A PRODUCT-QUALITY VIEW OF THE LINDER HYPOTHESIS

Juan Carlos Hallak*

Abstract—The Linder hypothesis has attracted substantial empirical research over decades. However, the evidence has failed to provide consistent support for it. This paper explains the failure. Building a theoretical framework in which, as in Linder’s theory, product quality plays the central role, I show that the Linder hypothesis is formally derived but holds only when formulated as a sector-level prediction. This prediction is then estimated using a sample of 64 countries in 1995. The results support the sectoral Linder hypothesis: controlling for the effect of intersectoral determinants of trade, countries of similar income per capita trade more intensely with one another.

I. Introduction

There is ample evidence that the quality of goods that countries produce and consume varies systematically with their income levels. On the production side, export prices are strongly correlated with countries’ income per capita, suggesting a positive relationship between per capita income and quality production (Schott, 2004; Hummels & Klenow, 2005; Hallak, 2006a). On the consumption side, quality demand is strongly correlated with household income (Bils & Klenow, 2001), suggesting a positive relationship, at the country level, between per capita income and quality consumption. This last relationship is also found in studies based on firm-level and country-level international trade data.1 The evidence of a systematic relationship between income per capita and product quality on the supply and demand sides points to a potentially important role of product quality as a determinant of bilateral trade patterns. While workhorse theories of international trade neglect this role, growing interest in product quality has spurred a substantial amount of theoretical research on the interaction between product quality and international trade.2

Linder (1961) introduced the idea of quality as an important determinant of the direction of trade long before empirical evidence of systematic cross-country variation in quality production and consumption was available. On the demand side of this theory, countries with high income per capita spend a larger fraction of their income on high-quality goods. On the supply side, countries develop a comparative advantage in the goods that are in high domestic demand. Thus, high- (low-) income countries develop a comparative advantage in the production of high- (low-) quality goods because those are the goods that are highly demanded in their domestic markets. Demand and supply are then combined to argue that the overlap of production and consumption patterns among countries of similar income per capita should induce them to trade more intensely with one another, a prediction commonly known as the Linder hypothesis. This hypothesis has attracted the attention of scholars for decades. However, despite the strong empirical evidence in support of both sides of Linder’s theory, empirical support for the hypothesis has not been robust across studies.

This paper provides an explanation for why consistent empirical support for the Linder hypothesis has been elusive so far, while simultaneously going back to Linder’s original emphasis on product quality as the driving force of the hypothesis. The paper builds a theoretical framework that captures the role of quality that Linder emphasized and delivers his conjecture as a formal prediction. In contrast to its standard formulation as a prediction for bilateral trade flows at the aggregate level, the Linder hypothesis is derived here as a prediction for bilateral flows at the sector level. Furthermore, aggregation across sectors is shown to introduce a systematic bias. Consistent with the theoretical predictions, the empirical results support the “sectoral Linder hypothesis” but do not support the Linder hypothesis as a prediction for aggregate trade.

The theoretical framework of the paper models countries as differing in their valuation of quality. In particular, a parameter governs the variation in countries’ expenditure shares across varieties of different quality. On the supply side, since several plausible theories relate quality production to per capita income, their common implications are captured by simply assuming a systematic relation between quality supply and per capita income. The interaction of demand for and supply of quality results in a gravity-type specification in which sectoral bilateral trade is a function of...
exporter and importer fixed effects, proxies for bilateral trade costs, and a “Linder term” measuring per capita income dissimilarity between pairs of countries. When both quality demand and quality supply increase with income per capita, as Linder postulated, the model predicts a negative coefficient on the Linder term. This prediction is the formal characterization of the sectoral Linder hypothesis.

I estimate the sectoral Linder hypothesis using a sample of bilateral trade flows in differentiated goods among 64 countries in 1995. Trade flows in the sample are classified at the three-digit SITC level. The empirical results support the theoretical prediction: countries with similar income per capita tend to trade more intensely with one another.

The sectoral Linder hypothesis can isolate the within-sector impact of product quality (for example, high-quality furniture versus low-quality furniture) because it controls for the between-sector impact of intersectoral determinants of comparative advantage (for example, furniture versus cars). In contrast, standard tests of the Linder hypothesis confound the impact of these two forces by using aggregate trade data. This induces a systematic upward bias in the estimated coefficient of the Linder term since patterns of sectoral specialization in manufacturing are strongly correlated with income per capita. The bias increases with the level of aggregation. In this paper, when the Linder hypothesis is estimated, aggregating all trade flows in differentiated products, the bias is sufficiently strong to reverse the estimated sign of the Linder effect.

The presence of the aggregation bias explains why a vast empirical literature testing the Linder hypothesis has been unable to find consistent support of its empirical validity. Linder’s work and the early studies that followed it initiated this literature, finding strong support for the hypothesis. However, these studies were flawed because they failed to control for geographical distance: countries of similar income per capita tend to be closer, and closer countries tend to trade more intensely with each other. Subsequent work has controlled for distance and other determinants of bilateral trade barriers, typically in a gravity-equation-type framework using aggregate trade data. This work finds only mixed support for the Linder hypothesis. A final set of studies tests the hypothesis using bilateral shares of intraindustry trade. Intraindustry shares also confound the effect of quality with the effect of intersectoral determinants of trade. In this case, however, both effects presumably operate in the same direction, inducing more intraindustry trade between pairs of countries with similar income per capita.

Several theoretical studies capture the key insights of Linder’s theory by modeling high-income countries as having a comparative advantage in the production of high-quality goods and consuming those goods in larger proportions. However, the models do not yield the Linder hypothesis as a theoretical prediction since they consider a world economy with only two countries, whereas the Linder hypothesis, a statement about differences in trade volumes between country pairs, involves at least four. A related theoretical literature also focuses on the aggregate Linder hypothesis but does not rely on product quality as its driving force. This literature emphasizes the role of intersectoral nonhomotheticities in demand, which can generate higher trade intensity between countries of similar income per capita if per capita income and comparative advantage in high-elasticity sectors are positively related. While intersectoral nonhomotheticities can potentially predict the “aggregate” Linder hypothesis, this prediction is not borne out in the data.

By developing a framework that captures Linder’s intuition about the link between product quality and the direction of trade, this paper solves the apparent contradiction between the appeal and the empirical relevance of the economic mechanisms Linder highlighted and the failure to find consistent empirical support for the Linder hypothesis. The data bear out the premises of Linder’s theory: countries of similar income have similar production and consumption patterns—they produce and consume goods of similar quality. But the hypothesized corollary that countries of similar income trade more with each other holds only at the sector level, where intersectoral determinants of trade can be controlled for. At the aggregate level, the Linder hypothesis does not follow from the premises that are supposed to imply it.

Although Linder used aggregate trade data to perform a first-pass test of his theory, he explicitly argued that the connection between income and quality demand should operate within rather than between sectors: “Qualitative product differences are not well brought out in empirical studies of consumer behavior along the lines first followed by Engel. The qualitative factor is submerged by taking broad groups of goods such as ‘food’ or ‘clothing’ ” (Linder, 1961, p. 95). His own work and the work of subsequent empirical researchers did not follow this early lead.

