

Choi, Yo Chul; Hummels, David; Xiang, Chong

**Working Paper**

## Explaining import variety and quality: The role of the income distribution

LIS Working Paper Series, No. 541

**Provided in Cooperation with:**

Luxembourg Income Study (LIS)

*Suggested Citation:* Choi, Yo Chul; Hummels, David; Xiang, Chong (2010) : Explaining import variety and quality: The role of the income distribution, LIS Working Paper Series, No. 541, Luxembourg Income Study (LIS), Luxembourg

This Version is available at:

<https://hdl.handle.net/10419/95434>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Explaining Import Variety and Quality: the Role of the Income Distribution

Yo Chul Choi<sup>\*</sup>  
David Hummels<sup>\*\*</sup>  
Chong Xiang<sup>\*\*</sup>

**Abstract:** In this paper we examine whether a generalized version of Flam and Helpman's (1987) model of vertical differentiation can reconcile three facts. One, countries import only a subset of available varieties. Two, import prices vary across exporters within narrow product categories. Three, US growth in both import variety and import price dispersion has occurred at the same time that the US income distribution has significantly widened. The generalized model maps cross-country differences in income distributions to variation in import variety and price variation. The theoretical predictions are examined and confirmed using panel data on import variety and prices, and detailed income distribution data from the Luxembourg Income Survey(LIS). Country pairs whose income distributions are growing more similar over time have growing similarity in the distribution of their import prices, and in the number of common export sources from which they buy.

\* Monetary Policy Analysis Team, Monetary Policy Dept., the Bank of Korea

\*\* Economics Dept., Purdue University

## 1. Introduction

The empirical literature on product differentiation in trade has grown rapidly, fueled by increased availability of detailed trade data and an enduring interest in the role of differentiation in determining trade flows. An important part of this new literature focuses on the role of quality differentiation in trade. Several authors have focused on the existence of substantial variation in import unit values (henceforth: prices) across exporters within narrowly defined goods categories. Prices covary in predictable ways with exporter characteristics (Schott 2004, Hummels and Klenow 2005), and trade costs (Hummels and Skiba 2004). Further, countries with high export prices have larger, not smaller, shares of the markets in which they sell (Hallak 2003). These facts point to the primacy of quality differentiation, as opposed to measurement error, as an explanation for price variation.

In this paper we examine whether a generalized version of Flam and Helpman's (1987) model of vertical differentiation can reconcile three additional facts. One, countries import only a subset of available varieties, and countries such as the US have exhibited rapid growth in import variety (Feenstra 1994, Broda and Weinstein 2005). Two, Schott (2004) shows not just that US import prices vary across exporters within narrow product categories, but that the degree of price dispersion has grown over time. Three, growth in both import variety and import price dispersion has occurred at the same time that the US income distribution has significantly widened.

While Flam and Helpman's North-South trade model covers a rich set of issues related to vertical product differentiation (e.g., income distribution effects, population growth, and technical progress), we focus on demand side implications linking consumer incomes to quality choice. In this model, high income consumers buy higher quality rather than higher quantities of a differentiated good. When focusing on international trade data and cross-country comparisons, it is not possible to see household consumption decisions. However, we show that the model can also be read in terms of income distributions, and it is here that we provide a theoretical contribution. We extend the Flam and Helpman model to the case of multiple vertically differentiated goods and multiple countries. We then derive implications of the model that are testable using data on national distributions of income and import prices.

We characterize the income distribution of a country in terms of its similarity to other importers and to the world income distribution. We show how this income similarity measure maps into observable variation in the number and price distribution of imported varieties. Two importers with dissimilar income distributions have fewer export partners in common, and a less similar import price distribution. Countries that are the least similar to the world income distribution import fewer varieties and have an import price distribution less similar to the rest of the world.

Model predictions are examined using panel data on trade and the income distribution involving 25 countries over 20 years. Our second contribution lies in employing data from the Luxembourg Income Study (LIS) to construct distributions of income both within and across countries. For our purposes it is not sufficient to employ conventional measures of within-country income inequality. Very poor countries might have extremely high degrees of income inequality according to standard measures such as the Gini coefficient or the 90/10 ratios of income.<sup>1</sup> However, these countries will span but a small portion of the world income distribution. In contrast, a high income country might have low within-country inequality according to the Gini or 90/10 measures, but span a much larger portion of the world income distribution.

The LIS income data enable us to compare time series variation in countries like the US, where the income distribution has widened to countries, such as Sweden, where it has not. We can also use cross-sectional data to contrast countries such as Russia and Mexico that have extreme within-country inequality but whose distribution spans a relatively small portion of the world income distribution, to countries that have much less within country inequality (Norway, Finland, Sweden) but whose income distribution spans more of the world distribution. We find strong confirmation of the theoretical predictions both in cross-sectional samples and in samples that employ purely within country time series variation in variables of interest.

---

<sup>1</sup> Dalgin, Mitra, and Trindade (2004) look at the role of income inequality in determining import demand, measuring income inequality using within-country Gini coefficients. They find that imports of luxury goods are increasing in the importing country's income inequality, imports of necessities decrease with it.

Previous authors have examined the Flam and Helpman implication, but looked only at the first moments of the income and price distributions. That is, countries with high mean income per capita buy goods with higher mean prices (Hallak 2003, Hummels and Skiba 2004). Looking at the entire distribution is interesting because it yields richer predictions about trade, and because it carries potentially important normative implications for variety growth. Our empirical findings suggest that a widening income distribution may be at least partly responsible for the growth in import variety we observe in the data. In models with representative agents and love of variety utility, variety growth translates into first order welfare gains in the economy. This is true in a setting of purely horizontal differentiation (Feenstra 1994, Romer 1994, Klenow and Rodriguez-Clare, Broda and Weinstein 2005), or in models that nest vertical differentiation into a homothetic love of variety framework (Hummels and Klenow 2005, Hallak and Schott 2005). However, in the Flam-Helpman world, variety growth can result from a widening of the income distribution with no particular welfare implication. Similarly, growth in variety at the low end of the price distribution can yield welfare gains for poor consumers, but not for consumers in the rest of the income distribution.

Our paper is also related to the literature on how non-homothetic preferences affect trade patterns (e.g. Markusen 1986, Hunter 1991, Mitra and Trindade 2005, Reimer 2005). Most of this work allows for differences in income-expenditure paths across broad industries or product categories, and relates cross-country differences in these expenditures to differences in mean incomes per capita. An exception is Dalgin, Mitra and Trindade (2004), who show that the imports of luxury goods increase with the importing country's Gini coefficient while the imports of necessities decrease with it. We differ from previous work in two respects. First, we are explicit about the role of quality differentiation as the source of the non-homotheticity and allow it to operate within rather than across product categories. Second, our theory requires us to examine data on income distributions both within and across countries. We show that income distribution measures focused on purely within-country inequality such as a Gini coefficient or the ratio of incomes at the 90<sup>th</sup> percentile / 10<sup>th</sup> percentiles are neither theoretically appropriate nor do they yield the predicted sign when taken to the data.

There is a rich theoretical literature describing positive and normative aspects of vertical product differentiation. Our work is closest to the ideas in Flam and Helpman 1987, and especially to Murphy and Schleifer's (1997) insight that rich countries may not trade with less developed countries unless they can produce the high quality goods demanded by high income consumers. Other authors have combined vertical differentiation with non-homothetic preferences and income distributions to shed light on many questions that are difficult for horizontal differentiation models to answer. They show that one country's income re-distribution policy may affect another country's income distribution (Flam and Helpman 1987, Matsuyama 2000), that absolute poverty and per capita growth can be sustained simultaneously in a fully integrated world economy (Funk 1998), that an export boom may push a country into industrialization in the presence of a large middle class (Murphy, Schleifer and Vishny 1989), and that an improvement in the productivity of one industry may trigger the take-off of a series of industries one after another (Matsuyama 2002). While we do not directly address these implications, our paper is a first step in taking the common elements of these models—the interactions of vertical differentiation with non-homothetic preferences and income distribution—to the data.

The paper proceeds as follows. Section 2 provides theorems linking a country's income and import price distribution. Section 3 explains the data set. The income distribution data come from LIS and the import data come from the UN trade database and Feenstra et al. (2005). Since comparable income distribution data are critical in this study, we describe in detail where we obtained the data and how we adjusted the original data and calculated income distribution measures. We also describe techniques for extracting “clean” import price signals from very noisy trade data. In section 4, we present the empirical results about the relationship between import quality demand and income distribution. Section 5 concludes.

## 2. The Model

We extend Flam and Helpman (1987) to a multi-country multi-good setting in order to generate empirical predictions relating an importer's income distribution to the distributions of prices for imported goods. We start with a closed economy model to build intuition and then extend our predictions to the open economy.

## 2.1. Closed Economy: Preferences and Income

There are two goods, a homogeneous numeraire good and a vertically differentiated good. Consumers are identical except for income level. Income is exogenously distributed across the population,  $N$ , according to the probability distribution function  $g(\cdot)$ , with support  $G$ .

A consumer of income  $I$  chooses quantities of the numeraire,  $y$ , and the desired quality,  $z \in [0, 1]$ , of a single unit of the differentiated good in order to maximize

$$u(y, z) = ye^{\alpha z} \quad \text{s.t.} \quad y + p(z) \leq I, \quad (1.1)$$

where  $\alpha > 0$ ,  $\alpha z$  is the elasticity of utility with respect to quality,  $p(z)$  is the price of the differentiated good with quality  $z$ , and the price of the numeraire is set to 1. We assume that income are sufficiently high so that every consumer consumes the differentiated good.

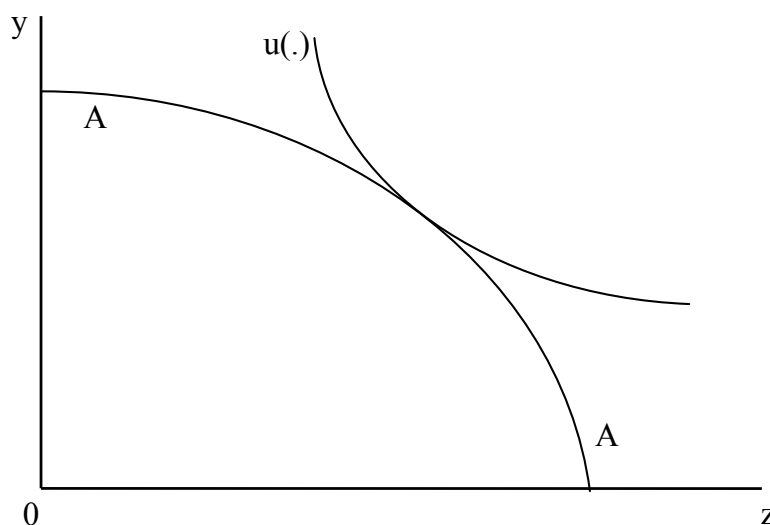


Figure 1.1 Consumer's Choice Problem in the Closed Economy

The marginal cost of producing quality  $z$  is:

$$MC(z) = e^{\gamma z} w \quad (1.2)$$

$w$  represents a cost component that is common to all the quality levels.  $e^{\gamma z}$  represents the cost component that is unique to quality  $z$  and implies that the marginal cost

increases exponentially with  $z$ .  $\gamma z$  is the elasticity of the marginal cost with respect to quality.

