^{D}). \)

The term in the square brackets is similar to the income similarity index in the equation for aggregate trade (2A), but this term depends not only on two countries’ relative income sizes but also on the sizes of production in the differentiated sectors of the countries \( (X_{i}^{D}, X_{j}^{D}) \). The income-share term for each country is “weighted” by the term \( ((X_{i}^{D} / Y_{i})/(X_{j}^{D} / Y_{j})) \), and this “weight” term indicates how large is the share of the differentiated sectors in that country’s total production (or GDP), relative to the average GDP share of the differentiated-sector production across the two countries. In other words, in this term income similarity between two countries is \textit{adjusted by symmetry of the two countries in production structure}. This term becomes larger as two countries become more similar in income size \textit{and} more symmetric in production structure. I thus call this term the \textit{production structure-adjusted income (or size) similarity}, and re-write the equation as follows:

\[ VT_{ij}^{D} / X_{ij}^{D} = e_{ij} \cdot sim_{ij}^{*}, \quad (2D) \]

\[ \text{where } sim_{ij}^{*} = \left[1 - \left(\frac{X_{i}^{D}}{X_{j}^{D}} / Y_{j}\right) \cdot \left(\frac{Y_{i}}{Y_{j}}\right)^{2} - \left(\frac{X_{j}^{D}}{X_{i}^{D}} / Y_{i}\right) \cdot \left(\frac{Y_{j}}{Y_{i}}\right)^{2}\right]. \]

Equation (2D) implies that the volume of bilateral trade \textit{in the differentiated sectors}, as a share of the \textit{total production} in those sectors, will be predicted by size similarity between the two trading
countries adjusted by how symmetric their production structures are. That is, the volume of bilateral trade will increase as the two countries are more similar in size and more symmetric in production.

**Volume of Trade and Size Similarity: Intuition**

Why is the volume of trade predicted to be larger as trading countries are more similar? The first thing to note is that size similarity determined relative vigorousness of interaction between countries to other country pairs, having the absolute size of their production and income adjusted. (This is expressed as the denominator on the left-hand side $X_{ij}^D$ and the term $e_{ij}$ on the right-hand side.) Under complete production specialization, all kinds of products of one country will be consumed by others, and thus the volume of international trade will be largest when both a pool—supplier’s production—and the power of absorption—demander’s income—are simultaneously large. In the model of aggregate trade assuming complete specialization in all sectors, however, since a country’s product equals its total income, the (relative) size of production of one country and the (relative) income of another country cannot be large at the same time. In that case, the (relatively) small seller cannot have much to export to the (relatively) large buyer; and on the other hand, the large seller cannot export much to its partner because the buyer’s income is not great enough. Therefore, the volume of trade will be the largest when two countries are equal in size.

In the case of sectoral trade of differentiated products, however, it could be the case that one country is larger in production and another country is larger in income. In fact, the production structure-adjusted similarity index in Equation (2D) indicates this possibility. The index will be even larger when, rather than two countries being similar in size and symmetric in
production, the countries are asymmetric, or rather, *counter-symmetric*, in income size and production; i.e., a country with smaller income is a dominant producer in that sector.\textsuperscript{11}

**Empirical Specifications**

For the empirical tests of the volume of trade equation derived in the previous section, I apply the following empirical specifications. For each specification, I derive an equation to be estimated for each of aggregate volume of trade (Equation (2A)) and trade in differentiated sectors (Equation (2D)). The results of estimation of the two cases, which are presented in a later section, are compared to show how the proposed alternative empirical approach differs from the conventional approach of aggregate trade.

**OLS estimation of log-linear form:**

As the benchmark specification, I apply OLS to the log-linearized form of the volume-of-trade equation. Recall the equation for the volume of trade (2A) and (2D), but considering other potential factors affecting trade flows:\textsuperscript{12}

\[
\frac{VT_{ijt}}{Y_{ijt}} = e_{ijt}^{\beta_{1}} \cdot \text{sim}_{ijt}^{\beta_{2}} \cdot \mu_{ij} \cdot \epsilon_{ijt}
\]

\[
\frac{VT_{ijt}^{D}}{X_{ijt}^{D}} = e_{ijt}^{\beta} \cdot \text{sim}_{ijt}^{* \beta_{2}} \cdot \mu_{ij} \cdot \epsilon_{ijt}
\]

Although the underlying theoretical model explains a core mechanism determining the volume of trade as Equations (2A) and (2D) suggest, real trade flows may be affected by other factors. For

\textsuperscript{11} As a numerical example, consider the following case: if Country $i$ is nine times larger in income than Country $j$ (i.e., $Y_i/Y_j = 0.9$); but the smaller Country $j$ has production in differentiated sectors that is 95% of its total income (GDP), while Country $i$’s production in those sectors is only 5% of its income. In this case, the index value is 0.64, which exceeds the maximum 0.5 of that was possible when the countries were perfectly equal and symmetric.

\textsuperscript{12} Script $t$ denotes time period, which in this study is the year.
example, the literature on the gravity equation suggests that international trade flows will be affected by geographic factors such as distance, border sharing, and commonness of language. The term $\mu_{ij}$ captures these factors that are specific to country pairs, as well as other unobserved potential country pair-specific but time-invariant factors affecting bilateral trade flows. The last term $\varepsilon_{ijt}$ captures idiosyncratic disturbances to recorded trade flows or measurement error, which is assumed to be log-normally distributed (i.e., $\log(\varepsilon) \sim N(0, \sigma_\varepsilon)$). Taking the logarithm of both sides of these two equations yields the following equations:

\[
\log(VT_{ijt} / Y_{ijt}) = \beta_1 \cdot \log(e_{ijt}) + \beta_2 \cdot \log(sim_{ijt}) + \mu_{ij} + \varepsilon_{ijt} \tag{3A}
\]

\[
\log(VT_{ijt}^D / X_{ijt}^D) = \beta_1 \cdot \log(e_{ijt}) + \beta_2 \cdot \log(sim^*_{ijt}) + \mu_{ij} + \varepsilon_{ijt} \tag{3D}
\]

Equation (3A) represents the volume of aggregated bilateral trade derived from Equation (2A), which is the same as the main empirical specification employed by Debaere (2005). Equation (3D), which is derived from (2D), is the alternative empirical specification that this paper proposes, which is designed to account for trade in the differentiated sectors. Both equations are estimated by fixed-effect OLS regression with country pair-specific dummies ($\mu_{ij}$). Time-specific dummies are also included for estimation in order to capture any trend in or shocks to trade flows that are common for all countries in the world.