Section II follows the standard approach for estimating the Linder hypothesis (using aggregate trade data) and finds

3 Deardorff (1984) and Leamer and Levinsohn (1995) survey most of the literature. Thursby and Thursby (1987) find strong support for the Linder hypothesis, but their results are not compelling since their study is based on a sample of only seventeen mostly OECD countries and thus does not exploit the most informative source of cross-country variation in per capita income levels. More important, the statistical significance of the estimates is likely to be severely overestimated since they use panel data without including fixed effects and assume no autocorrelation in the error term. McPherson, Redfearn, and Tieslau (2001) find support for the hypothesis when they use a tobit estimation procedure but do not find support for it when they use OLS.

4 Most notably Bergstrand (1990), using fourteen developed countries and a subset of manufacturing industries (SITC 71, 72, and 73), finds that the share of intraindustry trade grows with income similarity.


7 This type of nonhomotheticities could be empirically relevant yet be overshadowed by the (opposite) effect of supply-side intersectoral determinants of trade.
that it is strongly rejected. Section III develops the theoretical framework that delivers the sectoral Linder hypothesis. Section IV shows the empirical evidence supporting it. Section V provides theoretical and empirical evidence on the existence of a systematic aggregation bias. Section VI concludes.

II. The “Aggregate” Linder Hypothesis

The standard formulation of the Linder hypothesis states that countries with similar income per capita should trade more intensely with one another. This formulation is a prediction about bilateral trade flows aggregated across sectors, typically tested in a gravity-equation framework. To obtain comparable results to those in the literature, in this section I test this formulation of the hypothesis and show that it is not empirically supported.

Before presenting estimation results, this section describes some basic facts about the direction of world trade. Table 1 reports the volume of trade in 1995 between two groups of countries, rich and poor, divided according to their income per capita. The cut-off value used in the table is a (PPP-adjusted) GDP per capita of US$12,000.9 The 107 countries in the first row have a population above 1 million; of these, 26 are classified as rich and 81 as poor. Total trade (both exports and imports) among these 107 countries in 1995 is $5,049 billion, 97% of total world trade. High-income countries account for 81.1% of total exports ($4,097 billion) and 78.8% of total imports ($3,980 billion).

A preliminary assessment of the aggregate Linder hypothesis can be made by comparing import and export shares between groups, presented in the second set of columns. Rich countries buy 81.1% of their imports from other rich countries and the remaining 18.9% from poor countries. Poor countries buy 81.5% of their imports from rich countries and 18.5% from other poor countries. Apparently at odds with the predictions of the Linder hypothesis, rich (poor) countries as a group do not import relatively more from other rich (poor) countries. The last set of columns yields the same message viewed from the exporter side. Rich countries export (slightly) relatively more to poor countries, while poor countries export relatively more to rich countries.

The remainder of the paper focuses on a sample of 64 countries—those with population above 3 million and imports of differentiated goods above US$2 billion.9 This sample includes 26 rich countries and 38 poor countries.10 The information in the second row of Table 1 is analogous to the first row but is based only on trade flows among these 64 countries. As is the case in the larger sample, countries do not trade more with countries of similar per capita income. For example, while poor countries’ share of rich countries’ imports is 16.8%, their share of poor countries’ imports is 15.8%.11

Although informative about the structure of world trade, Table 1 cannot be used to assess the empirical validity of the Linder hypothesis. Countries of similar income per capita tend to be geographically close. Since trade costs are not controlled for, more intense trade between them may be due to their lower bilateral transport costs rather than to income similarity. The empirical literature has long acknowledged this concern and has addressed it by adopting a gravity-equation framework with bilateral trade expressed as a function of trading partners’ income and proxies for their bilateral trade costs, which is then augmented to include a “Linder term,” a measure of income dissimilarity between pairs of countries. A typical specification used for estimation is

$$\ln(imports_{od}) = \beta_0 + \beta_1 \ln(Y_o) + \beta_2 \ln(Y_d) + \beta_3 \ln(Dist_{od}) + \beta_4 I_{od} + \beta_5 Linder_{od} + \epsilon_{od},$$  (1)

9 Section IV describes data sources, classification of sectors into goods categories, and selection criteria for countries and sectors.
11 The choice of an alternative cut-off value of US$15,000, in which case Greece, Portugal, Taiwan, and South Korea switch to the group of poor countries, yields the same qualitative result.
where \( o \) denotes the exporter (origin), \( d \) denotes the importer (destination), \( Y_o \) (\( Y_d \)) is the GDP of country \( o \) (\( d \)), \( \text{Dist}_{od} \) is the geographical distance between countries \( o \) and \( d \), \( \text{I}_{od} \) is a vector of dichotomous proxy variables for trade costs, and \( \text{Linder}_{od} \) is the Linder term. Several Linder terms have been proposed in the literature. The Linder term that will be used as baseline in this study is \( \text{Linder}_{od} = (\ln y_o - \ln y_d)^2 \), where \( y_o \) (\( y_d \)) is the income per capita of country \( o \) (\( d \)). The Linder term is larger the more dissimilar are the two countries’ incomes. Therefore, the prediction of the Linder hypothesis is that \( \beta_5 < 0 \). In its simplest form, this equation is estimated using OLS.

I follow the empirical literature by estimating (a more general version of) equation (1) with OLS using the sample of 64 countries. Since theoretical derivations of the gravity equation (Eaton & Kortum, 2002; Anderson & Van Wincoop, 2003) indicate that its standard formulation should be augmented to include exporter and importer fixed effects, I estimate equation (1) including those terms, which then embed the effect of \( y_o \) and \( y_d \). The vector \( \text{I}_{od} \) includes dummies indicating whether a country pair shares a common border, a common language, a preferential trade agreement (PTA), a colony-colonizer relationship, or a common colonizer.

The results, presented in the first column of Table 2, strongly reject the Linder hypothesis. Instead of a negative coefficient on the Linder term, the estimated coefficient is not only positive but also significant at the 10% level. Countries with dissimilar income per capita appear to trade more with one another. The estimated coefficients on the other variables have the predicted sign in all cases. Lower trade costs, induced by a shorter distance or by sharing a border, common language, PTA, colonizer-colony relationship, or common colonizer, are associated with larger volumes of bilateral trade. The remaining columns of the table report results from estimating the same specification, including alternative Linder terms used in the literature: \( |y_o - y_d|, \ln |y_o - y_d|, \ln |y_o - y_d|, \text{and } |\ln y_o - \ln y_d| \). In contrast to the prediction of the Linder hypothesis, in all cases, the estimated coefficient is positive, and in all but one case, it is significantly positive.\(^{12} \)

The estimations of this section have the sole purpose of reproducing standard empirical tests of the Linder hypothesis. I show next that the failure to find support for the hypothesis cannot be attributed to the absence of the economic mechanisms described by Linder but rather to a misspecified empirical benchmark to test their impact on international trade.