We assume that there are perfectly competitive markets at each quality level so that  $p(z) = MC(z)$ . Figure 1.1 shows the utility maximization problem (1.1) for a consumer with income  $I$  and has  $y$  on the vertical axis and  $z$  on the horizontal axis. The budget constraint  $AA$  is concave because by equation (1.2), the higher is the quality level,  $z$ , the faster the price of the differentiated good,  $p(z)$ , increases with quality. When the indifference curve  $u(\cdot)$  is tangent to  $AA$ :

$$z = \frac{1}{\gamma} \left[ \log \frac{\alpha}{\alpha + \gamma} + \log I - \log w \right] \quad (1.3)$$

$$p(z) = aI, \text{ where } a = \frac{\alpha}{\alpha + \gamma} \quad (1.4)$$

Equation (1.4) indicates that a consumer with income  $I$  spends a fixed fraction

$a = \frac{\alpha}{\alpha + \gamma}$  of his income on (one unit of) quality  $z$ .

Summarizing, from equation (1.2) we have that for those qualities available to the market, higher qualities command higher prices. From (1.3), consumers with higher income purchase higher qualities. From (1.4), consumers with higher income pay higher prices. Suppose we are unable to observe prices of goods consumed at the household level, but are instead able to observe the distribution of prices consumed at the national level. These statements allow us to write the distribution of prices as a function of the country's income distribution.

Since optimal qualities (and therefore prices) are monotonically increasing in income, each income level has a unique quality it wishes to consume. For each quality  $z^*$  there is some income level  $I(z^*)$  for which  $z^*$  is the optimal quality. If there is no mass in the income distribution at  $I(z^*)$ , then  $z^*$  is not produced or consumed in equilibrium. Conversely, for every  $I(z^*)$  with positive mass, the quality  $z^*$  will be produced and consumed.<sup>2</sup> The number of people consuming  $z^*$  is precisely the number of persons with income  $I^*$ . As a consequence, the price distribution is a

---

<sup>2</sup> This is an implication of assuming no fixed costs of production and perfectly competitive markets.



direct mapping from the income distribution. The precise functional form of that mapping depends on the elasticities of marginal cost and marginal utility with respect to  $z$ .

This can be seen most clearly using a simple example. Suppose income is distributed log normally  $g \propto N(\mu, \sigma^2)$ . Since prices are strictly increasing in income we can rewrite income as an inverse function of prices, or  $I = \frac{p}{a}$ . The observed price distribution is also distributed log normally  $p \propto N(\mu a, \sigma^2 a^2)$ . The mean and variance of the price distribution are directly proportional to the mean and variance of the income distribution, respectively. As consumer gains from quality ( $\alpha$ ) rise, or the cost of producing quality ( $\gamma$ ) falls, the mean and variance of the price distribution also rise. For the more general income distribution  $g(\cdot)$  with support  $G$ , let  $f_p(I) = aI$ . Then we have a price distribution<sup>3</sup>

$$h(p(z)) = g\left(\frac{p(z)}{a}\right) \cdot \frac{1}{a} \quad \text{with support } H = f_p(G) \quad (1.5)$$

This idea can be made more general still. We need only that qualities and prices are strictly increasing in income. Suppose that  $f_p(I)$  takes a more general form than  $aI$ . Assume that  $f_p(I)$  is strictly increasing in income and that its inverse exists and is differentiable. Then equation (1.5) also takes a more general form:

$$h(p(z)) = g\left(f_p^{-1}(p_z)\right) \cdot [f_p^{-1}(p_z)]' \quad \text{with support } H = f_p(G) \quad (1.6)$$

## 2.2. Multiple Countries

With the closed economy intuition in hand, we can extend the model to a multi-country setting. There are  $C$  countries. Each country  $c$  has population  $N_c$ , with income distributed exogenously<sup>4</sup> according to the probability distribution function  $g_c(\cdot)$  with support  $G_c$ .

The number of people with income  $I$  from country  $c$  equals  $N_c g_c(I) dI$ , while the

---

<sup>3</sup> For example, if  $G = [0, b]$ , then  $H = [0, ab]$ .

<sup>4</sup> Distributing income exogenously allows us to focus on the role of national and world income distributions in determining quality demand, but we abstract from some feedback channels through which trade affects income, as in Flam and Helpman's seminal work.

number of people with income  $I$  worldwide equals  $\sum_c N_c g_c(I) dI$ . Let  $G_w$  be the support of the world income distribution (i.e.  $G_w$  is union of  $G_1, G_2, \dots, G_C$ ),  $N = \sum_c N_c$  be the world population, and  $\lambda_c = N_c / N$  be country  $c$ 's share in the world population. Then the world income distribution has the pdf:

$$g_w(\cdot) = \sum_c \lambda_c g_c(\cdot) \quad \text{with support } G_w. \quad (1.7)$$

The marginal cost of producing quality  $z$  in country  $c$  is:

$$MC_c(z) = e^{\gamma_c z} w_c \quad (1.8)$$

$w_c$  represents cost differences (due to factor price or Ricardian technology differences) that are common to all quality levels.  $e^{\gamma_c z}$  expresses the degree to which country  $c$  has a comparative advantage in high or low quality levels.

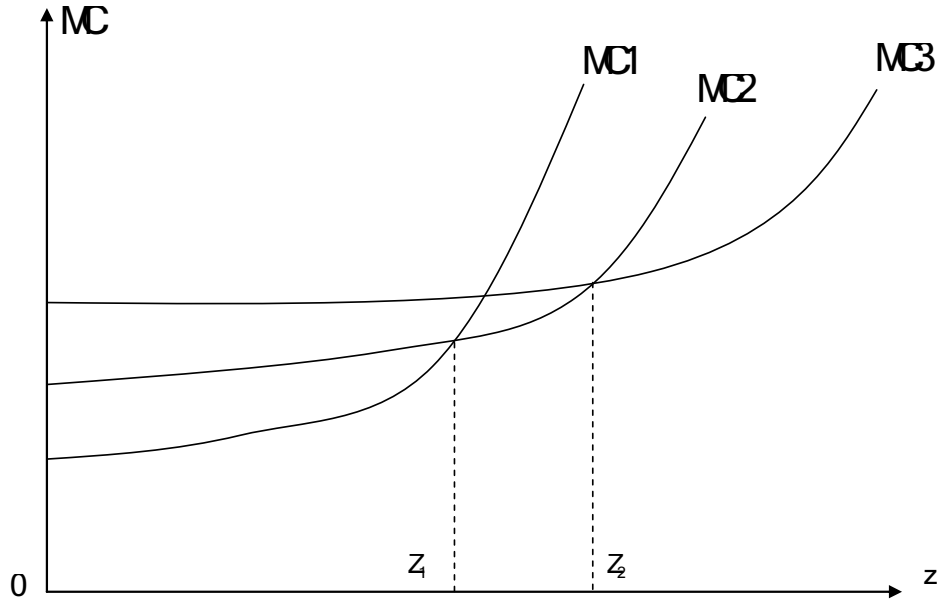


Figure 1.2 Marginal Cost and Comparative Advantage of Quality Production

For simplicity, we assume there are no trade costs. This means that consumers desiring quality  $z$  will buy it from the lowest marginal cost provider. This is illustrated in Figure 1.2, with marginal cost of quality measured on the vertical axis and the quality level on the horizontal axis.  $MC1 \sim MC3$  are the marginal cost curves for countries 1 ~ 3, with  $w_1 < w_2 < w_3$  and  $\gamma_1 > \gamma_2 > \gamma_3$ . This implies that countries 1, 2, and 3 have comparative advantages in the low, medium, and high ranges of quality,

respectively.

Let  $\Lambda_c \subset [0, 1]$  be the set of qualities that country  $c$  could produce.<sup>5</sup> Assume that as the number of countries,  $C$ , gets large, the “mass” of  $\Lambda_c$  goes to 0 for all  $c$

Assumption 1 Let  $m$  be a finite number. Then as  $C \rightarrow +\infty$ ,  $\int_{\Lambda_c} m dz \rightarrow 0$  for all  $c$ .

We re-visit this assumption when discussing empirical implementation in section 3.

Unlike in the closed economy case, having multiple providers of quality with varying  $MC_c(z)$  creates kinks in the budget set. Despite this, equations (1.3) and (1.4) hold with small adjustments. To illustrate this point, we use the two-country setting of Flam and Helpman (1987), where North and South have technologies  $MC_N(z) = e^{\gamma_N z} w_N$  and  $MC_S(z) = e^{\gamma_S z} w_S$ ,  $\gamma_N < \gamma_S$  and  $w_S < w_N$ . North has the comparative advantage in high qualities. Assume that  $z \in [0, 1]$ .

Figure 1.3 shows the utility maximization problem for a consumer with income  $I_d$  and  $u(\cdot)$  is the indifference curve. Figure 1.3 is similar to Figure 1.1 except that the budget constraint now has two segments. When quality is low, it is cheaper to produce the differentiated good in the South and so the budget constraint is determined by the Southern marginal cost (along the curve  $A_S T$ ). When quality is higher than point  $T$ , the budget constraint is determined by the Northern marginal cost (along curve  $TA_N$ ). The income level  $I_d$  is such that both segments of the budget constraint are tangent to the indifference curve; i.e. a consumer with income  $I_d$  is indifferent between buying the differentiated good from the North and buying it from the South. Let  $z_1$  and  $z_2$  be the quality levels associated with the tangent points. Then there is no demand for the qualities between  $z_1$  and  $z_2$ .

---

<sup>5</sup>  $\Lambda_c$  is not the set that country  $c$  actually produces at equilibrium because some qualities in  $\Lambda_c$  may face zero demand. See section 1.2.3.