The coefficients ($\beta_1, \beta_2$) in Equations (3A) and (3D) are estimated separately for the samples of OECD and non-OECD countries. This is for comparison to the previous studies (Hummels & Levinsohn (1995) and Debaere (2005)) that used the aggregate trade equation like (3D). These studies separated estimation for a group of OECD countries and for non-OECD, based on the understanding that intra-industry trade of differentiated products seems dominant in

---

13 See Appendix A for more details of the empirical approach of Debaere, as well as Hummels & Levinsohn (1995).

14 See the next section for the list of the countries included in each sample.
trade between OECD countries, while trade between non-OECD countries does not seem to be characterized by product differentiation. Their expectation was thus that the data should fit the model well for OECD countries but not for non-OECD countries. Although Hummels & Levinsohn found an unexpected result (data support the model for both groups), Debaere reexamined their estimation and as a result found empirical support for the model only for OECD countries, as expected. In contrast, this study focuses on trade in the sectors of differentiated products, which the monopolistic competition model aims to explain for any country in the world. Therefore, the strict hypothesis here based on the model is: $\beta_1 = \beta_2 = 1$. This hypothesis is tested using the proposed Equation (3D) and panel data for trade and production in differentiated sectors.

An empirical issue here in this specification is treatment of zero-trade observations. A considerable number of observations of country pairs in certain years with no bilateral trade in differentiated sectors are included in both OECD and non-OECD samples. In my samples, although zero-trade observations are less than one percent of the total sample size in the OECD sample, such observations comprise more than 60% in the non-OECD sample. (The details of the samples and data are described in the next section.) For the log-linear specification these zero-valued observations bring the problem of undefined log of zero. In the first estimation, I omit these zero-trade observations and use only observations with positive trade volume whose log-transformed variables have relevant values. However, I also estimate the equations including the zero-trade observations, by replacing zero in the trade volume of these observations with a small
positive number for taking the logarithm. This method was also used by Debaere (2005) for his main log-linear specification.

Non-linear estimation for zero-trade observations: Poisson quasi-maximum likelihood estimator:

Although replacing zero with a small number is a simple and convenient way to incorporate the information of those zero-trade observations for estimation, it is, of course, not a perfect way to handle the zero-trade observations. The value of zero may have to be kept as zero rather than being mimicked by another number. Moreover, the “replacing method” makes coefficient estimates sensitive to the choice of the small number for replacement, since one can arbitrarily overweight or underweight observations with zero value for estimation by lowering or increasing the number for zero. We do not know what is the “right” number to put for zero.

Both Hummels & Levinsohn (1995) and Debaere (2005) had zero-trade observations in their non-OECD sample. Hummels & Levinsohn used a level specification for the absolute volume of trade (not relative to production) for the non-OECD sample, rather than a log-linear equation, so that they kept zero trade as the value of zero in their OLS regression. Debaere, in addition to the “replacement method” with a log-linear equation, used a level specification for the trade per GDP for his non-OECD sample and applied Tobit regression with fixed-effect dummies for estimation. However, their approaches have some drawbacks. By using a level specification that has a multiplicative form of the two variables of interest (world income share and size similarity), they had to give up estimating the impacts of the two variables separately; or

---

15 This number must be smaller than the minimum observed non-zero value in the sample data. The minimum value of the bilateral trade volume per production in the data is 9.4e-9, so I have chosen 10^-9 (1.0e-9) for the positive small number. So, the (log of) zero-trade observations are proxied by log(10^-9).

16 They have no zero-trade observations in a OECD sample.

17 Their log-linear specification for the OECD sample also put the volume of trade itself on the LHS, rather than trade per production, which is not exactly the same equation as the ones presented in this paper and also used by Debaere. See Appendix A for further details of their empirical approach.
to see the effects separately, Debaere had to abandon a structural equation strictly deriving from the monopolistic competition model.

In this paper, an alternative method is employed, which both keeps the structure of the theoretical model in the estimation equation and enables separate estimation of coefficients for the two variables. It is the fixed-effect Poisson quasi-likelihood estimator (FE-PQMLE). The Poisson regression is usually applied for count data, but it is also applicable to non-negative continuous variables. Hausman, Hall and Griliches (1984) developed the conditional FE-PQMLE method in the panel data context, which has been shown by Wooldridge (1999) to be consistent and robust to distributional assumptions, under the assumption that the conditional mean of the dependent variable is an exponential-class function of the linear combination of regressors (i.e., \( E[y|x] = \alpha \cdot \exp(x \beta) \)). This FE-PQMLE method seems to match well the empirical analysis here. That is,

\[
\begin{align*}
VT^D_{ijt} / X^D_{ijt} & = e^{\beta_1 \cdot \text{sim}^*_ijt \cdot \mu_{ijt} + \varepsilon_{ijt}} \\
\Leftrightarrow VT^D_{ijt} / X^D_{ijt} & = \exp[\beta_1 \cdot \log(\varepsilon_{ijt}) + \beta_2 \cdot \log(\text{sim}^*_ijt) + \mu_{ijt}] + \varepsilon_{ijt}
\end{align*}
\]

(4D)

The main difference from the structure of the log-linear specification described above is that now the stochastic error term \( \varepsilon_{ijt} \) is additive in the level equation, instead of multiplicative. The specification for the aggregate trade equation is similar although I do not present it here.