### III. The “Sectoral” Linder Hypothesis: A Theoretical Framework Based on Quality

Product quality plays a central role in Linder’s theory. This section describes a theoretical framework that captures the essential components of that theory—a systematic relationship between countries’ per capita income and their supply of and demand for quality—and yields a formal derivation of the Linder hypothesis. In contrast to its standard formulation, the Linder hypothesis is shown to be valid only when formulated at the sector level, after controlling for intersectoral determinants of trade.

\(^{12} \)Estimation of equation (1), including \( Y_o \) and \( Y_d \) and excluding the exporter and importer fixed effects, yields an even stronger rejection of the Linder hypothesis. In all four specifications, the coefficient on the Linder term is positive and significant at the 1% level.
A. Demand for Quality

The demand side of the model is based on Hallak (2006a). Preferences are represented by a two-tier utility function. The upper tier is weakly separable into sectoral subutility indices \( \{ u_{dz} \}_{z=1, \ldots, Z} \) and \( \{ u_{dg} \}_{g=Z+1, \ldots, S} \), where \( d \) indexes countries (viewed as consumers and importers), \( z \) indexes differentiated-good sectors, and \( g \) indexes homogeneous-good sectors. The (sub)utility derived from consuming goods in differentiated-goods sector \( z \) is given by

\[
u_{dz} = \left[ \sum_{h \in \Omega_d} (\theta_b)^{\gamma_d} q_{dh} \right]^{\alpha_d-1/\alpha_d}, \quad \alpha_d>1, \quad \theta_b, \gamma_d > 0, \quad \sigma_z > 1
\]

where \( \Omega_d \) is the set of all varieties in sector \( z \), \( \theta_b \) is the quality of variety \( h \), \( q_{dh} \) is the quantity of variety \( h \) consumed by country \( d \), \( \gamma_d \) is the intensity of country \( d \)'s preference for quality (in sector \( z \)), and \( \sigma_z \) is the common elasticity of substitution among varieties in the sector. Demand in differentiated sectors is obtained in two stages. In the first stage, conditional on prices and income, the representative consumer chooses an allocation of expenditures across sectors \( \{ E_{dz} \}_{z=1, \ldots, Z} \) and \( \{ E_{dg} \}_{g=Z+1, \ldots, S} \). In the second stage, conditional on total expenditure \( E_{dz} \) in sector \( z \), he chooses expenditure \( e_{dh} \) on variety \( h \), which is given by

\[
e_{dh} = s_{dh} E_{dz} = \left( \frac{p_{dh}}{(\theta_b)^{\gamma_d} G_{dz}} \right)^{1-\sigma_z}, \quad s_{dh} = \sum_{r \in \Omega_z} \left( \frac{p_{dr}}{(\theta_b)^{\gamma_d} G_{dz}} \right)^{1-\sigma_z}
\]

where \( s_{dh} \) is the share of variety \( h \) in sectoral expenditure \( E_{dz} \), \( p_{dh} \) is the price of variety \( h \) faced by consumers in country \( d \), and \( G_{dz} = \sum_{r \in \Omega_z} \left( \frac{p_{dr}}{(\theta_b)^{\gamma_d} G_{dz}} \right)^{1-\sigma_z} \) is an exact consumption price index for sector \( z \) in country \( d \).

Product quality (\( \theta_b \)) is modeled as a utility (or demand) shifter. This utility shifter captures all attributes of a product that consumers value. Conditional on price, a higher quality increases the share spent on a given variety. The parameter \( \gamma \) measures intensity of preference for quality. Countries with higher \( \gamma \) are willing to pay more for a high-quality variety.\(^{13}\) The impact of \( \gamma \) on the share spent on variety \( h \) is given by

\[
\frac{\partial s_{dh}}{\partial \gamma_d} = (\sigma_z - 1) s_{dh} \ln \theta_b - \sum_{r \in \Omega_z} s_{dr} \ln \theta_r.
\]

Since \( \sigma_z > 1 \) and \( s_{dh} > 0 \), the sign of this derivative depends on the sign of the term in square brackets, that is, on whether the quality of variety \( h \) is above or below the (weighted) average quality of the sector. When the quality of variety \( h \) is higher (lower) than the average quality, the term in brackets is positive (negative). In that case, expenditure shares increase (decrease) with the intensity of preference for quality.

Variation across countries in the value of \( \gamma \) provides this demand system with the flexibility to capture cross-country variation in quality demand. Countries with higher \( \gamma \) spend a larger proportion of their income on high-quality goods. Such flexibility is convenient for capturing—in reduced form—the demand side of Linder’s theory, which relates countries’ demand for quality to income per capita. The demand system also embeds a relevant special case. When the intensity of preference for quality does not vary across countries (\( \gamma_d = \gamma \)), the demand system is equivalent to the demand system generated by Dixit-Stiglitz preferences.

Countries viewed as producers and exporters are indexed by \( o \). Exporter \( o \) produces \( N_{oz} \) symmetric varieties in sector \( z \). Denote by \( s_{od} \), the share defined in equation (3) when applied to a typical variety from country \( o \) in sector \( z \). Using this notation, denote by \( \omega_{odz} = N_{oz} s_{od} \), country \( o \)'s total share in country \( d \)'s sectoral expenditure. The amount of country \( d \)'s imports from country \( o \) in sector \( z \) is then

\[
\text{imports}_{od} = \omega_{od} E_{dz}, \quad \omega_{odz} = N_{oz} \left( \frac{p_{oz} \tau_{odz}}{G_{dz}} \right)^{1-\sigma_z},
\]

where the import price in country \( d \) of a variety produced by country \( o \) is expressed as the product of the export price, \( p_{oz} \), and the (iceberg) trade cost factor between countries \( o \) and \( d \), \( \tau_{odz} \). Equation (5) shows that the import share of country \( o \) decreases with the export price \( (p_{oz}) \) of its varieties and increases with their quality \( (\theta_{oz}) \).

To gain intuition about the role that cross-country differences in preference for quality play on the direction of trade, compare the shares of exporters \( o \) and \( o' \) in countries \( d \) and \( d' \)'s imports in sector \( z \), respectively, by focusing on the double ratio (or double difference if logarithms are taken on both sides of the equation):

\[
r_{o/od'/dz} = \frac{\omega_{od'/dz} / \omega_{o'od'/dz}}{\omega_{od} / \omega_{o'od}} = \frac{\tau_{od'/dz} / \tau_{o'od'/dz}}{1}.
\]

This ratio equals 1 when exporters \( o \)'s and \( o' \)'s shares in country \( d \)'s imports are proportional to their shares in country \( d' \)'s imports—numerator and denominator are equal. The ratio is greater than 1 when exporter \( o \)'s share relative to that of exporter \( o' \) is higher in country \( d' \)'s imports than it is in country \( d \)'s imports.