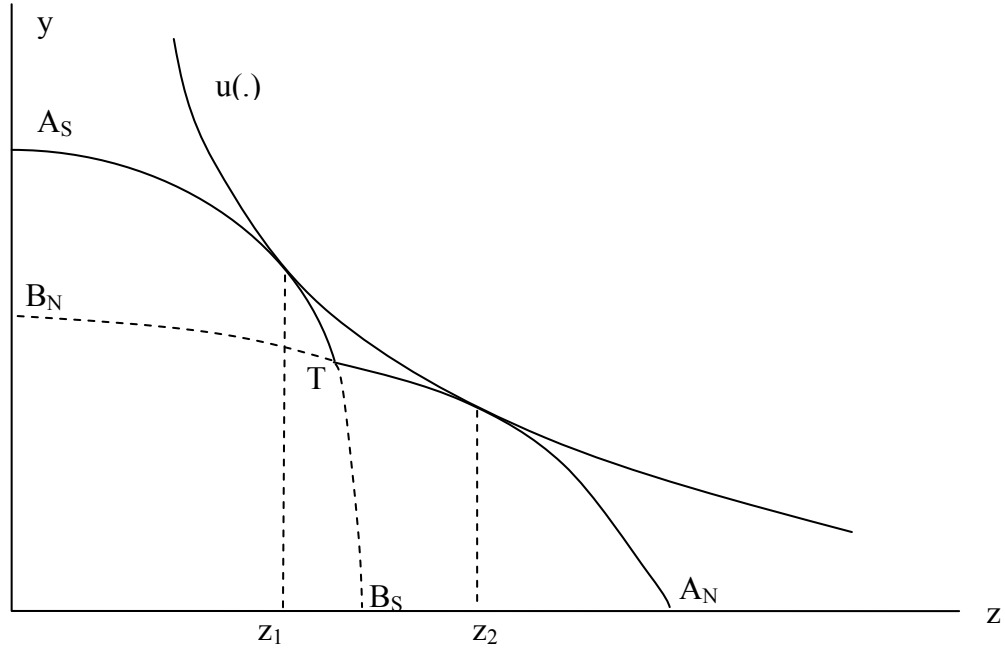


Figure 1.3 Consumer's Choice Problem in the Open Economy

However, for the qualities that are actually supplied to the market,  $[0, z_1] \cup [z_2, 1]$ , higher qualities still command higher prices, consumers with higher income still consume higher qualities, and equations (1.3) and (1.4) still hold, except that for the income below  $I_d$ ,  $\gamma_S$  and  $w_S$  replace  $\gamma$  and  $w$ , and for the income above  $I_d$ ,  $\gamma_N$  and  $w_N$  replace  $\gamma$  and  $w$ . This is an important point for the empirical work that follows. While the theory taken literally suggests a continuous distribution of prices, observable prices are necessarily discrete. This is because, one, money prices are not infinitely divisible, and two, in our data we observe average prices for goods sold by particular exporters rather than a distribution of exact transactions prices. The kinks in the budget set implied by Figure 1.3 are not empirically distinguishable from these other reasons we observe discrete prices. Accordingly, we will use discrete prices to estimate continuous price distributions.

Let  $j = 1, \dots, J$  index exporting countries and let  $G_j$  be the set of income with which a consumer buys the differentiated good from  $j$ . Since every consumer buys the differentiated good from somewhere,  $\cup_j G_j = G_w$  (recall that  $G_w$  is the support of the world income distribution). Then equation (1.4) becomes:

$$p(z) = a_j I \quad \text{for } I \in G_j, \text{ where } a_j = \frac{\alpha}{\alpha + \gamma_j} \quad (1.9)$$

Let  $f_{pj}(I) = a_j I$  and let  $G_{cj} = G_c \cap G_j$ .  $G_{cj}$  is the set of income with which a consumer in country  $c$  buys the differentiated good from exporter  $j$ . Then the price distribution of country  $c$  is still the transformation of the income distribution of  $c$ :

$$h_c(p(z)) = g_c \left( \frac{p(z)}{a_j} \right) \cdot \frac{1}{a_j} \quad \text{for } p(z) \in f_{pj}(G_{cj}), \text{ with support } H_c = \cup_j f_{pj}(G_{cj}) \quad (1.10)$$

To illustrate equation (1.10), consider the Flam and Helpman (1987) two-country example again. Suppose the support of the South's income distribution is  $[0, b_S]$ . Then the consumers with income  $[0, I_d]$  buy the differentiated good from the South and those with income  $(I_d, b_S]$  buy it from the North. The price distribution of the South is

$$g_S \left( \frac{p(z)}{a_S} \right) \cdot \frac{1}{a_S} \quad \text{for } p(z) \in [0, a_S I_d] \quad \text{and} \quad g_N \left( \frac{p(z)}{a_N} \right) \cdot \frac{1}{a_N} \quad \text{for } p(z) \in (a_N I_d, a_N b_S], \text{ with}$$

$$a_S = \frac{\alpha}{\alpha + \gamma_S} \quad \text{and} \quad a_N = \frac{\alpha}{\alpha + \gamma_N} .$$

Thus in the multi-country case, the support of the price distribution consists of disjoint intervals that correspond to the ranges of qualities *actually supplied in the world*. These intervals exclude the qualities that are not demanded by any income level and so not demanded by any country. Since these qualities have zero consumption shares and carry zero weights in the price distribution, equation (1.10) has the same intuition as (1.5), and like (1.5), still holds when  $f_{pj}(\cdot)$  takes a more general form than  $a_j I$ , provided that  $f_{pj}(\cdot)$  is strictly increasing in  $G_j$ , its inverse exists and is differentiable in  $G_j$ .

### 2.3. Measuring the Differences in Distributions

Equation (1.10) maps the price distribution of the differentiated good in country  $c$  to its income distribution, and implies that cross-country differences in the distribution of income will be reflected in the differences in the distribution of prices. We measure cross-country differences in income and price distributions using the following dis-similarity index.

**Definition 1** Dis-similarity Index (DSI): The dis-similarity index (DSI) for the pair of distributions with pdf's  $f_1(\cdot)$  and  $f_2(\cdot)$ , with supports  $S_1$  and  $S_2$  is

$DSI(f_1, f_2) \equiv \frac{1}{2} \int_{\underline{S}} |f_1(x) - f_2(x)| dx$  where  $\underline{S} = S_1 \cup S_2$ ,  $f_1(\cdot)$  is defined to be 0 for  $\underline{S} - S_1$  and  $f_2(\cdot)$  defined to be 0 for  $\underline{S} - S_2$ .

The DSI quantifies the difference between  $f_1(\cdot)$  and  $f_2(\cdot)$  by calculating the vertical distance between them at every point  $x$  and then aggregating these vertical distances. If  $f_1(\cdot)$  and  $f_2(\cdot)$  are dis-similar, they lie far away from each other, the vertical distances between them are large and so  $DSI(f_1, f_2)$  is large. Because both  $f_1(\cdot)$  and  $f_2(\cdot)$  are pdf's,  $DSI(f_1, f_2)$  exists and is bounded between 0 and 1.

Writing out the income similarity index explicitly, we have

$$IDSI(g_1, g_2) \equiv \frac{1}{2} \int_{\underline{G}} |g_1(\cdot) - g_2(\cdot)| dI \quad \text{where } \underline{G} = G_1 \cup G_2 \quad (1.11)$$

$g_1(I)$  is the height of country 1's income pdf at income level  $I$  and  $g_1(I)dI$  is the share of country 1's population that has income  $I$ . The income dissimilarity index simply measures the difference in population shares at each income level and then sums the difference over the support of the income distribution.

On the other hand, since the support of the price distribution consists of disjoint intervals (see section 2.2), the price dis-similarity index is the sum of integrals over these intervals. Let  $\underline{G}_j = G_{1j} \cup G_{2j}$  be the set of income with which a consumer in countries 1 and 2 buys the differentiated good from exporter  $j$ . Writing out the price dissimilarity index explicitly, we have<sup>6</sup>

$$PDSI(h_1, h_2) \equiv \frac{1}{2} \sum_{j=1}^C \int_{\underline{H}_j} \left| g_1 \left( \frac{p(z)}{a_j} \right) \cdot \frac{1}{a_j} - g_2 \left( \frac{p(z)}{a_j} \right) \cdot \frac{1}{a_j} \right| dp(z) \quad \text{where } \underline{H}_j = f_{pj}(\underline{G}_j) \quad (1.12)$$

We show that

**Proposition 1** Assume that trade costs are zero. Then  $PDSI(h_1, h_2) = IDSI(g_1, g_2)$  where 1 and 2 represent any country pair.

Behind Proposition 1 is a very simple idea. Prices are a one to one mapping from income, so the quantity consumed of a good with price  $p(z^*)$  is just the number of persons in a country with income  $I^*$ , and the share of good  $p(z^*)$  is just the population

---

<sup>6</sup> Note that  $\cup_j \underline{G}_j = \underline{G}$  and so  $\cup_j \underline{H}_j = \underline{H} = \underline{H}_1 \cup \underline{H}_2$ , where  $\underline{H}_1$  and  $\underline{H}_2$  are as defined in equation (1.10).

share of persons with income  $I^*$ . Thus the difference between the consumption shares for  $p(z^*)$  in countries 1 and 2 is just the difference in their population shares at  $I^*$ . When integrating over differences in the price distributions we simply recover the differences in the income distributions. We use two examples to illustrate Proposition 1 below.

For the first example, suppose that the income distribution in two countries has the same support,  $G_1 = G_2 = [0, b]$  and technology is identical across exporters,  $\gamma_j = \gamma$  and  $w_j = w$  for all  $j$ . Then  $\underline{G} = [0, b]$  and  $IDSI(\cdot) = \frac{1}{2} \int_0^b |g_1(\cdot) - g_2(\cdot)| dI$ . On the other hand,  $f_{pj}(\cdot) = aI$  for all  $j$  and so

$$PDSI(\cdot) = \frac{1}{2} \int_0^{ab} \left| g_1\left(\frac{p(z)}{a}\right) - g_2\left(\frac{p(z)}{a}\right) \right| \frac{1}{a} dp(z) = IDSI(\cdot) \quad (1.13)$$

For the second example, consider the Flam and Helpman (1987) setup of section 1.2.2 again. Suppose that the support of the North's income distribution is  $[0, b_N]$ ,  $b_N > b_S$  (recall that the South's income distribution has support  $[0, b_S]$ ). Then,

$$\begin{aligned} PDSI(\cdot) &= \frac{1}{2} \int_{a_N b_S}^{a_N b_N} g_N\left(\frac{p(z)}{a_N}\right) \frac{1}{a_N} dp(z) \\ &\quad + \frac{1}{2} \int_0^{a_S I_d} \left| g_N\left(\frac{p(z)}{a_S}\right) - g_S\left(\frac{p(z)}{a_S}\right) \right| \frac{1}{a_S} dp(z) \\ &\quad + \frac{1}{2} \int_{a_N I_d}^{a_N b_S} \left| g_N\left(\frac{p(z)}{a_N}\right) - g_S\left(\frac{p(z)}{a_N}\right) \right| \frac{1}{a_N} dp(z) \end{aligned} \quad (1.14)$$

and  $IDSI(\cdot) = \frac{1}{2} \int_{b_S}^{b_N} g_N(\cdot) dI + \frac{1}{2} \int_{I_d}^{b_S} |g_N(\cdot) - g_S(\cdot)| dI + \frac{1}{2} \int_0^{I_d} |g_N(\cdot) - g_S(\cdot)| dI$ . Each of the three terms in  $PDSI(\cdot)$  equals its counterpart in  $IDSI(\cdot)$  and so  $PDSI(\cdot) = IDSI(\cdot)$ .