Tobit estimation is also possible, but as mentioned above, in order to maintain the structure of the model in an empirical equation and also obtain the two estimates separately, a log-linear specification should be used, which potentially brings the problem of log of zero. However, in the specific data used in this study, bilateral trade whose volume is below one

---

18 Silva and Tenreyo (2005) applies QPMLE to the gravity equation with cross-section data, and Westerlund and Wilhelmsson (2006) uses the fixed-effect QPMLE with panel data.
thousand dollars are omitted; i.e., the data are “censored” at $1000. Using this feature of these data, Tobit estimation can be applied with the following specification:

\[
\log(VT_{ijt}^D) = \log(X_{ijt}^D) + \beta_1 \cdot \log(e_{ijt}) + \beta_2 \cdot \log(sim_{ijt}^*) + \mu_{ijt} + \epsilon_{ijt}
\]  

(5D)

\[
\log(VT_{ijt}^D) = \log(VT_{ijt}^{D*}) \text{ if } VT_{ijt}^{D*} > 1000
\]

\[
\log(VT_{ijt}^D) = \log(1000) \text{ if } VT_{ijt}^{D*} \leq 1000
\]

\( VT_{ijt}^D \) is the observed or recorded bilateral trade volume in the data, while \( VT_{ijt}^{D*} \) is the underlying actual trade volume. Two things to be noted for this estimation method are that (i) two countries’ total production \( X_{ijt}^D \) is now on the right-hand side of the equation instead of in the denominator of the left-hand side, and the coefficient for this term must be restricted to be one for estimation in order to maintain the structural equation form; and (ii) all the zero-trade observations in the data must have their values of trade replaced with $1000. This Tobit estimation is used here as a robustness check.19

Data

To estimate Equations (3A) through (5D) presented in the previous section, it is necessary to obtain the data on production in the differentiated sectors, as well as data on trade and GDP, for both OECD and non-OECD countries. Eighty-nine countries are selected based on the availability of the data on manufacturing production in the source described below. These countries all have population above one million. These sample countries are grouped into OECD and non-OECD according to the actual OECD membership as of 1973.

19 It should be noted, however, that unconditional fixed-effect Tobit model will be generally biased due to the problem of incidental parameters (Hsiao, 2003; pp.48-9, 243).
The data on GDP are from the World Bank (2005). I employ GDP measured in current U.S. dollars converted from the local currency units using the current exchange rate.\textsuperscript{20} The value of world total GDP for each year, which is also necessary for calculating the world income shares of the two trading countries \((e_i)\), is also available in the data source. It should be noted that this world total GDP includes the GDP of the countries that are not included in the samples of this paper.

For the production data, I rely on the United Nation’s industrial statistics (UNIDO (2003)), which contains the annual data on production of countries for manufacturing industries classified according to the 3-digit ISIC (Revision 2) for the years of 1960-2000. I select the data on gross output in current U.S. dollars.

The data on bilateral trade flows are from Feenstra \textit{et al.} (1997), which is supplemented by Feenstra (2000). The dataset contains bilateral trade flows (in thousands of current U.S. dollars) between each pair of countries in each commodity classified based on the 4-digit SITC (Revision 2) for the years 1970-1997.

\textit{Industry/commodity classifications for the production data and trade data}

Since the production data and the trade data are based on different classification schemes, mapping one classification onto the other is necessary in order to match the two datasets based on a common classification. Twenty-eight industries represented in the production data are classified according to the 3-digit ISIC, while over a thousand commodities for the trade data are classified according to the 4-digit SITC. Thus the mapping requires condensing the 4-digit SITC classification into the 3-digit ISIC. I map the trade data onto the 3-digit ISIC using the

\textsuperscript{20}I also use GDP data from Penn World Table version 6.1, which are in a purchasing power-adjusted combatible unit, for additional estimations for the benchmark specification. Appendix B describes the estimation and its results.
concordance information sourced from OECD, which is available on Jon Haveman’s web page (http://www.eiit.org).\textsuperscript{21}

To separate the differentiated sectors from other (non-differentiated) sectors, I follow Rauch (1999), which classifies the 4-digit SITC commodities into three categories based on product differentiation: goods traded on an organized exchange (homogeneous goods), reference priced goods, and differentiated goods. Although the production data classified according to ISIC cannot be simply matched to Rauch’s three categories, there are six 3-digit ISIC manufacturing industries whose corresponding 4-digit SITC commodities are all classified as differentiated goods by Rauch: 322 (wearing apparel), 324 (footwear), 332 (furniture), 356 (plastic products), 361 (pottery, china, and earthenware), and 362 (glass and products).\textsuperscript{22} I therefore calculate the values of production and trade in the differentiated sectors for each country in each year by aggregating gross output and trade, respectively, in these six “pure” differentiated ISIC manufacturing industries. These six “differentiated” manufacturing industries comprise 4.6\% of the world total trade on average, with annual shares ranging from 3.0 to 6.4\% during the sample period of 1970-1997. These shares in the total trade of sample countries are 4.5\% on average, ranging from 2.9 to 5.9\%.\textsuperscript{23}

\textsuperscript{21} While the ISIC for the production data is based on industrial activities, the SITC for the trade data is based on commodity characteristics. Since the two classifications are based on different principles from each other, the mapping cannot necessarily be one-to-one.

\textsuperscript{22} The steps for selecting the differentiated sector are as follows. The original trade data based on the 4-digit SITC are divided into four subsets of the data, each of which is for one of Rauch’s three goods classification and a group of unclassified commodities. For each subset the data are re-classified into the 3-digit ISIC using the mapping information. Combining the four subsets of the data, the shares of Rauch’s three categories in the world total volume of trade are calculated for each 3-digit industry and each year. 3-digit ISIC industries in which for all the sample years (1970-97) the share of Rauch’s differentiated goods in the total industrial trade is 100\% are selected for the “differentiated sector.” The six 3-digit industries listed here satisfy this criterion.

\textsuperscript{23} Note that these differentiated sectors cover only manufacturing industries, so the share of these six industries in the world aggregate trade is not large.
The resulting panel data cover the 89 countries, 20 of which are in the OECD group and 69 are the non-OECD, for the period of 1970-1997. I cannot have the panel balanced for the entire 89 units and 28 time periods due to the lack of data for one or more variables for some countries in some years. However, I keep the panel unbalanced in order to retain as many observations in the sample as possible. The complete list of the countries and years included in the sample is shown in Table 1. As mentioned in the previous section, in both OECD and non-OECD samples zero bilateral trade values are observed for some country pairs in some years. In the OECD sample 28 observations out of the total 3,630 have zero trade in the differentiated sectors, and in the non-OECD sample 9,165 observations out of the total 14,565 are of zero trade.