Abstracting (temporarily) from trade costs to simplify the exposition \( \tau_{od'/dz} / \tau_{o'od'/dz} = 1 \) and using the definitions of \( \omega_{odz} \) and \( s_{odz} \), we can express this ratio as
Equation (7) shows that the existence of quality differences between countries $o$ and $o'$ is not sufficient to induce more intense trade between specific country pairs. In the benchmark case with no cross-country variation in $\gamma (\gamma_{dc} = \gamma_{d'c})$, the numerator and denominator of equation (7) are equal even in the presence of quality differences ($\theta_{oz} \neq \theta_{o'z}$). In that case, $r_{oo'd'd'} = 1$, implying that importers $d$ and $d'$ buy goods from exporters $o$ and $o'$ in the same proportions. In contrast, when there exists cross-country variation in $\gamma (\gamma_{dc} \neq \gamma_{d'c})$, quality differences do influence the direction of trade. For example, if exporter $o$ produces goods of higher quality than exporter $o'$ and importer $d$ has a more intense preference for quality than importer $d'$, then $r_{oo'd'd'} > 1$. In this case, the country with higher $\gamma$ (country $d$) will import relatively more from the country that produces higher-quality goods (country $o$). Although both importers prefer high-quality rather than low-quality goods, this preference is more intense for importers with higher $\gamma$.

Equation (5) is the basis of the empirical specification that will be used for estimation. I assume that trade costs are determined by $\tau_{odc} = (Dist_{od})^\gamma_{odc}K_{odc} e^{v_{odc}}$, where $I_{odc}$ is the vector of dummy variables defined in section II and $v_{odc}$ is a random disturbance to bilateral trade costs. Substituting $\tau_{odc}$ for this expression in equation (5) and taking logarithms, we obtain

$$
\ln \text{imports}_{odc} = \varphi_{oz} + \psi_{oz} - \beta_{oz} n_{odc} + \varphi_{o'z}I_{o'z} + \beta_{o'z}n_{o'zd'}.
$$

Equation (9) does not impose any parameter restriction on $\mu_z$. For example, it allows the benchmark case with no cross-country differences in demand for quality ($\mu_z = 0$) discussed above, in which case the distribution of import shares is common across importers—up to differences due to trade costs. The demand side of Linder’s theory, however, asserts that rich countries consume higher-quality goods. This prediction implies that $\mu_z > 0$.\textsuperscript{16}

Combining equations (8) and (9), we obtain

$$
\ln \text{imports}_{odc} = \varphi_{oz} + \psi_{oz} - \beta_{oz} n_{odc} + \varphi_{o'z}I_{o'z} + \beta_{o'z}n_{o'zd'} + \varphi_{z} + \beta_{z}n_{odc} + \varphi_{o'z}I_{o'z} + \beta_{o'z}n_{o'zd'} + \varphi_{z} + \beta_{z}n_{odc} + \varphi_{o'z}I_{o'z} + \beta_{o'z}n_{o'zd'}.
$$

Since product quality ($\theta_{oz}$) is not observable, equation (10) cannot be estimated as such. Hallak (2006a) uses export unit values as indicators of product quality, which implies focusing on only the demand side of Linder’s theory. However, since the Linder hypothesis is based on the interaction of demand and supply forces, a formal characterization of quality relation of their high-quality varieties is presumably less appealing, for a majority of world consumers, than the price-quality relation of the low-quality varieties offered by low-income countries. The former are therefore net importers and the latter net exporters in this sector. In Machinary, the same pattern of within-sector comparative advantage is consistent with the opposite pattern of between-sector comparative advantage, making high-income countries net exporters in this sector and low-income countries net importers.

\textsuperscript{14} This prediction is also implied by standard Dixit-Stiglitz preferences. To see this, note that we can define $\tilde{\theta}_{oz} = 0$; as a preference parameter, common across importers, for goods from country $o$.

\textsuperscript{15} It is worth distinguishing here the notions of within-sector and between-sector comparative advantage. Within sectors, countries have conventional comparative advantage in either low-quality or high-quality varieties. Either case is compatible with a between-sector comparative advantage or disadvantage in the sector as a whole. The country will have a between-sector comparative advantage if the cost of the goods it produces, relative to their quality, is sufficiently low to make it a net exporter in the sector. This will happen when a majority of world consumers find the price-quality relation offered by the country more appealing than the price-quality relation offered by other countries. For example, high-income countries can be expected to have a within-sector comparative advantage in high-quality Apparel. However, the price-quality relation of their high-quality varieties is presumably less appealing, for a majority of world consumers, than the price-quality relation of the low-quality varieties offered by low-income countries. The former are therefore net importers and the latter net exporters in this sector. In Machinary, the same pattern of within-sector comparative advantage is consistent with the opposite pattern of between-sector comparative advantage, making high-income countries net exporters in this sector and low-income countries net importers.

\textsuperscript{16} Aggregate consumption of high-quality goods should also depend on the distribution of income (Choi, Hummels, & Xiang, 2009). The connection between income distribution and the Linder hypothesis is explored empirically by François and Kaplan (1996) and Dalgin, Mitra, and Trindade (2008), although their studies do not address the role of product quality.
Linder’s theory requires relating quality supply to income per capita.

B. Supply of Quality

Linder argued that high-income countries have a comparative advantage in the production of high-quality goods. Several theories can explain a systematic relationship between per capita income and quality production. For example, a Ricardian view of quality specialization predicts that richer countries will export high-quality goods if their productivity advantage in those goods is relatively larger. Alternatively, a Heckscher-Ohlin view of quality specialization predicts that rich countries, which tend to be capital abundant, will export high-quality goods if they are capital intensive. Linder, in turn, proposed that closeness to demand for high-quality products, which occurs disproportionately in high-income countries, provides them with a comparative advantage in the production of those goods. Finally, Vernon’s (1966) product cycle theory17 can also explain high-income countries’ comparative advantage in high-quality production if innovations consist of upgrading the quality of existing varieties.18 The prediction of the Linder hypothesis about the direction of trade can be founded on any of these theories of supply-side determinants of quality production. Their common implications are captured here by simply postulating a systematic relationship between quality supply and per capita income,

$$\ln \theta_{oz} = \delta_z + \delta_o \ln y_o + s_{oz},$$  

where $s_{oz}$ is a stochastic disturbance. The supply side of Linder’s theory predicts that $\delta_z > 0$.

C. Demand and Supply Interaction: The Sectoral Linder Hypothesis

Substituting equation (11) into equation (10), we obtain:

$$\ln \text{imports}_{odz} = \varphi_{oz} + \psi_{dz} - \beta_{Dz} \ln \text{Dist}_{od} + \beta_{Lz} \ln y_o + \ln y_d + \epsilon_{odz},$$

where $\beta_{Dz} = \delta_z \eta_z$, $\beta_{Lz} = \delta_o \xi_z$, $\beta_{yz} = \delta_z \mu_z$, and $\epsilon_{odz} = \delta_o \mu_z s_{oz} + \delta_{dz} v_{odz}$. The disturbances $s_{oz}$ and $v_{odz}$ are assumed to be uncorrelated with the regressors; therefore, so is $\epsilon_{odz}$.