Two things are noteworthy about Proposition 1. First, even with multiple differentiated goods, the price dissimilarity index will be the same for each product. We illustrate this point using the closed-economy setting of section 1.2.1 below. Equations (1.3) through (1.6) can be easily extended to the case of multiple differentiated goods. Let  $k = 1 \dots K$  index the differentiated goods and  $z_k$  denote the quality of good  $k$ . The consumer preferences are

$$u = ye^{\sum_k \alpha_k z_k} . \quad (1.15)$$

The marginal cost of each differentiated good is given by equation (1.2), though  $\gamma$  may differ across goods. In this case, consumers spend a fixed fraction of their income on each differentiated good just as in equation (1.4), and the remaining implications go through for each differentiated good. In particular, the mapping from prices to income for good  $k$  is given by

$$p_k(z) = \frac{\alpha_k}{\alpha_k + \gamma_k} I \quad (1.16)$$

Clearly, the slopes of the price-income relationship will differ across goods  $k$ .<sup>7</sup> Goods with high marginal utility of quality and/or low marginal cost of quality will have a steep price-income slope. For a given distribution of income, these goods will have price distributions that have higher means and variances (see section 1.2.1). However, the price DSI is the difference between two countries' price distributions and this removes the variation in the price-income slopes across goods. Thus:

**Corollary 1** In the case of multiple differentiated products, for a given pair of countries  $c$  and  $c'$ ,  $PDSI^k(h_c^k, h_{c'}^k) = IDSI(g_1, g_2)$  for every differentiated product  $k$ .

Corollary 1 suggests that we pool the observations of all the products in implementing regression (1.18).

Finally, we also find it useful to compare the income and price distributions of a country  $c$  with the world distributions. Because every consumer consumes one unit of the differentiated good, the world price distribution has the pdf:

$$h_w(.) = \sum_c \lambda_c h_c(.) \text{ with support } H_w = \cup_c H_c \quad (1.17)$$

where  $\lambda_c$  is the population share of country  $c$  and  $H_c$  is as defined in equation (1.10).

**Corollary 2**  $PDSI^k(h_c^k, h_w^k) = IDSI(g_1, g_w)$  for every country  $c$  and product  $k$ .

### 3. Empirical Implications

The generalized model has four implications that can be taken to the data, first by comparing pairs of importers and second by comparing each importer to the world.

---

<sup>7</sup> Bils-Klenow (2001) call these slopes "Quality Engel Curves" and use household data to estimate how they differ across a set of durable goods.



When comparing pairs of importers, as the bilateral income dissimilarity index (equation (1.11)) rises

1. The number of common export partners falls.
2. The bilateral price dissimilarity index (1.12) for a particular commodity rises.

Both predictions contain related information. Suppose each quality is produced by only a single exporter, and a particular exporter produces quality  $z^*$ . Then two importers will buy from that exporter only if both have some population with income  $I(z^*)$  (where  $z^*$  is the optimal quality for income  $I(z^*)$ ). The smaller the overlap between two importer's income distribution, the fewer the exporters they will have in common. This is prediction one.

Prediction two is somewhat different and stronger. Again suppose that each quality is produced by only a single exporter. Prediction two suggests that the import *share* of an exporter producing quality  $z^*$  will be more similar for two importers who have similar population shares at income  $I(z^*)$ . It is stronger because it contains information both on which qualities of the differentiated product should have positive trade shares as well as the magnitude of the trade share.

However, unlike prediction one, prediction two focuses only on the distribution of prices as opposed to the names of the exporters. This is useful because it is conceivable that two exporters (e.g. Mexico, China) could have identical technology and therefore in equilibrium specialize in an identical spectrum of quality. In this case, our theory does not tell us from which exporter two importers (US, Japan) will buy quality  $z^*$ , only that they will buy from someone. That is, the US could buy from Mexico and Japan from China, and we would fail prediction one, but since Mexican and Chinese qualities (and prices) are identical, we would pass prediction two.

It is also useful to compare an importer to the rest of the world. When an importer's multilateral income dissimilarity index rises

1. The number of export partners in a particular commodity falls.
2. The multilateral price dissimilarity index rises.

The logic of these implications is similar to the bilateral case, we include them primarily because the literature on import variety and quality has not typically focused on bilateral pairs. The multilateral comparison has a more direct analogue in work that examines the level and growth of product variety across importers.

Caveat: Domestic Sales

Proposition 1 relates the difference in the price distributions between countries 1 and 2 to the difference in their income distributions. It implies that if the income distribution of country 1 is dis-similar to 2, the price distribution of country 1 is also dis-similar to 2. Thus if we run the following regression across country pairs for the differentiated good:

$$PDSI(h_c, h_{c'}) = \alpha_0 + \beta IDSI(g_c, g_{c'}) + e_{cc'} \quad (1.18)$$

we will get  $\alpha_0 = 0$  and  $\beta = 1$ .

There is a subtlety in implementing regression (1.18). Each quality  $z$  is supplied by a single country and so the supplier country does not import  $z$  from abroad but buys it from itself. However, we do not observe this portion of the supplier country's price distribution in our data because our data covers imports but not domestic sales. Furthermore, some quality levels might be non-traded (i.e. country  $c$  supplies certain quality levels to itself and no other country). Both imply that we measure the price distributions with error and may obtain biased estimate for  $\beta$ . However, the presence of non-traded qualities compresses the price distribution we observe into a narrower range than the true price distribution, and this reduces the price dis-similarity index and so biases our  $\beta$  estimate downward. On the other hand, thanks to Assumption 1, as the number of countries,  $C$ , gets larger, each country supplies a narrower range of traded goods, and so the bias caused by not observing the domestic sales of traded goods gets smaller. Therefore:

**Proposition 2** Assume that trade costs are zero. When the number of countries,  $C$ , is large, the estimated  $\beta$  in regression (1.18) using our data is smaller than the true  $\beta$ .

Proposition 2 implies that we expect to have  $\beta > 0$  in regression (1.18).

It is useful to contrast model predictions with the predictions of a Melitz (2003) style model of horizontal differentiation featuring firm-level heterogeneity in productivity plus fixed costs of trade. The Melitz model is a plausible alternative explanation for growth in import variety and import price dispersion. With fixed costs of trade only a subset of high productivity (low cost) varieties can sell a sufficiently high quantity to earn non-negative profits from exporting. As trade costs drop over time, as they have in the US import market, the number of imported varieties increases. Firm heterogeneity in productivity also implies heterogeneity in

prices at each point in time. Further, since growth in import variety occurs by adding lower productivity firms at the extensive margin, variety growth will also lead to increased price dispersion for imports.

The problem with a horizontal differentiation explanation for these facts is that high import prices in this model are due to low productivity and high marginal costs of production, which would appear to contradict Schott's (2004) finding that high US import prices originate in high income (and presumably, high productivity) exporters. It also follows from the logic of the Melitz model that these high cost firms would have a relatively low share of the market, in contrast to Hallak's (2003) findings that high export prices are correlated with large market shares. Finally, Melitz, like all homothetic CES utility models allows no role for the within country income distribution.

Still, the market size effects suggested by this framework suggest additional explanatory variables. We include in our regressions measures of market size and trade costs (for the multilateral regressions) or relative market size and distance (for the bilateral regressions). This gives us four estimating equations.

For the bilateral comparisons we have

$$\ln(N_{cc't}^k) = \alpha_{cc'}^k + \beta_1 \ln(IDSI_{cc't}) + \beta_2 \left| \ln\left(\frac{GDP_{ct}}{GDP_{c't}}\right) \right| + \beta_3 \ln DIST_{cc't} + \varepsilon_{cc't}^k \quad (1.19)$$

where the dependent variable  $N_{cc't}^k$  is the number of common export partners between  $c$  and  $c'$ .

$$\log(PDSI_{cc't}^k) = \alpha_{cc'k} + \beta_1 \log(IDSI_{cc't}) + \beta_2 \left| \log\left(\frac{GDP_{ct}}{GDP_{c't}}\right) \right| + \beta_3 \ln DIST_{cc't} + \varepsilon_{cc't}^k \quad (1.20)$$

For the comparisons to the world as a whole we have

$$\ln(N_{ct}^k) = \alpha_c^k + \beta_1 \ln(IDSI_{cwt}) + \beta_2 \ln(GDP_{ct}) + \beta_3 \ln MP_{ct} + \varepsilon_{ct}^k \quad (1.21)$$

Where the dependent variable is the number of export partners from which importer  $c$  buys product  $k$ , and  $MP$  is the "market potential" of  $c$ , that is, the GDP weighted average of distance from each export market.

$$\ln(PDSI_{cwt}^k) = \alpha_c^k + \beta_1 \ln(IDSI_{cwt}) + \beta_2 \ln(GDP_{ct}) + \beta_3 \ln MP_{ct} + \varepsilon_{ct}^k \quad (1.22)$$

We implement the first two regressions both in cross-sections for each wave, and in panel including country-pair x product fixed effect (for bilateral comparisons). For the other two, we implement regressions only in cross-sections for each wave.<sup>8</sup>

### 3.1. Income Data

Cross-country and inter-temporal comparisons in income distribution are limited because internationally comparable data are not easy to obtain. We employ data from the Luxembourg Income Study (LIS). LIS data are a compilation of the income survey data files of several countries, made comparable by rearranging/reclassifying income measures from national household budget surveys. Another widely used dataset of income distributions is Deininger and Squire (1996) and its extensions by the World Bank (DSWB).<sup>9</sup> We employ the LIS data because: 1. the LIS provides *percentile* level income while the DSWB provides *quintile* level income shares; and 2. the LIS is more consistent and better suited for cross-country and cross-time comparisons of income distribution (Atkinson and Brandolini 2001, Deaton 2003). The tradeoff is that the LIS covers a smaller number of countries than the DSWB data. However, because our theory is defined in terms of pair-wise comparisons of countries we can still generate considerable cross-sectional variation in income comparisons.

LIS provides standardized measures of household income for a set of 30 countries at roughly 5 year intervals for the period 1979-2001 (Wave 1 - Wave 5)<sup>10</sup>. Some countries have data missing from one or two of the five waves.<sup>11</sup> The literature has shown that percentile income levels tend to follow smooth trends (e.g. Dollar and

---

<sup>8</sup> We do not run panel regressions for (1.21) and (1.22) since we are not expected to obtain any meaningful results. If we apply country x product fixed effect for comparisons to the world, almost all the variations are absorbed through these dummies. The reason is that a country's income/price positions in the world (multilateral PDSI/ IDSI) are very stable over time, and it leaves no room to capture the time-series variations significantly.

<sup>9</sup> For other income distribution data see Chen and Ravallion (2001), Bourguignon and Morrisson (2002) and Milanovic (2002).