Figures 1-A through 2-D show some plots of the data used for empirical analysis. These figures plot the volume of bilateral trade per production against size similarity (production structure-adjusted index for the differentiated sectors). All the variables are in log scale and mean-differenced, which correspond to the benchmark fixed-effect OLS specification. To avoid having thousands of dots in a cloud, one country pair is chosen from each group: UK-Japan from the OECD sample (Figures 1-A and 1-D) and Columbia-Philippines from the non-OECD sample (Figures 2-A and 2-D). Each point in a graph represents the data for a year for the designated country pair, and the graphs thus show variation across time. The vertical and horizontal lines indicate zeros, which are the means of the mean-differenced variables. The left panels (1-A, 2-A) plot the variables for aggregate trade, while the right panels (1-D, 2-D) plot those for differentiated-sector trade. The graphs show that the positive relationship between the volume of bilateral trade and size similarity becomes clearer by focusing on the differentiated sectors.
Empirical Results

*Fixed-effect OLS of the benchmark log-linear specification:*

First estimation is the fixed-effect OLS regression of the log-linear equations (3A) and (3D) using only non-zero trade observations. The results are presented in Table 2. The left columns show the estimation results of Equation (3A) for aggregate trade, and the right columns provide the results of Equation (3D) for trade in the differentiated sectors. The lower half of each column shows the results of the test of the hypothesis that the volume of trade is strictly proportional to both size similarity of trading countries and the world income share of the countries, as the underlying economic model suggests. The test results are shown in the probabilities (P-values) of non-rejection, or acceptance, of the following three hypotheses: (i) the estimate of the coefficient for the similarity index \( \beta_2 \) will be equal to one; (ii) the estimate of the coefficient for the world income share of trading countries \( \beta_1 \) will equal one; and (iii) the two coefficients will jointly be equal to one.\(^{24}\) The estimation of the aggregate trade equation (3A) indicates that the impact of size similarity on bilateral trade is significant for both OECD and non-OECD countries, but the coefficient estimate is considerably lower than one. That is, the volume of aggregate trade will increase as trading countries become more equal in size, but the increase in trade will be less than proportional. On the other hand, the estimation of the equation for trade in differentiated sectors shows that the relationship between (production structure-adjusted) size similarity and the volume of trade is not only significant but also close enough to strict proportionality to support the theoretical model. The hypothesis of the coefficient for the similarity index being one is accepted with fairly high probability for both OECD and non-OECD. These results seem to suggest that the monopolistic competition model explains

\(^{24}\) The P-values are based on F-tests of the Wald statistics.
international trade of differentiated products well. The results for aggregate trade seem also reasonable since although some products are traded under complete specialization, others are incompletely specialized in production, for which the sizes of countries does not predict bilateral trade volume by itself. The “noise” of these incompletely specialized sectors in the aggregate trade may make the link between trade volume and size similarity weaker.25

Table 3 shows the results of the fixed-effect OLS estimation of the same equations but including zero-trade observations by the “replacement method.” The estimation of the aggregate equation (3A) shows a similar result for OECD countries to the previous case, the result for non-OECD countries is different. The estimated coefficient for income similarity is very small (almost zero) and insignificant. This result agrees with Debaere’s (2005) finding in his benchmark regression using the same methodology. However, the estimation of the equation for the differentiated-sector trade provides a different picture. First, for OECD countries, the coefficient estimate for the adjusted similarity index not only maintains its significance but also is close to one, with which the hypothesis of being equal one is accepted with fairly high probability. Therefore, for the OECD sample the data seem to support the monopolistic competition model even with zero-trade observations. Secondly, for non-OECD countries, the estimated marginal impact of size similarity is now significant at the level of one percent, unlike the case of aggregate trade; that is, size similarity matters. However, the estimate is not as large as one, and thus the hypothesis of equaling one that is suggested by the model is rejected. On the other hand, considering that nearly 63% of the observations in the non-OECD sample are of zero-trade, this rejection of the coefficient equaling one may not have to be regarded as non-

25 Evenett and Keller (2002) derive a similar conclusion using one-way trade (export or import) gravity estimation. They obtain a coefficient estimate for income (GDP) in the gravity equation that is less than unity, which they interpret as evidence of incomplete production specialization in sectors.
favorable evidence for the monopolistic competition model, but rather be showing the limitation of the standard linear estimation method to incorporate zero-trade data.

In both estimations with and without zero-trade observations, however, the coefficient estimate for the other variable, the world income share of the two trading countries, is insignificant for OECD countries. This may be because the world income share of each country pair does not vary much across time for these countries, and the fixed-effect estimator uses the time-series variation of the observations for a country pair. For the non-OECD group, the estimated coefficient is significant but basically smaller than one, which the model suggests. This may present a consistent finding in the literature on the gravity equation (for example, McCallum (1995)), which is the so-called a “missing trade puzzle”; i.e., trade flows in reality are less than trade models predict (in this case, the trade flows in the data are smaller than those predicted with the income coefficient $\beta_1$ equal to one).

Non-linear estimations with zero-trade data:

The results of the estimation of the structural equation for the volume of trade by the standard linear method seem to suggest that the monopolistic competition model indeed nicely explains bilateral trade flows of differentiated products, and that size similarity, with the relative production structures of trading countries taken into account, predicts well the volume of trade in differentiated sectors adjusted by countries’ production size. In this subsection, alternative non-linear estimation methods, which have been described in the second section, are attempted to handle zero-trade observations without “replacing” with other non-zero value.