The parameter of interest is the coefficient on the term interacting the trading partners’ per capita incomes, $\beta_{yz} = \delta_z \mu_z \delta_o$. Since $\delta_o > 0$, the sign of $\beta_{yz}$ corresponds to the sign of $\mu_z \delta_z$. Linder’s theory postulates that richer countries consume higher-quality goods ($\mu_z > 0$) and produce higher-quality goods ($\delta_z > 0$). The interaction of these forces then implies that $\beta_{yz} > 0$. This prediction is in fact equivalent to the prediction of the Linder hypothesis as typically specified in empirical exercises. The equivalence is easy to show; basic algebraic manipulation of the term $Linder_{odz} = (\ln y_o - \ln y_d)^2$ yields

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

This expression, substituted back into equation (12), results in

$$\ln \text{imports}_{odz} = \varphi_{oz} + \psi_{dz} - \beta_{Dz} \ln \text{Dist}_{od} + \beta_{Lz} \ln y_o + \ln y_d + \epsilon_{odz},$$

where the exporter and importer fixed effects absorb, respectively, the first two terms of equation (13) and $\beta_{Lz} = - (1/2) \beta_{yz}$. The last equality implies the equivalence between the prediction that $\beta_{yz} > 0$ in equation (12) and the prediction that $\beta_{Lz} < 0$ in equation (14).

The results of this section demonstrate that the Linder hypothesis can be derived from a theoretical framework that captures both the demand and supply mechanisms that Linder originally postulated. They also show that a gravity-equation specification, augmented with a Linder term, is appropriate for testing the role of quality described by the theory when it is formulated at the sector level. I henceforth refer to such formulation as the sectoral Linder hypothesis.

IV. The “Sectoral” Linder Hypothesis: Empirical Results

A. Data

Estimating equation (14) requires bilateral trade flows at the sector level, which I obtain from the World Trade Flows data set (Feenstra, 2000). This data set breaks down trade flows up to the four-digit SITC (Rev. 2) level of aggregation. However, I define sectors at the three-digit SITC level because information at the four-digit level is often missing.

I follow Rauch’s (1999) classification of sectors into three categories—differentiated, reference priced, and homogeneous—and use his “liberal” classification because it is more stringent in the classification of goods as differentiated. When a three-digit sector includes four-digit subsectors that belong to different categories, the three-digit sector is broken down accordingly, each part including only the relevant four-digit sectors.

The criterion for selecting the countries in the sample attempts to balance two considerations. On the one hand, including more countries increases sample size and estimation precision. On the other hand, concentrating on relatively large countries decreases the proportion of bilateral country pairs with zero trade at the sector level, which
prevents zero-trade observations from dominating the sample. The sample consists of 64 countries, all of those with a population larger than 3 million, and with more than $2 billion imports of differentiated goods. I also drop very small sectors, keeping only sectors with a volume of trade (among the 64 selected countries) above $2 billion. In the final sample are 116 differentiated sectors, 56 reference-priced sectors, and 39 homogeneous sectors.21

The variable distance measures great circle distance between capital cities and comes from Shatz (1997). Dummies for border, common language, colonizer-colony relationship, and common-colonizer relationship were constructed using the CIA Factbook. Only “official” languages are taken into account in the construction of the common language variable, except for Malaysia-Singapore, which is coded here as sharing a common language. Colonial links are considered only if the colonizer-colony relationship was still in force after 1922. The indicator variable for preferential trade agreement includes PTAs in force and with substantial coverage in 1995: Andean Pact, ASEAN, CACM, EFTA, EEA, EU, MERCOSUR, NAFTA, Australia–New Zealand, EC–Turkey, EFTA–Turkey, EC–Israel, EFTA–Israel, and United States–Israel. Data on PPP GDP come from the World Bank WDI.

B. Estimation Issues

Equation (14) predicts bilateral trade at the sector level. Aggregation of trade flows, as is commonly done in the literature, implicitly forces the parameters of this equation to be equal across sectors. In particular, this practice precludes the use of sector-specific exporter and importer fixed effects to control for intersectoral determinants of trade. I thus estimate equation (14) by sector. I also estimate pooling the observations across sectors but allowing cross-sector variation in all the parameters and fixed effects except for the Linder coefficient—that is, I impose only $\beta_L z^c = \beta_L$. While in the first case, the estimation yields sector-by-sector estimates of the parameter of interest, in the second case, the estimation yields a single estimate of this parameter.

I focus on differentiated goods because those are the goods for which the assumptions of the theory most clearly apply. Equation (14) can also be estimated using intermediate-good sectors if we interpret equation (2) as a production function of a final good. The interpretation of the sectoral Linder hypothesis in that case is that richer countries consume a larger proportion of high-quality intermediate inputs, presumably as a requirement to produce high-quality final goods.

Estimation of gravity-type equations using OLS is known to suffer from a potential selection bias, as the OLS procedure drops bilateral pairs with zero trade. To deal with this problem, I use a generalized tobit estimation with random and unobserved censoring value. The estimation strategy is based on the idea that zero values of trade are due to the presence of fixed costs of exporting. The magnitude of fixed costs is unobservable but can be modeled as a function of observable variables. In particular, I postulate a censoring equation of the following form:

$$\ln c_{odz} = \delta_{oz} + \delta_{dz} \ln Dist_{od} + \delta_{Iod} + \delta_{xz} \ln Y_o + \delta_{mz} \ln Y_d + u_{odz},$$

(15)

where the (unobserved) censoring value is determined by the level of trade that generates sufficient profits to cover the fixed costs, $F_{odz}$. Given the constant elasticity of demand, the censoring value is proportional to those costs ($c_{odz} = \sigma z F_{odz}$). The vector $\mathbf{I}_{od}$ includes the same dummy variables as in equation (14), and $u_{odz}$ is normally distributed random disturbance. Joint estimation of the parameters of equations (14) and (15), respectively denoted the “imports equation” and the “fixed-cost equation,” can be performed using maximum likelihood (ML) estimation under the assumption that the distribution of the two errors is bivariate normal. This estimation strategy is explained in more detail in Hallak (2006a).20

C. Estimation Results

Equation (14) is first estimated by OLS separately for each of the 116 differentiated sectors. Panel A of table 3 provides a summary report of the estimated coefficients, by sign (columns 1 and 2) and significance level (columns 3 to 5). Standard errors are robust to heteroskedasticity in all specifications. The results support the empirical validity of the sectoral Linder hypothesis. The estimated coefficient on the Linder term is negative, as predicted, in more than two-thirds of the sectors (82) and positive in less than one-third of the sectors (34). The coefficient is negative and significant at the 5% level in approximately half of the sectors (59), and it is positive and significant in less than one-fifth of the sectors (21). The variables that control for trade costs affect trade volumes in the predicted direction: shorter distance promotes trade, as does sharing a common border, common language, PTA, colonial relationship, or common colonizer. Column 6 shows the median magnitude of the estimated coefficients. The median has the predicted sign for all variables.

The last column of the table reports the parameter estimate and standard error for the coefficient on the Linder term when all observations are pooled (but not aggregated) across sectors and the coefficient is constrained to be the same across sectors.21 The magnitude of the estimated coefficient is substantially smaller than the median value of

---

20 Two alternative methods for dealing with zero-trade observations are proposed by Silva and Tenreyro (2006) and by Helpman, Melitz, and Rubinstein (2008).

21 The coefficients on all other variables are not constrained to be equal across sectors. Summary measures of those estimates are not reported to save space.
the sectoral estimates. Nevertheless, the effect of the Linder term is still negative, and it is significantly different from 0 at the 1% level.