<sup>10</sup> The wave years differ across countries. Wave 1 is around 1980, wave 2 is around 1985, ..., and wave 5 is around 2000. Before wave 1, historical databases were available for five countries (Canada, Germany, Sweden, the United Kingdom, and the United States). We call the historical databases wave 0 (around 1975).

<sup>11</sup> Ten countries are missing one year of data. They are Australia, Belgium, Finland, France, Italy, Luxemburg, Mexico, Netherlands, Poland and Spain. Three countries (Austria, Ireland and Switzerland) are missing two.

Kraay 2002, Sala-i-Martin 2005), which is consistent with the patterns we see in the LIS data. Accordingly, we use linear estimates of income trends by percentile to construct distributions for the missing waves.<sup>12</sup>

We employ the most commonly used measure of income employed in the analysis of income inequality, disposable household income (DPI). DPI is disposable monetary income after direct taxes and including transfer payments. Wealth is ignored except to the extent that it is represented by cash interest, rent, and dividends. The data do not provide a comprehensive measure of income, typically excluding much of capital gains, imputed rents, home production, and most in-kind income. Further, we ignore indirect taxes and the benefits from public spending such as those from health care, education, or most housing subsidies. Data are available in current year local currency values. We convert the data to constant year US dollars using the PPP data from Penn World Tables 6.1.

DPI data are available at the level of households rather than consumers. Since household sizes vary, and consumption needs vary by age, we adjust the income measure using an adult equivalence scale (AES). Total household income is divided by the number of equivalent adults in order to get a measure of household “equivalent” income. Buhmann et al. (1988) propose a succinct parametric approximation to equivalence scales which summarizes the wide range of scales in use:

$$\text{Adjusted Income (EY)} = \text{DPI} / \text{Household Size}^E .$$

The equivalence elasticity,  $E$ , is a parameter representing the economies of scale.  $E$  ranges from 0 (perfect economies of scale, no adjustment) to 1 (no economies of scale, per capita income). We employ the LIS Equivalence Scale ( $E=0.5$ ). It is the most commonly used method among researchers who study income inequality using the LIS database.<sup>13</sup>

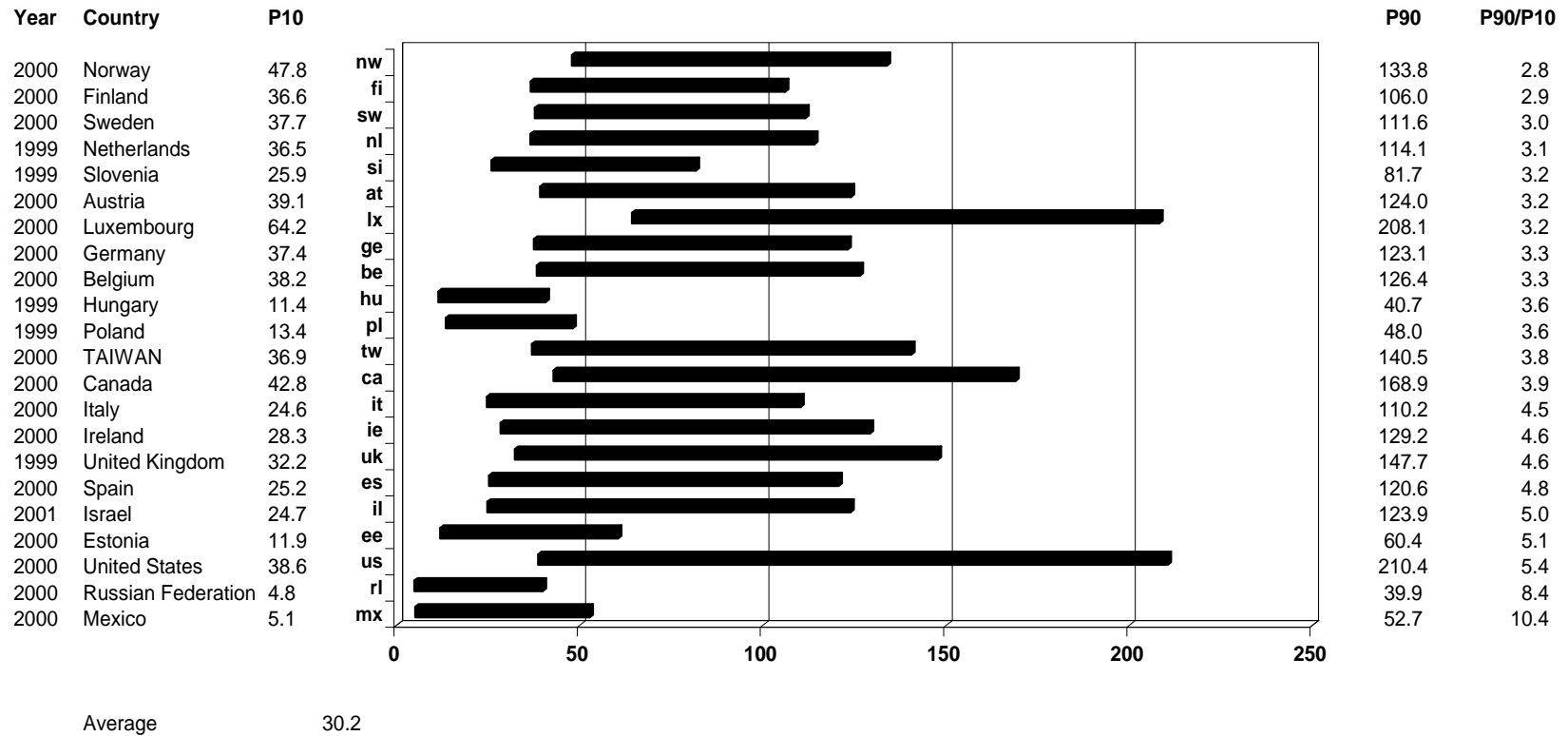
Figure 1.4 shows the income dispersion of the countries in Wave 5 (2000). For each country, income is measured relative to the US median income of \$24,094 (we set this median income to 100) and we plot the range of income starting at the 10<sup>th</sup>

---

<sup>12</sup> For each country we try both linear and log linear trends and then pick the trend with a higher  $R^2$ . For the 10 countries the average  $R^2$  is 0.89, and 5 countries have  $R^2$  higher than 0.95. For the 3 countries the average  $R^2$  is 0.92, and 2 countries have  $R^2$  higher than 0.95.

<sup>13</sup> An alternative popular approach explicitly employs data on the number of adults and children in the household. This approach is only feasible for a subset of our data.

percentile (P10) and ending at the 90<sup>th</sup> percentile (P90). An often-used measure of income dispersion is the 90/10 Decile Ratio (the Decile Ratio henceforth),  $P90/P10$ . We have arranged the countries in ascending order of their Decile Ratios. Norway has the least income dispersion with a Decile Ratio of 2.8 while the United States, Russia and Mexico have the most income dispersion with Decile Ratios of 5.4, 8.4 and 10.4. Thus there are large differences in income dispersion across countries in our data.



Notes: P10 and P90 numbers are the ratio of 10th percentile, 90th percentile PPP Income (in 2000 US\$) against US median PPP income. Data are arranged by ascending order of the decile ratio (P90/P10).

Source: Author's calculations using the LIS database and PWT6.1.

Figure 1.4 PPP Income Distribution Comparisons in Wave 5 (Around 2000)

While the Decile Ratio as a measure of income dispersion has some obvious appeal (e.g., insensitivity to top/bottom coding, ease of understanding), it has the disadvantage of focusing on only two data points in the distribution and so provides no information about how much of the world income distribution a country's income spans. This can be seen by comparing US and Mexico in Figure 1.4. The Decile Ratio for Mexico is nearly twice that of the US, but as the income levels for Mexico are lower than those for the US, Mexico's income distribution spans a much smaller range than the US and much of the income distributions for the US and Mexico does not overlap at all. The case involving US and Russia is similar.

In contrast, our measure of differences in the income distribution (equation (1.11)) employs data from all points in the income distribution and explicitly compares both the level and distribution of two countries income. When two countries distributions lie far away from each other (e.g. US and Mexico in Figure 1.4), the vertical distances between them are large at each point in the distribution. Two countries with identical income distributions have a bilateral dissimilarity index of 0. Countries with completely disjoint distributions (e.g. US and Russia in Figure 1.4) have a bilateral dissimilarity index of 1.

Finally, our theory requires us to construct and then compare income distributions across countries. To construct a continuous income distribution from the discrete income data we perform a non-parametric kernel estimation of its income distribution using the "kdensity" command in STATA (Deaton 1997). We use STATA's default kernel, the Epanechnikov, and STATA's default bandwidth,<sup>14</sup> and evaluate the density of the distribution at \$100 intervals from \$100 to \$150,000. The world income distribution of each wave is then constructed using equation (1.7).

### 3.2. Data on Import Prices and their Distribution

The trade data to test (1.19)-(1.22) come from the world trade flows database (the WTF) (Feenstra et. al., 2005) and the United Nations (UN) trade database. Since the LIS

---

<sup>14</sup> The choice of kernel tends to be relatively unimportant in practice (e.g. DiNardo and Tobias 2001) and STATA's default bandwidth is based on Silverman (1986)'s optimal bandwidth.



income distribution data exist for every five years from wave 0 (around 1975) to wave 5 (around 2000) we use import data for 1980, 1985, 1990, 1995, and 1999.<sup>15</sup>

Using these data it is straightforward to count the number of exporters from whom an importer has purchased a product in a given year, or the number of common exporters from whom two importers have purchased a good. The price data are more problematic. One, these data are riddled with measurement error. Two, the measurement error is likely to be importer-specific. In particular, the “prices” are really unit values (value/quantity) and the quantity units are unknown but are likely to be importer specific.<sup>16</sup> Third, quantity (but not value) data are missing from many of the importer’s reports.

Accordingly, we use a data cleaning procedure designed to extract exporter-specific signals from the noise of the raw data. We observe prices for some subset of importer  $c$  – exporter  $j$  pairs. We regress these on importer-product and exporter-product fixed effects, and bilateral distance to sweep out Alchian-Allen effects in pricing (Hummels-Skiba 2005).

$$\ln p_{cjt}^k = \alpha_{ct}^k + \alpha_{jt}^k + \beta_t^k \text{Distance}_{cj} + e_{cjt}^k \quad (1.23)$$

The importer-product fixed effects absorb variation in unit values arising from differences in units. We also experiment with a version in which we only control for importer-product and exporter-product fixed effects.

We use the exporter-product fixed effects as our measure of “true” export prices,  $\hat{p}_{jt}^k = \hat{\alpha}_{jt}^k$ . As Table 1.1 shows, these estimated prices are very highly correlated with exporter characteristics such as per capita income, and the capital-labor ratio, consistent with the findings in Schott (2003) and Hummels-Klenow (2005). This is true both before and after the estimation in (1.23), but the fitted prices have very high R2 (greater than 0.9). The estimated prices are also positively correlated with US imports price data.