I first discuss the results, shown in Table 4, of Tobit estimation with country pair- and year-specific dummies. Note that, as described in the preceding section, the sectoral production
term, which is the denominator on the LHS in the benchmark log-linear equation, is now on the RHS as a regressor, but the coefficient for it is restricted to one. The dependent variable, log($VT$), is lower-censored at the value of log(1000), which is the minimum recorded volume of trade in the data. Focusing on the similarity index, the basic results agree with those in the benchmark estimation, and the main message remains here. That is, for the differentiated equation, the estimated coefficient is significant and close to one (the hypothesis of one-ness cannot be rejected at the 10% level of significance) for both country groups, though the estimate is much smaller or insignificant in the aggregate trade equation. By testing with disaggregated trade in differentiated sectors, the monopolistic competition model is now supported.26

Let us turn to the results of the FE-PQML estimation, which are shown in Table 5. In fact, the results do not agree with those in the previous two regressions. For OECD countries, the estimates do not differ between the aggregate and differentiated-sector specifications, in both of which the size of the coefficient for the similarity index is significantly smaller than one. For the non-OECD group, the result shows that the data may be fit to the model rather better for aggregate trade than for trade in the differentiated sectors. However, considering the inconsistency in the results between the benchmark estimation and FE-PQMLE, even for the OECD sample in which only 0.8% of the total observations (28 out of 3,630) are of zero-trade, this might suggest that zero trade should be treated in a trade-theoretically more sound way.

Conclusion

In this paper, I have re-examined an empirical test of the monopolistic competition model of international trade, which implies that the volume of trade between countries will increase as

---

26 The estimated impact of country pair’s income ($e_{ij}$) is not similar to that in the benchmark estimation. Especially, for the differentiated-sector equation, the coefficient estimate for this term is insignificant for non-OECD, and even significantly negative for OECD.
trading countries become more equal or similar in income size. The existing literature has tested this implication using aggregate trade, for which the empirical equation takes a very simple form but requires restrictive assumptions, such as perfect production specialization and the tradability of products in all sectors, to be derived from the model. However, understanding that international trade includes goods that are non-differentiated (or homogeneous) and produced by multiple countries (i.e., specialization is incomplete), which are “out of the scope” of the monopolistic competition model, using aggregate trade data would introduce noise and thus may reduce the accuracy of the test. This paper has proposed an alternative test of the monopolistic competition model that focuses on the trade of differentiated products, which the model exactly aims to describe. The tested equation has been carefully re-examined and re-derived for trade in differentiated sectors. The volume of trade of differentiated products can still be predicted by similarity between trading countries in their income size, but this income similarity must be adjusted by a factor indicating the extent to which the two countries are symmetric in production structure, or how similar the share of differentiated products in total economy is between the two countries. This alternative sectoral volume-of-trade equation has been empirically tested using data on trade and production at the industry level that are classified into product categories based on an available classification of traded goods. The results of the benchmark estimation, as well as of the Tobit estimation, show that, regardless of country groups, the change in the volume of bilateral trade in differentiated sectors seems to be proportional to the change in production structure-adjusted size similarity between two trading countries. This supports the implication of the monopolistic competition model.
References


Appendix A

This appendix is to review empirical approaches in the two preceding studies; Hummels and Levinsohn (1995) and Debaere (2005). Both studies use some versions of the empirical equation for the volume of aggregate bilateral trade, which are derived from the monopolistic competition model based on the two assumptions A1 and A2 described in the second section in the text of this paper. Please refer to that section for derivation of the equation.

Hummels & Levinsohn use the following specification:

for OECD: \[ \log(VT_{ijt}) = \beta \cdot \log(Y_{ijt} \cdot sim_{ij}) + \eta_{ijt} + \varepsilon_{ijt} \] (*)

for non-OECD: \[ VT_{ijt} = \beta \cdot (Y_{ijt} \cdot sim_{ij}) + \mu_{ij} + \varepsilon_{ijt} \] (**)

where \( sim_{ij} = \left[ 1 - \left( Y_i / Y_j \right)^2 - \left( Y_j / Y_i \right)^2 \right] \).

Some points should be noted, in terms of differences from the specifications introduced in this paper. First, they use the (log of) the absolute volume of trade as the dependent variable, rather than the volume of trade per income (GDP) as in Equation (3A) in the text. A country pair’s income is put on the RHS as the product term with the income similarity index in their specification. Therefore, they estimate one coefficient for the product term of income and the similarity index\(^1\); and this is the second point. Thirdly, they assume, as Helpman (1987) did, that the world income share of a pair of two countries does not change (at least much) across years, so the term for world income share, \( e_{ij} \), is merged into the country pair-specific dummies \( \eta_{ij} \) in their equation for OECD. (In the equation for non-OECD, the time-invariant income share term will be absorbed by the slope coefficient \( \beta \).) They estimate the log-linear equation (*) for OECD, but use the level equation (**) for non-OECD since they have zero-trade observations in their

\(^1\) Imposing the restriction that the coefficients for the two elements are the same is not a problem by itself, since the model suggests that the both elements are strictly proportional to the volume of trade. However, Debaere claims an econometric problem in this approach, as described later.
non-OECD data (using level variables enables them to avoid the problem of the log of zero).
They apply pooled OLS, random-effect OLS, and fixed-effect OLS regressions to both equations,
using balanced panel data on bilateral (aggregate) trade for 14 OECD countries for 1962-1983
for (*) and the data for 14 non-OECD countries for 1962-1977 for (**). In any estimation, they
obtain an estimate for $\beta$ that is positive and significant for both country groups.

Debaere starts with a claim that their significant coefficient estimate, which is counter to
expectation for non-OECD countries, may be driven by a high correlation between the volume of
trade and the size of two countries, rather than their similarity. He argues that although the
similarity index may not at all relate to, and thus be totally independent of, the volume of trade,
the coefficient estimate for the product of income size and the similarity index would be
significant if the other part of the term, the size of total income, is highly correlated to the
volume of trade. This is indeed highly likely since in general the absolute volume of trade of
large countries is greater than that of small countries.\(^{ii}\) Therefore, he proposes an empirical
equation that uses the volume of bilateral trade adjusted by the size of income of a country pair,
which is the same as Equation (3A) in this paper. His benchmark estimation is fixed-effect OLS
for Equation (3A), and he also includes year-specific dummies. For zero-trade observations in
his non-OECD sample, he applies the “replacement method” mentioned in the text of this
paper.\(^{iii}\)

In addition to Equation (3A), he uses the following two level specifications for
estimation:

\(^{ii}\) However, it should be noted that Hummels and Levinsohn seem to have noticed this issue by
themselves. In fact, as they mentioned in a footnote (Hummels and Levinsohn, 1995; pp.808, footnote 14),
they also estimated an equation separating the term for income size ($Y_{ij}$) from the similarity index, from
which they concluded that the similarity index is still significant.