It is possible to quantify the impact of the Linder effect on bilateral trade volumes by considering the term $\exp[\hat{\beta}_L (\ln y_o - \ln y_d)^2]$. For example, using the estimate $\hat{\beta}_L = -0.0435$ obtained in the pooled regression, we can compare this term across origin countries of U.S. imports. Substituting for the corresponding per capita incomes in the above equation, we find that the United States is predicted to import from Switzerland 10% more than it imports from Colombia, 20% more than it imports from the Philippines, 40% more than it imports from Pakistan, and 70% more than it imports from Nigeria. Also, any country is predicted to import 14.7% more from a country with identical income per capita than from a country two (log) standard deviations apart in the 64-country sample income distribution (for example, Finland and Ecuador or Argentina and India).

Panel B of table 3 displays the baseline estimates of the Linder effect—those that result from estimating the censoring model using the ML estimator. The censoring model yields qualitatively similar results, although now there are fewer sectors with negative estimates of the Linder term (74) and with negative and significant estimates of this term (55). Also, there is a larger number of sectors with positive estimated coefficients (42) and with positive and significant estimated coefficient (27). The absolute value of the median coefficient on this term is also substantially smaller (0.0396) and is close in magnitude to the pooled sample OLS estimate. The estimates for the other controls in the imports equation have the expected sign in most sectors. They also have the expected sign in the fixed-cost equation, except for common border and importer GDP, which are estimated to have a positive effect on the fixed costs of exporting.

The distribution of OLS and ML coefficient estimates is presented in figure 1. In both graphs, the 116 sectors are arranged along the horizontal line in ascending order according to the size of the estimated coefficient. The middle curve represents the point estimates, while the upper and lower curves represent the standard errors.

### Table 3—Sectoral Linder Hypothesis: Differentiated Goods—OLS and ML Estimates

<table>
<thead>
<tr>
<th>Sign</th>
<th>Positive</th>
<th>Negative</th>
<th>Positive</th>
<th>Not Significant</th>
<th>Negative</th>
<th>Median</th>
<th>Pooled Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. OLS estimation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linder term</td>
<td>34</td>
<td>82</td>
<td>21</td>
<td>36</td>
<td>59</td>
<td>-0.0714 (0.0111)</td>
<td></td>
</tr>
<tr>
<td>Distance</td>
<td>0</td>
<td>116</td>
<td>0</td>
<td>0</td>
<td>116</td>
<td>-1.0154</td>
<td></td>
</tr>
<tr>
<td>Border</td>
<td>107</td>
<td>9</td>
<td>58</td>
<td>58</td>
<td>0</td>
<td>0.3827</td>
<td></td>
</tr>
<tr>
<td>Common language</td>
<td>114</td>
<td>2</td>
<td>101</td>
<td>15</td>
<td>0</td>
<td>0.5766</td>
<td></td>
</tr>
<tr>
<td>PTA</td>
<td>103</td>
<td>13</td>
<td>69</td>
<td>47</td>
<td>0</td>
<td>0.4648</td>
<td></td>
</tr>
<tr>
<td>Colonial link</td>
<td>112</td>
<td>4</td>
<td>88</td>
<td>28</td>
<td>0</td>
<td>0.7742</td>
<td></td>
</tr>
<tr>
<td>Common colony</td>
<td>98</td>
<td>18</td>
<td>29</td>
<td>86</td>
<td>1</td>
<td>0.4026</td>
<td></td>
</tr>
<tr>
<td><strong>B. ML estimation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imports equation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linder term</td>
<td>42</td>
<td>74</td>
<td>27</td>
<td>34</td>
<td>55</td>
<td>-0.0396</td>
<td></td>
</tr>
<tr>
<td>Distance</td>
<td>0</td>
<td>116</td>
<td>0</td>
<td>0</td>
<td>116</td>
<td>-1.1375</td>
<td></td>
</tr>
<tr>
<td>Border</td>
<td>108</td>
<td>8</td>
<td>58</td>
<td>58</td>
<td>0</td>
<td>0.3950</td>
<td></td>
</tr>
<tr>
<td>Common language</td>
<td>115</td>
<td>1</td>
<td>107</td>
<td>9</td>
<td>0</td>
<td>0.6556</td>
<td></td>
</tr>
<tr>
<td>PTA</td>
<td>90</td>
<td>26</td>
<td>57</td>
<td>55</td>
<td>4</td>
<td>0.2544</td>
<td></td>
</tr>
<tr>
<td>Colonial link</td>
<td>114</td>
<td>2</td>
<td>101</td>
<td>15</td>
<td>0</td>
<td>0.8601</td>
<td></td>
</tr>
<tr>
<td>Common colony</td>
<td>101</td>
<td>15</td>
<td>46</td>
<td>70</td>
<td>0</td>
<td>0.4859</td>
<td></td>
</tr>
<tr>
<td>Fixed-cost equation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance</td>
<td>89</td>
<td>27</td>
<td>38</td>
<td>78</td>
<td>2</td>
<td>0.0901</td>
<td></td>
</tr>
<tr>
<td>Border</td>
<td>98</td>
<td>18</td>
<td>23</td>
<td>93</td>
<td>0</td>
<td>0.4571</td>
<td></td>
</tr>
<tr>
<td>Common language</td>
<td>10</td>
<td>106</td>
<td>1</td>
<td>79</td>
<td>36</td>
<td>-0.2687</td>
<td></td>
</tr>
<tr>
<td>PTA</td>
<td>14</td>
<td>102</td>
<td>2</td>
<td>63</td>
<td>51</td>
<td>-0.4298</td>
<td></td>
</tr>
<tr>
<td>Colonial link</td>
<td>31</td>
<td>85</td>
<td>2</td>
<td>99</td>
<td>15</td>
<td>-0.4789</td>
<td></td>
</tr>
<tr>
<td>Common colony</td>
<td>43</td>
<td>73</td>
<td>0</td>
<td>105</td>
<td>11</td>
<td>-0.1643</td>
<td></td>
</tr>
<tr>
<td>Exporter GDP</td>
<td>27</td>
<td>89</td>
<td>2</td>
<td>71</td>
<td>43</td>
<td>-0.0935</td>
<td></td>
</tr>
<tr>
<td>Importer GDP</td>
<td>79</td>
<td>37</td>
<td>44</td>
<td>63</td>
<td>9</td>
<td>0.0598</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Estimation of equation (14) in panel A. Estimation of censoring model equations (14) and (15) in panel B. The Linder term is $(\ln y_o - \ln y_d)^2$. Columns 1 and 2 provide a breakdown of the total number of sectors by sign of the estimated coefficient. Columns 3 to 5 provide a breakdown by sign and significance. Heteroskedasticity-robust standard errors in all regressions. Clustering by country pair across sectors in pooled regression. Standard errors in parentheses.