---

<sup>15</sup> The WTF 2000 data has some technical problems, so we used 1999 data instead.

<sup>16</sup> See Hummels-Klenow (2005) for a discussion of this problem. Importers might report quantities in weight terms, either kg or pounds, while others use counts.

Table 1.1 Raw Unit Prices vs the Estimated Prices

	Dependent Var	
	$p_{ijk}$	phat
ln pgdp(j)	0.246 *** <i>0.034</i>	0.233 *** <i>0.034</i>
Constant	-0.971 *** <i>0.316</i>	-2.950 *** <i>0.320</i>
Number of Obs.	497,208	497,208
R <sup>2</sup>	0.48	0.94

Notes: This table reports the results of regressions of the log of raw / estimated(phat) unit export prices on exporter per capita GDP in Wave 5. Product fixed effects are applied and standard errors are adjusted for exporter clustering. Standard errors are in Italics.

\*\*\* refers to statistical significance at the 1 percent levels.

Data Sources: UN trade database, Feenstra et al. (2005), WDI 2005.

Table 1.2 shows the correlation coefficients between the US Census' unit imports prices and the estimated prices. The US Census imports data are at the HS 10-digit level. We matched HS 10-digit into SITC 5-digit level products using the concordance file. And the SITC 4-digit level raw unit prices are compiled using SITC 5-digit level imports values and quantities. There are two sets of estimated prices; one controlled for bilateral distance and the other not controlled for it. The correlation with raw unit prices is higher in the estimated prices without controlling for distances.

Table 1.2 Correlation between the US Import Prices and the Estimated Prices

1) Estimation of phat with distance		2) Estimation of phat without distance	
	Coefficients		Coefficients
All Waves	0.1636	All Waves	0.3307
Wave 1	0.1057	Wave 1	0.1927
Wave 2	0.1502	Wave 2	0.2938
Wave 3	0.1917	Wave 3	0.3396
Wave 4	0.1646	Wave 4	0.3332
Wave 5	0.1725	Wave 5	0.3512

With these estimated prices in hand we can now construct the desired price distribution for each importer. For each importer  $c \times$  product  $k$  it is possible to have trade

with every exporter  $j$  that ships to at least one importer. We weigh the prices  $\hat{p}_{jt}^k$  by the share  $s_{cjt}^k$  of that exporter in importer  $c$ 's trade. This includes zero shares. Then using that discrete distribution we estimate a nonparametric kernel to smooth the price distribution. Finally, we take differences in price distributions (either relative to a particular importer  $c$  or to the world) following equation (1.12)

#### 4. Empirical Results

As Tables 1.3 – 1.6 show, the coefficients of the log of the income dissimilarity index ( $\beta_1$ ) have the expected signs and are significant for regressions (1.19) ~ (1.22).

First, table 1.3 reports the results of regression (1.19) that looks at the correlation between the bilateral income dissimilarity and the number of common export partners. Since the set of countries whose LIS income data are available varies from wave to wave and we exploit only the cross-section variations, we use the full sample of countries in columns (1) to (5). For the panel regression, however, we use the (smaller and more homogeneous) sample of countries whose data are available for waves 1~5. We see in columns (1) to (5) that the coefficients of the log of the income dissimilarity indices are all negative and significant. It matches our expectation. At a given point in time, if the income distribution of the pairs of importers is more dissimilar from each other, the number of common export partners is smaller. The results on GDPgap show mixed results across waves, while the coefficients of the log of distance are all positive and significant for all waves. When the pairs of countries are located farther away, the number of common imports products becomes smaller. Column (6) is the result of the panel regression. It shows that the number of common export partners is negatively and significantly related to the bilateral income dissimilarity index. That is to say, over time, as the income distribution of the pairs of countries becomes more dissimilar to each other, the number of common export partners that these countries source a given product becomes smaller. Again this is consistent with our theory.

Table 1.4 reports the results of regression (1.20) that looks at the correlation between the bilateral income and bilateral price dissimilarity. Again, the first five

columns are cross-section variations by wave and the last column is the time-series variations using the panel data. We see in columns (1) to (5) that the coefficients of the log of the income dissimilarity indices are all positive and significant. It is consistent with our theory. That is to say, at a given point in time, if the incomes of the pairs of importers are more dissimilar from each other, the prices of a particular importing commodity are more dissimilar. Furthermore, it also implies that the import share of a product  $k$  is more dissimilar for two importers who have dissimilar population share at income  $I(k)$  at which level optimal quality choice is  $k$ . The coefficients of the GDPgap and the log of distance are all positive and significant for all waves. The import price distributions of two countries are more dissimilar when two countries' are more differ in country size and farther away from each other. The result in column (6) shows that the bilateral price dissimilarity index is positively and significantly related to the bilateral income dissimilarity index. It implies that, over time, as the income distribution of the pairs of countries becomes more dissimilar to each other, the imports price distribution of a given commodity for these countries becomes also dissimilar. The panel regression results in tables (1.3) and (1.4) can explain the fact that growth in both import variety and import price dispersion has occurred at the same time that the US income distribution has significantly widened.

Table 1.5 reports the results of regression (1.21) that looks at the multilateral income dissimilarity and the number of common export partners. The results are very similar to table 1.3. We see in columns (1) to (5) that the coefficients of the log of the income dissimilarity indices are all negative and significant. At a given point in time, if the income distribution of country  $c$  is more dissimilar to the world income distribution, this country sources a given product from a smaller number of exporters. The coefficient on the log GDP is positive and significant, which means that a bigger country sources a given product from a larger number of exporters. The results on the log of the market potential give mixed results across waves. Finally, table 1.6 reports the results of regression (1.22) that looks at the correlation between multilateral income dissimilarity and multilateral price dissimilarity. We see the results very similar to table 1.4. By wave cross-section regressions in columns (1) to (5) give the positive and significant

coefficients on multilateral IDSI. It suggests that if a country's income distribution is more dissimilar to the world income distribution, this country's imports price distribution is also more dissimilar to the world price distribution for a given product. The coefficients on the log GDP is all negative and significant, which means that a bigger country's imports demand price distribution is more similar to the world price distribution. On the other hand, the coefficient on the log of the market potential is positive (Our theory does not tell about the sign of the coefficient of the market potential).

Table 1.3 Bilateral Income Similarity and Common Export Partners

	Dependent Var: $\ln N(c,c')$					
	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Panel
$\ln \text{IDSI}(c,c')$	-0.137 *** <i>0.003</i>	-0.150 *** <i>0.004</i>	-0.240 *** <i>0.002</i>	-0.271 *** <i>0.002</i>	-0.128 *** <i>0.002</i>	-0.045 *** <i>0.002</i>
$\text{GDPgap}(c,c')$	-0.016 *** <i>0.002</i>	-0.0005 <i>0.002</i>	0.047 *** <i>0.001</i>	-0.027 *** <i>0.001</i>	0.002 <i>0.001</i>	0.223 *** <i>0.009</i>
$\ln \text{Dist}(c,c')$	-0.106 *** <i>0.001</i>	-0.032 *** <i>0.002</i>	-0.080 *** <i>0.001</i>	-0.035 *** <i>0.001</i>	-0.055 *** <i>0.001</i>	-0.753 *** <i>0.023</i>
Constant	1.262 *** <i>0.015</i>	0.813 *** <i>0.015</i>	1.295 *** <i>0.011</i>	1.104 *** <i>0.010</i>	0.799 *** <i>0.010</i>	5.282 *** <i>0.158</i>
Number of Obs.	62,897	78,173	111,357	171,585	162,219	337,954
R <sup>2</sup>	0.16	0.11	0.20	0.17	0.10	0.38

Notes: This table reports the results of regression (1.19). All variables are in natural logs,  $N(c,c')$  is the number of common exporter partners between importer  $c$  and  $c'$ ,  $\text{IDSI}(c,c')$  is the Income Dis-Similarity Index,  $\text{GDPgap}(c,c')$  is the absolute value of the  $\ln(\text{GDP})$  differences (i.e.,  $\text{ABS}\{\ln(\text{GDP}_c/\text{GDP}_{c'})\}$ ), and  $\text{Dist}(c,c')$  is the physical distance (kilometers). Product dummies are applied in each wave regression, and importer-pair and product dummies are applied in the panel regression. The panel uses 15 common countries from wave 1 to wave 5. Standard errors are in Italics. \*\*\*, \*\*, \* refers to statistical significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Data Sources: UN trade database, Feenstra et al. (2005), LIS, World Bank, PWT6.1.

Table 1.4 Bilateral Income and Price Similarity

	Dependent Var: $\ln \text{PDSI}(c,c')$					
	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Panel
$\ln \text{IDSI}(c,c')$	0.040 *** <i>0.004</i>	0.034 *** <i>0.005</i>	0.087 *** <i>0.003</i>	0.065 *** <i>0.002</i>	0.031 *** <i>0.002</i>	0.010 *** <i>0.003</i>
$\text{GDPgap}(c,c')$	0.051 *** <i>0.002</i>	0.058 *** <i>0.002</i>	0.044 *** <i>0.002</i>	0.036 *** <i>0.001</i>	0.020 *** <i>0.001</i>	-0.059 *** <i>0.013</i>
$\ln \text{Dist}(c,c')$	0.089 *** <i>0.002</i>	0.065 *** <i>0.003</i>	0.065 *** <i>0.002</i>	0.056 *** <i>0.001</i>	0.067 *** <i>0.001</i>	0.405 *** <i>0.034</i>
Constant	2.748 *** <i>0.020</i>	2.877 *** <i>0.019</i>	2.750 *** <i>0.013</i>	2.904 *** <i>0.011</i>	2.940 *** <i>0.011</i>	0.736 *** <i>0.234</i>
Number of Obs.	61,574	77,903	112,361	175,238	165,164	336,007
R <sup>2</sup>	0.23	0.23	0.21	0.20	0.22	0.20

Notes: This table reports the results of regression (1.20). All variables are in natural logs,  $\text{PDSI}(c,c')$  is the Price Dis-Similarity Index between importer  $c$  and  $c'$ ,  $\text{IDSI}(c,c')$  is the Income Dis-Similarity Index,  $\text{GDPgap}(c,c')$  is the absolute value of the  $\ln(\text{GDP})$  differences (i.e.,  $\text{ABS}\{\ln(\text{GDP}_c/\text{GDP}_{c'})\}$ ), and  $\text{Dist}(c,c')$  is the physical distance (kilometers). Product dummies are applied in each wave regression, and importer-pair and product dummies are applied in the panel regression. The panel uses 15 common countries from wave 1 to wave 5. Standard errors are in Italics. \*\*\*, \*\*, \* refers to statistical significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Data Sources: UN trade database, Feenstra et al. (2005), LIS, World Bank, PWT6.1.