\(^{iii}\) He does not mention what is the number he uses for replacement of zeros.
\[ VT_{ijt} / Y_{ijt} = \beta \cdot (e_{ijt} \cdot sim_{ijt}) + \mu_{ij} + \epsilon_{ijt} \]

\[ VT_{ijt} / Y_{ijt} = \beta_1 \cdot e_{ijt} + \beta_2 \cdot sim_{ijt} + \mu_{ij} + \epsilon_{ijt} \]

Both level equations are estimated by fixed-effect OLS for OECD countries and by Tobit with fixed-effect dummies for non-OECD countries.

Debaere constructs a balanced panel from the data on bilateral (aggregate) trade and GDP for the same 14 countries and 12 countries for the OECD and non-OECD groups, respectively, as Hummels & Levinsohn choose. (2 countries out of Hummels & Levinsohn’s non-OECD sample are dropped in Debaere, due to data availability.) The covered period is from 1970 to 1989. The results of his benchmark OLS with the log-linear specification leads him to conclude that the monopolistic competition model is supported for OECD countries but not for non-OECD countries, as he expects (and Hummels and Levinsohn initially expected).iv

Finally, the list of the OECD and non-OECD countriesv that are included in both Hummels & Levinsohn and Debaere is as follows. These 14 OECD countries were initially selected by Helpman (1987), which the two papers follow. Two non-OECD countries marked by * are excluded in Debaere’s study due to data availability.

<table>
<thead>
<tr>
<th>OECD countries (14)</th>
<th>Non-OECD countries (14*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>Brazil</td>
</tr>
<tr>
<td>Belgium</td>
<td>Cameroon</td>
</tr>
<tr>
<td>Canada</td>
<td>Netherlands</td>
</tr>
<tr>
<td>Germany</td>
<td>Columbia</td>
</tr>
<tr>
<td>Denmark</td>
<td>Congo*</td>
</tr>
<tr>
<td>France</td>
<td>Cote d’Ivoire*</td>
</tr>
<tr>
<td>Ireland</td>
<td>Greece</td>
</tr>
<tr>
<td></td>
<td>South Korea</td>
</tr>
<tr>
<td></td>
<td>United Kingdom</td>
</tr>
<tr>
<td></td>
<td>United States</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

iv Although Debaere insists that the results of other estimations support this conclusion, for me those other results look quite mixed.

v In fact, these 14 “non-OECD” countries include two countries that have been original OECD members since 1961: Greece and Norway.
Appendix B

This appendix is to show the results of the benchmark log-linear estimation using an alternative measure of income size, which is the purchasing power-adjusted GDP expressed in world compatible units (PPP-GDP). I employ GDP in current US dollars as the primary measure of the income sizes of countries since both trade and production data are measured in this unit. However, one might worry about the possibility that the current-dollar GDP includes non-traded sectors, which may be unequally valued across countries, and it may thus be inflated or deflated as the measure of a country’s purchasing power of tradable products. PPP-GDP might better measure the incomes of countries in such a case. Therefore, in this appendix I use PPP-GDP for the measure of the income sizes of countries to estimate the benchmark equation. The data on PPP-GDP are from Penn World Table version 6.1.\textsuperscript{vi} It should be noted, however, that I keep other variables (trade and production) still measured in the current US dollars. Especially, for the specification for aggregate trade, GDP that expresses a nation’s aggregate production (for example, the denominator on the LHS) is kept in current US dollars although GDP as income is now in PPP-adjusted value.

The results of estimation without zero-trade observations are shown in Table B-1, and those including zero-trade observations are in Table B-2. One remark is that, unlike the results with current-dollar GDP, the coefficient estimate for the world income share ($e_{ijt}$) is now large and exceeds one in most cases.$\textsuperscript{vii}$ However, the two measures of $e_{ijt}$ are highly positively correlated in both country groups (the correlation coefficient of the two variables (in logs) is 0.99 for the OECD sample and 0.96 for non-OECD). A possible reason is that the variance of the PPP

\textsuperscript{vi} PPP-GDP data in the Penn World Table are available for fewer countries than the countries for which current-dollar GDP data are available in World Development Indicators. Because of this, the “world GDP” in PPP units, which is for calculating each country pair’s income share, includes fewer counties than does the world GDP in current dollars.

\textsuperscript{vii} This is similar to what is seen in the estimation in Debaere (2005).
measure is smaller than the current-dollar measure, which means that change in the PPP measure corresponding to the same change in the volume of trade per production (LHS) is smaller than that in the current-dollar measure, so that the estimated coefficient becomes larger. In addition, by using PPP-GDP the size of the coefficient estimate for the similarity index also becomes larger than that with the current-dollar measure,\textsuperscript{viii} while it does not seem to be the case for the non-OECD group.

The main message, however, seems not different from the estimation with the current-dollar measure, especially when zero-trade observations are excluded: By focusing trade in differentiated sectors, the estimate for the similarity index is likely to be one, which supports the monopolistic competition model.