***, **, * Significant at the 1%, 5%, and 10% level, respectively.
lower bands display, respectively, two standard deviations above and below the point estimates. The graphs show a smooth distribution of point estimates across sectors. The distribution is also quite symmetric around the median estimates, which are negative, as predicted by theory (the median estimates for OLS and ML correspond to those presented in table 3).25

Although the theoretical framework of this paper suggests the use of a particular Linder term, alternative Linder terms can be used to assess the robustness of the results. I estimate equation (14) using the three alternative Linder terms described in section II and find that the results are almost unchanged.26

The fact that estimated coefficients are positive and significant in a number of sectors (21 and 27 sectors when OLS and the censoring model are, respectively, used to estimate) raises the concern that forces correlated with the interaction of the country pair’s per capita incomes but other than those highlighted by Linder’s theory might have a substantial influence on the results. This concern can be addressed by testing the sectoral Linder hypothesis using alternatively reference-priced and homogeneous sectors. As I discuss next, since the assumptions of the theory do not match the characteristics of these goods as they do for goods in differentiated sectors, the results of testing the Linder hypothesis in those alternative cases can provide further evidence as to whether the Linder mechanisms drive the main results.

The sectoral Linder hypothesis is derived under the joint assumption that quality supply and quality demand are systematically related to income per capita. On the supply side, despite the fact that quality differences often exist even among goods that are not classified as differentiated, those differences are likely to be smaller for reference-priced sectors and yet even smaller for homogeneous sectors. Further, even if quality differences exist in those sectors, quality supply is less likely to be correlated with income per capita, mostly so in the case of homogeneous sectors. It is reasonable to think, for example, that among manufactured goods (typically differentiated or reference priced), specialization of a country in the high-quality end is the result of a high proportion of (human or physical) capital to labor. High-income countries, which are usually abundant in capital, then produce high-quality capital-intensive manufactures. In contrast, among agricultural and mineral commodities (typically homogeneous), high-quality output is often linked to a country’s endowment of a specific natural resource whose availability is not necessarily related to the country’s per capita income. This suggests that the supply-side assumptions of the sectoral Linder hypothesis should weaken as we move from differentiated to reference-priced goods and should weaken considerably more as we move from the latter goods to homogeneous goods.

On the demand side, horizontal differentiation is a more appropriate characterization of differentiated sectors than it is of homogeneous sectors. To the extent that quality differences exist in the latter sectors, richer countries might still consume relatively more high-quality products and thus import relatively more from the countries that produce them. However, as there is no justification for a “love for variety” assumption in this case, they will be expected to import, among countries producing high-quality goods, only from those with the lowest price (net of trade costs). Therefore, the bilateral predictions of equation (14) will capture only an average effect.

The combination of a weak relationship between per capita income and quality supply and between per capita income and quality demand implies that the theoretical results of the paper should not apply to homogeneous sectors. In the case of reference-priced sectors, it is not a priori obvious whether the supply and demand relationships predicted by the theory are sufficiently strong to induce bilateral patterns of trade consistent with the Linder hypothesis, leaving its validity as an empirical question.

---

25 A weighted average of the estimated coefficients using total trade in the sector among the 64 countries in the sample as weights yields $-0.0832$ and $-0.0457$ for OLS and ML, respectively, indicating almost no correlation between the estimated sectoral Linder coefficient and sector size.

26 Those results are not reported here but can be found in Hallak (2006b, table 5).
Table 4 compares the distribution of coefficient estimates for the baseline Linder term for the three groups of sectors (results for differentiated goods reproduce those previously reported). Although the theoretical prediction is ambiguous for referenced-priced sectors, the results for those sectors are still consistent with the predictions of the sectoral Linder hypothesis. When OLS is used to estimate, the coefficient is negative in four-fifths of the sectors (43 out of 56) and is negative and significant in more than one-third of the sectors (21). When the censoring model is used to estimate, the coefficient is negative in two-thirds of the sectors (37) and is negative and significant in fewer than half of the sectors (24). Compared to the results for differentiated sectors, a larger proportion of reference-priced sectors displays no significant Linder effect. The discrepancy between the OLS and ML results is not substantial here; for example, the median estimate is very similar in both cases. Finally, the estimation using the pooled sample yields a significantly negative coefficient estimate, close in magnitude to the estimate for differentiated goods. In the case of homogeneous goods, the results are radically different. The estimated coefficient is more often positive than negative, more often significantly positive than significantly negative, and not significant in most sectors. In addition, both the median coefficient and the pooled coefficient are positive (and significant in the latter case). The failure to find Linder effects for homogeneous sectors is consistent with the fact that the characteristics of those sectors violate the assumptions of the theory. Therefore, these results provide further evidence that the interaction of supply and demand for quality stressed by the theory drives the estimates of the Linder effect found for differentiated sectors rather than other unobserved factors correlated with per capita income. The estimates are consistent with the Linder hypothesis only in goods categories where we expect income per capita to be correlated with quality supply and quality demand.

V. Aggregation Bias

This section shows that standard tests of the Linder hypothesis, performed using aggregate trade data, suffer from a systematic bias. Section VA shows that the estimated Linder coefficient increases with the level of aggregation and even reverses its sign when aggregation is high. Section VB demonstrates that failure to control for intersectoral determinants of trade generates an aggregation bias that can account for this sign reversal.

A. Empirical Evidence on the Aggregation Bias

Table 5 displays the results of estimating the Linder hypothesis using differentiated sectors defined alternatively at the three-digit level (116 sectors), two-digit level (36 sectors), and one-digit level (6 sectors). The table also reports estimates of the Linder hypothesis aggregating trade flows for all differentiated sectors. When trade flows are aggregated at the two-digit level, the results are very similar to those obtained at the three-digit level (shown in the first row) in terms of both magnitude and distribution of estimates across significance levels. However, when trade flows are aggregated at the one-digit level, the results no longer support the Linder hypothesis. In particular, the estimated sign switches to positive for both the median of the sectoral estimates and the estimate of the pooled regression. Finally, the reversal of the estimated Linder effect is even starker when all trade in differentiated sectors is aggregated into one sector. In that case, the Linder coefficient is positive and significant at the 1% level. The ML estimates show qualitatively similar patterns.
To describe the bias, we focus on the double ratio \( r_{oo'}d'd' \) defined in equation (6). This ratio is later interpreted as an odds ratio. Double-differencing the estimating equation (14), we obtain

\[
\ln r_{oo'}d'd' = \beta_L^* + \beta_L \tilde{y}_{oo'}d'd' - 2\beta_L (\ln y_o - \ln y_{o'}) (\ln y_d - \ln y_{d'}) + \epsilon_{oo'}d'd',
\]

where tildes denote double differences. We use equation (16) to describe the nature of the bias as it is isomorphic to equation (14) with respect to \( \beta_L^* \).