Table 1.5 Multilateral Income Similarity and the Number of Export Partners

	Dependent Var: ln N (c)				
	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
ln IDSI(c,W)	-0.239 *** <i>0.018</i>	-0.364 *** <i>0.014</i>	-0.347 *** <i>0.016</i>	-0.184 *** <i>0.015</i>	-0.065 *** <i>0.013</i>
ln GDP(c)	0.311 *** <i>0.013</i>	0.389 *** <i>0.013</i>	0.452 *** <i>0.013</i>	0.236 *** <i>0.008</i>	0.291 *** <i>0.008</i>
ln MP(c)	-0.144 *** <i>0.012</i>	-0.155 *** <i>0.013</i>	-0.202 *** <i>0.013</i>	0.026 *** <i>0.007</i>	-0.005 <i>0.007</i>
Constant	-3.663 *** <i>0.163</i>	-5.179 *** <i>0.126</i>	-5.544 *** <i>0.120</i>	-6.445 *** <i>0.110</i>	-7.513 *** <i>0.098</i>
Number of Obs.	9,385	14,048	14,986	17,706	17,014
R <sup>2</sup>	0.44	0.43	0.47	0.51	0.55

Notes: This table reports the results of regression (1.21). All variables are in natural logs, N(c) is the number of c's exporter partners, IDSI(c,W) is the Income Dis-Similarity Index of c against the world, and MP(c) is the market potential of c. Product dummies are applied in each wave regression.

Standard errors are in Italics. \*\*\*, \*\*, \* refers to statistical significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Data Sources: UN trade database, Feenstra et al. (2005), LIS, World Bank, PWT6.1.



Table 1.6 Multilateral Income and Price Similarity

	Dependent Var: ln PDSI(c,W)				
	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
ln IDSI(c,W)	0.118 *** <i>0.018</i>	0.165 *** <i>0.013</i>	0.181 *** <i>0.014</i>	0.097 *** <i>0.012</i>	0.058 *** <i>0.011</i>
ln GDP(c)	-0.125 *** <i>0.013</i>	-0.152 *** <i>0.012</i>	-0.149 *** <i>0.011</i>	-0.079 *** <i>0.006</i>	-0.100 *** <i>0.007</i>
ln MP(c)	0.080 *** <i>0.012</i>	0.071 *** <i>0.012</i>	0.074 *** <i>0.011</i>	0.003 <i>0.006</i>	0.014 ** <i>0.006</i>
Constant	4.303 *** <i>0.163</i>	5.094 *** <i>0.117</i>	4.833 *** <i>0.104</i>	5.151 *** <i>0.093</i>	5.563 *** <i>0.083</i>
Number of Obs.	8,994	13,432	14,430	17,027	16,367
R <sup>2</sup>	0.32	0.41	0.37	0.37	0.38

Notes: This table reports the results of regression (1.22). All variables are in natural logs, PDSI(c,W) is the Price Dis-Similarity Index of c against the world, IDSI(c,W) is the Income Dis-Similarity Index, and MP(c) is the market potential of c. Product dummies are applied in each wave regression.

Standard errors are in Italics. \*\*\*, \*\*, \* refers to statistical significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Data Sources: UN trade database, Feenstra et al. (2005), LIS, World Bank, PWT6.1.

## 5. Conclusions

In this paper, we investigate the relationship between quality demand and income distribution. We re-interpret Flam and Helpman's model of quality differentiation in terms of distributions of incomes and prices. We derive predictions that relate the number and price distribution of imported varieties to an importer's income distribution relative to the world or to other importers.

To test these predictions we employ microdata on income from household surveys for many countries over a 20 year span to construct income distributions within and across countries. We show that pairs of importers whose income distributions look more similar have more export partners in common and a more similar import price distribution. Similarly, importers whose income distribution looks more like the world buy from more exporters and have an import price distribution that looks more like the world.

## References

- Atkinson, Anthony B. and Andrea Brandolini, 2001. "Promise and Pitfalls in the Use of 'Secondary' Data-Sets: Income Inequality in OECD Countries as a Case Study", *Journal of Economic Literature*, 39 (3), 771-799.
- Atkinson, A.B., Lee Rainwater, and Timothy Smeeding, 1995, *Income Distribution in OECD countries: Evidence from the Luxembourg Income Study*, *Social Policy Studies* No. 18, Paris, France: OECD.
- Bils, Mark and Peter J. Klenow, 2001, Quantifying Quality Growth, *American Economic Review*, 91(4), pp. 1006-2001.
- Bourguignon, Francois and Christian Morrisson, 2002. "Inequality Among World Citizens: 1820-1992", *American Economic Review*, 92(4), 727-744.
- Buhmann, Brigitte et al., June 1988, Equivalence Scales, Well-Being, Inequality, and Poverty: Sensitivity Estimates across Ten Countries Using the Luxembourg Income Study (LIS) Database, *Rev. Income Wealth*, 34(2), pp. 115-42.
- Chen, Shaohua and Martin Ravallion, 2001. "How Did the World's Poorest Fare in the 1990s?", *Review of Income and Wealth*, 47(3), 283-300.
- Dalgin, M., D. Mitra, and V. Trindade, Inequality, September 2004, Nonhomothetic Preferences, and Trade: A Gravity Approach, *NBER Working Paper* No. 10800, pp. 1-34.
- Deaton, Angus, 2003. "Health, Inequality, and Economic Development", *Journal of Economic Literature*, 41(1), 113-158.
- Deaton, A., 1997, *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*, The Johns Hopkins University Press, Baltimore.
- Deaton, A. and J. Muellbauer, 1986, On Measuring Child Costs: With Applications to Poor Countries, *Journal of Political Economy*, 94, pp. 720-744.
- Deaton, A., J. Ruiz-Castillo, and D. Thomas, 1989, The Influence of Household Composition on Household Expenditure Patterns: Theory and Spanish Evidence, *Journal of Political Economy*, 97, pp. 179-200.

- Deininger, Klaus and Lyn Squire, 1996. "A New Data Set Measuring Income Inequality". *World Bank Economic Review*, 10, 565-591.
- DiNardo, John and Justin L. Tobias, 2001. "Nonparametric Density and Regression Estimation", *Journal of Economic Perspectives*, 15(4), 11-28.
- Dollar, David and Kraay, Aart, 2002. "Growth is Good for the Poor", *Journal of Economic Growth*, 7, 195-225.
- Dornbusch, R., S. Fishcher, and P. A. Samuelson, 1977, Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods, *American Economic Review*, 67(5), pp. 823-839.
- Falvey, R. E., 1981, Commercial Policy and Intra Industry Trade, *Journal of International Economics* 11, pp. 495-511.
- Falvey, R. E. and H. Kierzkowski, 1987, Product Quality, Intra-Industry Trade and (Im)perfect Competition, in Kierzkowski, H. (ed.), *Protection and Competition in International Trade: Essays in Honor of W. M. Corden*, Basil Blackwell, Oxford and London.
- Feenstra, R. C., R. E. Lipsey, H. Deng, A. C. Ma and H. Mo, 2005, World Trade Flows, 1962-2000, *NBER Working Paper* No. 11040.
- Flam, H. and E. Helpman, 1987, Vertical Product Differentiation and North-South Trade, *American Economic Review*, 77(5), pp. 810-822.
- Funk, Peter, 1998. "Satiation and Underdevelopment", *Journal of Development Economics* 57, 319-341.
- Gottschalk, P. and T.M. Smeeding, 1997, Cross-National Comparisons of Earnings and Income Inequality, *Journal of Economic Literature*, XXXV(June), pp. 633-687.
- Hallak, J. C., 2003, The Effect of Cross-Country Differences in Product Quality on the Direction of International Trade, *EIIT 10<sup>th</sup> Annual Conference Paper*.
- Hallak, J. C., 2004, Product Quality, Linder, and the Direction of International Trade, *NBER Working Paper*, No. 10877.

- Helpman, E. and P. Krugman, 1985, *Market Structure and Foreign Trade*, MIT Press, Cambridge (Mass.).
- Heston, Alan, Robert Summers and Bettina Aten, October 2002, Penn World Table Version 6.1, Center for International Comparisons at the University of Pennsylvania (CICUP).
- Hummels, D. and P. J. Klenow, 2002, The Variety and Quality of a Nation's Trade, NBER Working Paper No.w8712, pp. 1-32.
- Hummels, D. and P. J. Klenow, 2005, The Variety and Quality of a Nation's Trade, *American Economic Review* Vol. 95, No 3, pp. 704-723.
- Hunter, Linda, 1991. "The Contribution of Nonhomothetic Preferences to Trade", *Journal of International Economics*, 30 (3-4), 345-358.
- Johnson, D. and T. Smeeding, 1997, Measuring the Trends in Inequality of Individuals and Families: Income and Consumption, Paper presented to the American Economic Association, January 6.
- Khandelwal, A., 2005, Product Quality and Competition in International Trade, *EIIT 12<sup>th</sup> Annual Conference Paper*.
- Krugman, Paul R., 1979, Increasing Returns, Monopolistic Competition, and International Trade, *Journal of International Economics* 9(4), pp. 469-479.
- Markusen, James, 1986. "Explaining the Volume of Trade: An Eclectic Approach", *American Economic Review* 76(5), 1002-1011.
- Matsuyama, K., 2000, A Ricardian Model with a Continuum of Goods under Nonhomothetic Preferences: Demand Complementarities, Income Distribution, and North-South Trade, *Journal of Political Economy*, 108(6), pp. 1093-1120.
- Matsuyama, Kiminori, 2002. "The Rise of Mass Consumption Societies", *Journal of Political Economy*, 110(5), 1035-1070.
- Melitz, Marc J., November 2003, The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity, *Econometrica* Vol. 71, No. 6, pp. 1695-1725.
- Milanovic, Branko, 2002. "True World Income Distribution, 1988 and 1993: First

- Calculation Based on Household Surveys Alone”, *Economic Journal*, 112, 51-92.
- Mitra, Devashish and Vitor Trindade, 2005. “Inequality and Trade”, *Canadian Journal of Economics* 38(4), 1253-1271.
- Muellbauer, J., 1977, Testing the Barten model of Household Composition Effects and the Cost of Children, *Economic Journal*, 87, pp. 460~487.
- Murphy, K. M. and A. Shleifer, 1997, Quality and Trade, *Journal of Development Economics*, 53, pp. 1-15.
- Phipps, S. A., 1998, What Is the Income “Cost of a Child”? Exact Equivalence Scales for Canadian Two-Parent Families, *Review of Economics and Statistics* 80, (1), pp. 157-64.
- Rainwater, L., 1974, *What Money Buys: Inequality and the Social Meanings of Income*. New York: Basic Books.
- Sala-i-Martin, Xavier, 2005. “The World Distribution of Income: Falling Poverty and ... Convergence, Period”, forthcoming, *Quarterly Journal of Economics*.
- Schott, P. K., 2004, Across-Product versus Within-Product Specialization in International Trade. *Quarterly Journal of Economics*, Vol. 119, Issue 2, pp. 647-678.
- Silverman, B.W., 1986. “Density Estimation for Statistics and Data Analysis”, London: Chapman and Hall.
- Smeeding, T. M., 2000, Changing Income Inequality in OECD Countries: Updated Results from the Luxembourg Income Study (LIS), *Luxembourg Income Study Working Paper* No. 252.
- Stokey, N. L., 1991, The Volume and Composition of Trade Between Rich and Poor Countries, *Review of Economic Studies*, 58(1), pp. 63-80.