\textsuperscript{viii} This is also seen in Debaere’s result.
Table 1: List of Countries & Periods included in the sample groups

<table>
<thead>
<tr>
<th>OECD (20 Countries)*</th>
<th>Non-OECD (69 Countries)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Country</strong></td>
<td><strong>Periods</strong></td>
</tr>
<tr>
<td>Austria</td>
<td>1970-97</td>
</tr>
<tr>
<td>Germany (West)</td>
<td>1971-84</td>
</tr>
<tr>
<td>Denmark</td>
<td>1970-91</td>
</tr>
<tr>
<td>Italy</td>
<td>1970-91</td>
</tr>
<tr>
<td>Turkey</td>
<td>1970-95</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* OECD/non-OECD grouping is according to the OECD membership as of 1973.
Table 2: Estimation Results using **Positive-Trade Observations Only**
(Income measure: GDP in current US dollars)

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>Aggregate specification Eq.(3A)</th>
<th>Differentiated-sector specification Eq.(3D)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OECD</td>
<td>Non-OECD</td>
</tr>
<tr>
<td>log(similarity)</td>
<td>0.428*** (0.133)</td>
<td>0.561*** (0.169)</td>
</tr>
<tr>
<td></td>
<td>(s.e.)</td>
<td>(s.e.)</td>
</tr>
<tr>
<td>log(world income share)</td>
<td>-0.162 (0.164)</td>
<td>0.611*** (0.150)</td>
</tr>
<tr>
<td></td>
<td>(s.e.)</td>
<td>(s.e.)</td>
</tr>
<tr>
<td>R-square</td>
<td>0.02 (0.15)</td>
<td>0.01 (0.08)</td>
</tr>
<tr>
<td>(within R-square)</td>
<td>(0.15)</td>
<td>(0.08)</td>
</tr>
<tr>
<td># observations</td>
<td>3,602</td>
<td>5,400</td>
</tr>
</tbody>
</table>

(Alternative Hypotheses Tests)

<table>
<thead>
<tr>
<th>P-values of F-tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>coef. for similarity = 1</td>
</tr>
<tr>
<td>coef. for income share = 1</td>
</tr>
<tr>
<td>coef. for similarity = coef. for i-share = 1</td>
</tr>
</tbody>
</table>

Notes: All variables are in logarithms. OECD sample includes 20 countries and non-OECD sample includes 69 countries, both for years 1970-97. The samples exclude observations with zero trade. Country pair-specific and year-specific dummies are included in the regressions. Standard errors are clustered by country pair. ***, **, * indicate the significance levels at 1%, 5%, and 10%, respectively. The lower parts of the table show the results of the Wald test for the hypotheses of each coefficient equaling one and the two coefficients jointly equaling one, in p-values.

Table 3: Estimation Results using **All Observations**
(Income measure: GDP in current US dollars)

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>Aggregate specification Eq.(3A)</th>
<th>Differentiated-sector specification Eq.(3D)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OECD</td>
<td>Non-OECD</td>
</tr>
<tr>
<td>log(similarity)</td>
<td>0.397*** (0.138)</td>
<td>0.062 (0.242)</td>
</tr>
<tr>
<td></td>
<td>(s.e.)</td>
<td>(s.e.)</td>
</tr>
<tr>
<td>log(world income share)</td>
<td>-0.155 (0.169)</td>
<td>0.105** (0.254)</td>
</tr>
<tr>
<td></td>
<td>(s.e.)</td>
<td>(s.e.)</td>
</tr>
<tr>
<td>R-square</td>
<td>0.01 (0.09)</td>
<td>0.07 (0.04)</td>
</tr>
<tr>
<td>(within R-square)</td>
<td>(0.09)</td>
<td>(0.04)</td>
</tr>
<tr>
<td># observations</td>
<td>3,630</td>
<td>14,565</td>
</tr>
</tbody>
</table>

(Alternative Hypotheses Tests)

<table>
<thead>
<tr>
<th>P-values of F-tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>coef. for similarity = 1</td>
</tr>
<tr>
<td>coef. for income share = 1</td>
</tr>
<tr>
<td>log(sim) = log(i-share)</td>
</tr>
<tr>
<td>coef. for similarity = coef. for i-share = 1</td>
</tr>
</tbody>
</table>

Notes: All variables are in logarithms. OECD sample includes 20 countries and non-OECD sample includes 69 countries, both for years 1970-97. The samples include observations with zero trade volume. Country pair-specific and year-specific dummies are included in the regressions. Standard errors are clustered by country pair. ***, **, * indicate the significance levels at 1%, 5%, and 10%, respectively. The lower parts of the table show the results of the Wald test for the hypotheses of each coefficient equaling one and the two coefficients jointly equaling one, in p-values.
### Table 4: Results of Tobit Estimation: using All Observations  
(Income measure: GDP in current US dollars)

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>Aggregate specification</th>
<th>Differentiated-sector specification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OECD</td>
<td>Non-OECD</td>
</tr>
<tr>
<td>similarity index</td>
<td>0.447*** (0.138)</td>
<td>0.110 (0.143)</td>
</tr>
<tr>
<td>world income share</td>
<td>0.241*** (0.086)</td>
<td>0.386** (0.159)</td>
</tr>
<tr>
<td>Pseudo R-square</td>
<td>0.83</td>
<td>0.31</td>
</tr>
<tr>
<td># observations</td>
<td>3,630</td>
<td>14,565</td>
</tr>
</tbody>
</table>

**Alternative Hypotheses Tests**

<table>
<thead>
<tr>
<th>P-values of F-tests</th>
<th>coef. for similarity = 1</th>
<th>coef. for income share = 1</th>
<th>coef. for similarity = coef. for i-share = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>OECD</td>
<td>0.000</td>
<td>0.000</td>
<td>0.122</td>
</tr>
<tr>
<td>Non-OECD</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

**Notes:** All variables are in logarithms. Log of industrial production (or aggregate production = GDP) is included as a regressor, but the coefficient is restricted to one. OECD sample includes 20 countries and non-OECD sample includes 69 countries, both for years 1970-97. All observations are included, and left-censored at the value of ln($1000). Country pair-specific and year-specific dummies are included in the regressions. ***, **, * indicate the significance levels at 1%, 5%, and 10%, respectively. The lower parts of table shows the results of the Wald test for the hypotheses of each coefficient equaling one and the two coefficients jointly equaling one, in p-values.