Consider two sectors: \( z = 1, 2 \). For simplicity, let us focus on the sector \( \beta_L^* = \beta_L \) and \( \tilde{y}_{oo'}d'd' = \tilde{y}_{oo'}d'd' = 0 \). Averaging equation (16) across the two sectors, we obtain

\[
\ln r_{oo'}d'd' = -2\beta_L \tilde{y}_{oo'}d'd' + \epsilon_{oo'}d'd',
\]

where \( \ln r_{oo'}d'd' = \frac{1}{2} (\ln r_{oo'}d'd' + \ln r_{oo'}d'd'z) \) is the (log) geometric average of the sectoral odds ratios, \( \tilde{y}_{oo'}d'd' = (\ln y_o - \ln y_{o'}) (\ln y_d - \ln y_{d'}) \), and \( \epsilon_{oo'}d'd' = \frac{1}{2} (\epsilon_{oo'}d'd' + \epsilon_{oo'}d'd'z) \). If the OLS estimator of \( \beta_L \) in the baseline specification (14) is consistent, so is the OLS estimator of \( \beta_L \) in equation (17).

The appendix demonstrates that if countries of similar income per capita have similar intersectoral patterns of trade, then \( u_{oo'}d'd'A \) is positively correlated with \( -2\tilde{y}_{oo'}d'd' \) and thus induces an upward bias in the estimation of the Linder coefficient \( \beta_L \). Here, the intuition for why aggregation induces systematic bias is provided with an illustrative example.

Consider trade flows between two exporters, the United States and China, and two importers, Switzerland and South Africa, in two different two-digit sectors, Industrial Machinery (SITC 72) and Apparel (SITC 84). Since U.S. per capita income is higher than China’s, while Switzerland’s per capita income is higher than South Africa’s, U.S. quality and Switzerland’s intensity of preference for quality are presumed to be higher in both sectors.

Reported trade flows in 1995 between exporters United States (\( o \)) and China (\( o' \)) and importers Switzerland (\( d \)) and South Africa (\( d' \)) in each of the two sectors are displayed in the first set of rows of table 6. The three vertical panels of this table can be thought of as frequency tables between an exporter variable that takes values \{ \( o, o' \) \} and an importer variable that takes values \{ \( d, d' \) \}. The first set of rows reports exporter shares in import markets, where entry \( od \) is the probability that $1 of \( d' \)’s imports comes from \( o \). The last row reports the double ratio, which in this context is an odds ratio. As predicted by theory, the odds ratio is higher than 1 in both sectors (1.44 in Machinery and 3.32 in Apparel), indicating that in each of them, Switzerland is more likely to import from the United States than is South Africa. The last set of columns shows the effect of aggregation. While the properly constructed odds ratio, \( r_{oo'}d'd' \), is constrained to lie inside the bounds defined by the sectoral odds ratios, the aggregate odds ratio \( r_{oo'}d'd'A \) lies outside those bounds. Furthermore, its value is below 1.
(0.29), indicating a reversal in the direction of association between the exporter and importer variables. This result is a manifestation of Simpson’s paradox, a well-known case of association reversal due to aggregation (Simpson, 1951). Here the reversal is driven by the fact that countries with similar income per capita also have similar sectoral trade patterns. In the aggregate, Switzerland imports relatively more from China because it is a relatively large importer in the sector in which China is a large exporter, Apparel.

Now we can better understand the results of table 5. First, determinants of trade patterns are often similar across threedigit sectors in the same two-digit category. For example, net exporters of Apparel (SITC 84) are often net exporters of both Outer Garments, Mens, of Textile Fabrics (SITC 842), and Outer Garments, Womens, of Textile Fabrics (SITC 843). Thus, the exporter and importer fixed effects that we include when we estimate at the two-digit level can still reasonably control for determinants of trade at the three-digit level. However, when we aggregate at higher levels, the exporter and importer fixed effects are unable to control for the omitted determinants of trade, which are drastically different across two-digit sectors. Thus, in this case, aggregation generates substantial bias, which is maximized when we aggregate all trade in differentiated goods.

VI. Conclusion

Despite the persistent appeal of the Linder hypothesis, a large body of empirical work testing its validity has failed to find robust support for it. This failure is the consequence of using a misspecified empirical benchmark, the gravity equation estimated with aggregate data. Building a theoretical framework that captures the main components of Linder’s theory, this paper shows that the Linder hypothesis should be formulated at the sector level, where intersectoral determinants of trade can be controlled for. The sectoral Linder hypothesis is tested and confirmed empirically. Further, it is shown that aggregation across sectors induces a systematic bias against finding support for Linder’s quality-based theory.

REFERENCES


### APPENDIX

This appendix derives necessary and sufficient conditions for the existence of aggregation bias. Denote by \( x_{a d} \) the exports of country \( a \) to country \( d \) in sector \( z \), and consider for simplicity a case with no Linder effect: \( r^{z}_{a a^{'}} = 1 \) for \( z = 1, 2 \). When the odds ratio is 1 in both sectors,

\[
\frac{x_{a d}}{x_{a d}} = x_{a d} \frac{x_{a d}}{x_{a d}} \Rightarrow x_{a d} = ax_{a d}, \quad x_{a d} = ax_{a d}, \quad a > 0
\]

and

\[
\frac{x_{a d}}{x_{a d}} = x_{a d} \frac{x_{a d}}{x_{a d}} \Rightarrow x_{a d} = cx_{a d}, \quad x_{a d} = cx_{a d}, \quad c > 0.
\]

Since the sectoral odds ratios equal 1, the geometric average \( r^{z}_{a a^{'}} \) also equals 1. Then, \( u_{a d} = 0 \) if \( r^{z}_{a a^{'}} = 1 \), which implies that

\[
\frac{x_{a d} + x_{a d}}{x_{a d} + x_{a d}} \leq \frac{x_{a d} + x_{a d}}{x_{a d} + x_{a d}},
\]

Using (19) and (20), this last inequality can be expressed as

\[
(a-c)x_{a d}x_{a d} < (a-c)x_{a d}x_{a d}.
\]

Without loss of generality, assume that \( y_a > y_{a^{'}} \) and \( y_a > y_{a^{'}} \) so that \( \bar{y}_{a d} = 0 \). To show that aggregation bias occurs if intersectoral patterns of trade are related to per capita income, assume, also without loss of generality, that higher income per capita is associated with a comparative advantage in sector \( z = 1 \). This condition can be formally expressed as

\[
\frac{x_{a d}}{x_{a d}} = x_{a d} \frac{x_{a d}}{x_{a d}} \Rightarrow x_{a d} = cx_{a d}, \quad x_{a d} = cx_{a d}, \quad c > 0.
\]

The first inequality states that since exporter \( a \) has a comparative advantage in sector 1, it exports relatively more in that sector compared to exporter \( a' \). The second inequality states that since importer \( d \) has a comparative advantage in sector 1, it imports relatively less in that sector compared to importer \( d' \). These inequalities can be simplified using equations (19) and (20) to

\[
x_{a d}x_{a d} > x_{a d}x_{a d} \quad \text{and} \quad c < a.
\]

Conditions (22) imply that equation (21) also holds. Therefore, they are sufficient to induce a downward bias in the aggregate odds ratio:

\[
u^{z}_{a a^{'}} = 0. \quad \text{Since} \quad -2\ln u^{z}_{a a^{'}} < 0 \quad (\text{because} \quad \bar{y}_{a d} > 0), \quad u^{z}_{a a^{'}} \text{ and } -2\ln u^{z}_{a a^{'}} \text{ are positively correlated generating an upward bias in the estimate of } \beta_z.
\]