## Theory Appendix

### 1. Proof of Proposition 1

The income dis-similarity index in equation (1.10) can be re-written as  $\text{IDSI}(\cdot) = \frac{1}{2} \sum_{j=1}^C \int_{\underline{G}_j} |g_1(\cdot) - g_2(\cdot)| dI$ . Since  $f_{pj}(I) = a_j I$  is strictly increasing in  $\underline{G}_j$ ,  $\int_{\underline{G}_j} |g_1(\cdot) - g_2(\cdot)| dI = \int_{\underline{H}_j} |g_1(\frac{p(z)}{a_j}) - g_2(\frac{p(z)}{a_j})| dp(z)$  for all  $j$ . Thus  $\text{IDSI}(\cdot) = \frac{1}{2} \sum_{j=1}^C \int_{\underline{H}_j} |g_1(\frac{p(z)}{a_j}) - g_2(\frac{p(z)}{a_j})| \frac{1}{a_j} dp(z) = \text{PDSI}(\cdot)$ .

Suppose  $f_{pj}(\cdot)$  takes a more general form than  $a_j I$ , but  $f_{pj}(\cdot)$  is strictly increasing in  $G_j$ , its inverse exists and is differentiable in  $G_j$ . Then  $\text{PDSI} = \frac{1}{2} \sum_{j=1}^C \int_{\underline{H}_j} |g_1(f_{pj}^{-1}(p(z))) - g_2(f_{pj}^{-1}(p(z)))| [f_{pj}^{-1}(p(z))]' dp(z)$ . The same logic as above goes through and  $\text{PDSI}(\cdot) = \text{IDSI}(\cdot)$ .

### 2. In the multi-country case, higher qualities command higher prices and consumers with higher income consume higher qualities

Suppose the optimal quality choices for consumers with income levels  $I_1$  and  $I_0$  are  $z_1$  and  $z_0$ ; we show that (a) if  $z_1 > z_0$ ,  $p(z_1) > p(z_0)$  and (b) if  $I_1 > I_0$ ,  $z_1 > z_0$ .

(a) is easy to show. Suppose  $p(z_1) < p(z_0)$ . Then  $z_1$  not only has higher quality but also lower price, and so the consumers with income  $I_0$  will buy  $z_1$  rather than  $z_0$ , and this is a contradiction. Thus higher qualities command higher prices.

To show (b), note first that consumer  $I_1$  can afford  $z_0$  and so  $z_1 \geq z_0$ . Now suppose that  $z_0$  is supplied by country  $c_0$ . Let  $z_1^*$  be consumer  $I_1$ 's optimal choice of quality under the constraint that all the qualities higher than  $z_0$  are produced in  $c_0$ . By equation (4)  $z_1^* > z_0$ . Is quality  $z_1^*$  actually produced in country  $c_0$ ? If yes,  $z_1 = z_1^*$  and we are home. If no, it must be because another country, say  $c_1$ , can produce  $z_1^*$  at a lower marginal cost than  $c_0$ . Then consumer  $I_1$  can still afford to buy  $z_1^*$  and  $z_1^* < z_1$  ( $z_1$  is consumer  $I_1$ 's optimal unconstrained quality choice). Thus  $z_0 < z_1^* < z_1$  and again we are home. Thus consumers with higher income consume higher qualities.

### 3. Proof of Proposition 2:

Let  $h_1'(\cdot)$  denote the price distribution of country 1 that we observe in our data. Suppose country 1 supplies the set of prices  $H_1^S$  (there is a one-to-one mapping between prices and qualities). Let the probability “mass” of  $H_1^S$  be  $P(H_1^S) = \int_{H_1^S} h_1(p(z)) dp(z)$

where  $h_1(\cdot)$  is as defined in equation (1.10). Then  $h_1'(p(z)) = \frac{1}{1 - P(H_1^S)} h_1(p(z))$  if  $p(z) \in$

$H_1 - H_1^S$  (this is when country 1 buys the quality with price  $p(z)$ , quality  $p(z)$  henceforth, from abroad) and  $h_1'(p(z)) = 0$  if  $p(z) \in H_1^S$  (this is when country 1 buys quality  $p(z)$  from itself).

Next, consider the various components of the set  $H_1$ .

(a) The set  $H_1 - H_1^S$  is a subset of the set  $\cup_d H_d^S$  where  $d = 2, \dots, C$  (i.e. when country 1 buys quality  $p(z)$  from abroad, it must buy it from some other country). Consider the set  $B \equiv (H_1 - H_1^S) \cap H_2^S$ . Country 1's price distribution is not missing from our data and

$h_1'(\cdot) = \frac{1}{1 - P(H_1^S)} h_1(\cdot)$ ; but country 2's price distribution is missing from our data and so

$h_2'(\cdot) = 0$  and  $|h_1'(\cdot) - h_2'(\cdot)| = \left| \frac{1}{1 - P(H_1^S)} h_1(\cdot) \right|$ . This differs from the true value  $|h_1(\cdot) -$

$h_2(\cdot)|$ . However, by Assumption 1:

$$\int_B |h_1'(\cdot) - h_2'(\cdot)| dp(z) \rightarrow 0 \text{ as } C \rightarrow +\infty \quad (\text{A1})$$

On the other hand, for the rest of the set  $H_1 - H_1^S$ , we observe country 2's price distribution as well and so  $h_2'(\cdot) = h_2(\cdot)$ . Thus over the set  $H_1 - H_1^S$ , we observe:

$$\begin{aligned} \int_{H_1 - H_1^S} |h_1'(\cdot) - h_2'(\cdot)| dp(z) &= \\ &= \int_{H_1 - H_1^S - B} \left| \frac{1}{1 - P(H_1^S)} h_1(\cdot) - h_2(\cdot) \right| dp(z) + \int_B |h_1'(\cdot) - h_2'(\cdot)| dp(z) \\ &\approx \int_{H_1 - H_1^S} \left| \frac{1}{1 - P(H_1^S)} h_1(\cdot) - h_2(\cdot) \right| dp(z) \end{aligned} \quad (\text{A2})$$

Where the  $\approx$  is by equation (A1).

(b) The set  $H_1^S$  contains two kinds of qualities. The first kind is not traded: country 1



supplies them to itself and no other country. Let  $H_1^D$  be the collection of them. Because the existence of  $H_1^D$  is driven by demand and not by trade cost (we assume zero trade cost),  $H_1^S \cap (\cup_d H_d^S) = \emptyset$ , where  $d \neq c$  (i.e. the set  $H_1^D$  is unique to country  $c$  and not demanded by any other country). The second kind of qualities in the set  $H_1^S$  is traded: country 1 supplies them to itself and all the other countries (recall that each quality level  $z$  is supplied by only a single country). Let  $H_1^X$  be the collection of them. Then  $H_1^S = H_1^D \cup H_1^X$  and the probability “mass” of  $H_1^S$  can be decomposed as:  $P(H_1^S) = P(H_1^D) + P(H_1^X)$ .

(b1) For the set  $H_1^X$ , as  $C$  gets large, by Assumption 1, each country supplies a smaller and smaller segment of the quality spectrum. Thus  $P(H_1^X) \rightarrow 0$ ,  $P(H_1^S) \approx P(H_1^D)$  and:

$$\int_{H_1^X} |h_1'(\cdot) - h_2'(\cdot)| dp(z) \rightarrow 0 \text{ as } C \rightarrow +\infty \quad (\text{A3})$$

Equation (A2) becomes:

$$\int_{H_1 - H_1^S} |h_1'(\cdot) - h_2'(\cdot)| dp(z) \approx \int_{H_1 - H_1^S} \frac{1}{1 - P(H_1^D)} |h_1(\cdot) - h_2(\cdot)| dp(z) \quad (\text{A4})$$

(b2) For the set  $H_1^D$ , the observed distributions  $h_2'(\cdot)$  and  $h_1'(\cdot)$  are both 0. The true distribution  $h_2(\cdot)$  is also 0. Thus:

$$\begin{aligned} \int_{H_1^D} |h_1'(\cdot) - h_2'(\cdot)| dp(z) &= 0 \\ \int_{H_1^D} |h_1(\cdot) - h_2(\cdot)| dp(z) &= \int_{H_1^D} h_1(\cdot) dp(z) = P(H_1^D) \end{aligned} \quad (\text{A5})$$

(c) Therefore, for the set  $H_1$ , by equations (A3) and (A4), we observe:

$$D \equiv \int_{H_1} |h_1'(\cdot) - h_2'(\cdot)| dp(z) = \int_{H_1 - H_1^S} \frac{1}{1 - P(H_1^D)} |h_1(\cdot) - h_2(\cdot)| dp(z) \quad (\text{A6})$$

whereas the true value is:

$$D^* \equiv \int_{H_1} |h_1(\cdot) - h_2(\cdot)| dp(z) = \int_{H_1 - H_1^D} |h_1(\cdot) - h_2(\cdot)| dp(z) + P(H_1^D) \quad (\text{A7})$$

Then:

$$D = \int_{H_1 - H_1^S} |h_1(\cdot) - h_2(\cdot)| dp(z) + \frac{P(H_1^D)}{1 - P(H_1^D)} \int_{H_1^D} |h_1(\cdot) - h_2(\cdot)| dp(z)$$

$$\begin{aligned}
&\leq \int_{H_1-H_1^s} |h_1(\cdot) - h_2(\cdot)| dp(z) + \frac{P(H_1^D)}{1-P(H_1^D)} \int_{H_1-H_1^s} h_1(\cdot) dp(z) \\
&= \int_{H_1-H_1^s} |h_1(\cdot) - h_2(\cdot)| dp(z) + P(H_1^D) = D^* \tag{A8}
\end{aligned}$$

Analogously, over the set  $H_2$ ,

$$\int_{H_2} |h_1'(\cdot) - h_2'(\cdot)| dp(z) \leq \int_{H_2} |h_1(\cdot) - h_2(\cdot)| dp(z)$$

Therefore, the bias caused by the existence of  $H_1^D$  implies that the dis-similarity index we observe in our data is smaller than the true dis-similarity index and so the estimate of  $\beta$  for regression (1.13) is biased towards 0.