### Table 5: Results of Poisson QML Estimation (PQMLE): using All Observations  
(Income measure: GDP in current US dollars)

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>Aggregate specification</th>
<th>Differentiated-sector specification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OECD</td>
<td>Non-OECD</td>
</tr>
<tr>
<td>log(similarity)</td>
<td>0.628*** (0.109)</td>
<td>0.862*** (0.296)</td>
</tr>
<tr>
<td>log(world income share)</td>
<td>0.102 (0.109)</td>
<td>0.652*** (0.131)</td>
</tr>
<tr>
<td># observations</td>
<td>3,630</td>
<td>12,329</td>
</tr>
</tbody>
</table>

**Alternative Hypotheses Tests**

<table>
<thead>
<tr>
<th>P-values of F-tests</th>
<th>coef. for similarity = 1</th>
<th>coef. for income share = 1</th>
<th>coef. for similarity = coef. for i-share = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>OECD</td>
<td>0.003</td>
<td>0.641</td>
<td>0.039</td>
</tr>
<tr>
<td>Non-OECD</td>
<td>0.000</td>
<td>0.008</td>
<td>0.007</td>
</tr>
</tbody>
</table>

**Notes:** The LHS variable is in level, while all the RHS variables are in logarithm. OECD sample includes 20 countries and non-OECD sample includes 69 countries, both for years 1970-97. The samples includes observations with zero trade. The conditional fixed-effect PQML estimation follows Hausman et al. (1984), including time-specific dummies. The estimation does not use observations for groups with only one-period observation or no positive observations in any period. Bootstrapped standard errors are used. ***, **, * indicate the significance levels at 1%, 5%, and 10%, respectively. The lower parts of table shows the results of the Wald test for the hypotheses of each coefficient equaling one and the two coefficients jointly equaling one, in p-values.
Figures 1: Volume of Bilateral Trade vs. Similarity Index:
for OECD pair: **UK and Japan** (in log scales; mean-differenced)

1-A: Aggregate  
1-D: Differentiated

![Graph 1-A: Aggregate](image1-a.png)  
![Graph 1-D: Differentiated](image1-d.png)

Figures 2: Volume of Bilateral Trade vs. Similarity Index:
for non-OECD pair: **Columbia and Philippines** (in log scales; mean-differenced)

2-A: Aggregate  
2-D: Differentiated

![Graph 2-A: Aggregate](image2-a.png)  
![Graph 2-D: Differentiated](image2-d.png)

* Size similarity index (for 1-A and 2-A) or production structure-adjusted similarity index (for 1-D and 2-D) is on the horizontal axis, and the volume of trade per production on the vertical. All the variables are mean-differenced (for fixed-effect OLS). The vertical and horizontal lines indicate zero.
### Table B-1: Estimation Results using Positive-Trade Observations Only

(Income measure: GDP in PPP international dollars)

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>Aggregate specification Eq.(3A)</th>
<th>Differentiated-sector specification Eq.(3D)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OECD Non-OECD</td>
<td>OECD Non-OECD</td>
</tr>
<tr>
<td>log(similarity)</td>
<td>1.20*** (0.231)</td>
<td>1.46*** (0.376)</td>
</tr>
<tr>
<td></td>
<td>0.650*** (0.184)</td>
<td>0.793*** (0.253)</td>
</tr>
<tr>
<td>log(world income share)</td>
<td>1.01*** (0.370)</td>
<td>1.25 (0.950)</td>
</tr>
<tr>
<td></td>
<td>1.55*** (0.222)</td>
<td>0.952** (0.382)</td>
</tr>
<tr>
<td>R-square</td>
<td>0.15 (0.18)</td>
<td>0.09 (0.09)</td>
</tr>
<tr>
<td>(within R-square)</td>
<td>0.00 (0.13)</td>
<td>0.00 (0.09)</td>
</tr>
<tr>
<td># observations</td>
<td>3,602 4,896</td>
<td>3,602 4,896</td>
</tr>
</tbody>
</table>

(Alternative Hypotheses Tests)

P-values of F-tests

- coef. for similarity = 1: 0.395 (0.370) 0.226 (0.293)
- coef. for income share = 1: 0.971 (0.456) 0.792 (0.494)
- log(sim) = log(i-share): 0.000 (0.000)

Notes: All variables are in logarithms. OECD sample includes 20 countries and non-OECD sample includes 69 countries, both for years 1970-97. The samples exclude observations with zero trade. Country pair-specific and year-specific dummies are included in the regressions. Standard errors are clustered by country pair. ***, **, * indicate the significance levels at 1%, 5%, and 10%, respectively. The lower parts of table shows the results of the Wald test for the hypotheses of each coefficient equaling one and the two coefficients jointly equaling one, in p-values.

### Table B-2: Estimation Results using All Observations

(Income measure: GDP in PPP international dollars)

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>Aggregate specification Eq.(3A)</th>
<th>Differentiated-sector specification Eq.(3D)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OECD Non-OECD</td>
<td>OECD Non-OECD</td>
</tr>
<tr>
<td>log(similarity)</td>
<td>1.15*** (0.240)</td>
<td>1.48*** (0.467)</td>
</tr>
<tr>
<td></td>
<td>0.883** (0.353)</td>
<td>0.792*** (0.293)</td>
</tr>
<tr>
<td>log(world income share)</td>
<td>1.13** (0.456)</td>
<td>2.28* (1.19)</td>
</tr>
<tr>
<td></td>
<td>2.18*** (0.438)</td>
<td>3.19*** (0.494)</td>
</tr>
<tr>
<td>R-square</td>
<td>0.13 (0.12)</td>
<td>0.03 (0.05)</td>
</tr>
<tr>
<td>(within R-square)</td>
<td>0.12 (0.05)</td>
<td>0.09 (0.11)</td>
</tr>
<tr>
<td># observations</td>
<td>3,630 12,656</td>
<td>3,630 12,656</td>
</tr>
</tbody>
</table>

(Alternative Hypotheses Tests)

P-values of F-tests

- coef. for similarity = 1: 0.538 (0.724) 0.310 (0.285)
- coef. for income share = 1: 0.774 (0.007) 0.285 (0.000)
- log(sim) = log(i-share): 0.000 (0.000)
- coef. for similarity = coef. for i-share = 1: 0.819 (0.005) 0.387 (0.000)

Notes: All variables are in logarithms. OECD sample includes 20 countries and non-OECD sample includes 69 countries, both for years 1970-97. The samples include observations with zero trade. Country pair-specific and year-specific dummies are included in the regressions. Standard errors are clustered by country pair. ***, **, * indicate the significance levels at 1%, 5%, and 10%, respectively. The lower parts of table shows the results of the Wald test for the hypotheses of each coefficient equaling one and the two coefficients jointly equaling one, in p-values.