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

Yo Chul Choi
David Hummels
Chong Xiang

Purdue University

August 2006

Abstract: We examine a generalized version of Flam and Helpman's (1987) model of vertical differentiation that maps cross-country differences in income distributions to variations in import variety and price distributions. The theoretical predictions are examined and confirmed using micro data on income from the Luxemburg Income Study for 30 countries over 20 years. The pairs of importers whose income distributions look more similar have more export partners in common and more similar import price distributions. Similarly, the importers whose income distributions look more like the world buy from more exporters and have import price distributions that look more like the world.

Keywords: quality differentiation; income distribution; variety; import price distribution; Luxemburg Income Study.

JEL classification: F1, D3.

Acknowledgements: The authors would like to thank Purdue CIBER and the NSF for financial support and seminar participants at the Midwest International Meetings, the Federal Reserve Boards of Governors and Purdue, Oregon and Oregon State Universities for helpful comments.

1. Introduction

There is a large literature examining how international trade affects a nation's income distribution, but there is relatively little empirical work examining the reverse channel. This is in large part because trade models commonly rule out income effects in order to focus attention on supply considerations such as factor endowments or scale economies. To the extent that richer demand structures with non-homothetic preferences are employed they operate at the level of broad industries, for example, allowing poor countries to devote relatively large income shares to commodity foodstuffs. In this paper we investigate how the distribution of income within and across countries shapes patterns of consumption and international trade in quality differentiated varieties within narrow product categories.

Our starting point is Flam and Helpman's (1987) model of quality differentiation in trade and we focus on the model's demand side implications linking consumer incomes to quality choice. As in Flam-Helpman, goods can be quality differentiated at some cost so that higher prices reflect higher quality, and consumers use marginal income to buy higher qualities rather than higher quantities of a differentiated good. This provides an equilibrium mapping in which prices of goods consumed are rising in household income.

This prediction is consistent with household evidence on consumer durables purchases. Bils and Klenow (2001) use survey data for the US that reports household income and purchase prices and estimate positive price-income slopes (or, "Quality Engel Curves"). Our interest lies in cross-country comparisons where household consumption choices are unobservable. We show that the model can be written in terms of national income and price distributions which are, with some effort, observable. We provide a theorem showing that the difference in two countries' price distributions for the quality differentiated good is equal to the difference in their income distributions. Put another way, countries with more similar income distributions have more similar product price distributions and import from a larger number of common exporters.

We also extend Flam and Helpman (1987) to the case of multiple differentiated goods and multiple countries with different technologies. In this case the quality-price relationship and price-income slopes vary endogenously across goods. However, our result linking cross-country differences in price distributions to differences in income distributions goes through precisely as before because the differencing removes cross-commodity variation in the quality-price and price-income relationships.

To examine our model's predictions we must first construct theoretically appropriate income and price distributions that are comparable within and across countries. This disqualifies conventional and easily obtained measures of income differences that exploit exclusively within country variation (such as Gini coefficients or income decile ratios) or cross country variation (such as per capita income).[1] We employ internationally comparable household income data from the Luxembourg Income Study (LIS) for 30 countries and 20 years. The LIS provides us with percentile level household income data from which we construct income distributions for cross-country as well as inter-temporal comparisons.

We construct our price distributions using international trade data. Previous authors have shown that prices vary substantially across exporters and covary with exporter

[1] Two countries might have similar degrees of within-country income inequality according to standard measures such as the Gini coefficient or the 90/10 ratios of income, but have very different income means. Similarly, two countries might have similar mean per capita income but difference variances. In both instances, our theory would predict significant differences in price distributions across the two countries.

characteristics such as per-capita income and per worker supplies of capital and skilled labor (Schott 2004, Hummels and Klenow 2005). Further, countries with high export prices have larger, not smaller, shares of the markets in which they sell (Hallak 2005). These facts point to the primacy of quality differentiation, as opposed to measurement error, as an explanation for measured price variation. For each product we observe from which exporters an importer buys, along with each exporter's price and share of trade, and from these construct price distributions for each importer and product.

We find strong support for our model. The difference in importers' price distributions are closely linked to the difference in their income distributions and the importers with similar income distributions also have more export partners in common. Our results are consistent with Murphy and Shleifer'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. Both findings hold when we make pairwise comparisons of importers or when comparing an importer to the world as a whole, and they are stronger for consumer goods than for capital or intermediate goods. Finally, our results only hold up when we use the theoretically appropriate measures of differences in income distributions; conventional measures of within-country income inequality produce coefficients with the wrong signs in our regressions.

Our work relates to the literatures on product variety in trade, quality differentiation and non-homothetic preferences. Most of the empirical work in the product variety literature employs horizontal differentiation models in which representative agent consumers have love-of-variety preferences (e.g. Helpman and Krugman 1985). These models have several characteristics that we highlight to contrast with our model. One, since consumers desire all foreign varieties, these models invoke fixed costs of trade to explain why, in the data, countries import only small subsets of available varieties. Two, utility is increasing in the number of varieties, so greater variety implies greater welfare gains from trade. Three, consumers allocate expenditure shares independently of income. Looking across markets, expenditure shares differ only because trade costs alter relative prices of different varieties. In our model, a household desires a single quality differentiated variety, while an economy as a whole desires subsets of the world's varieties dictated by its income distribution. There are no fixed costs of trade but importers will choose not to access the foreign varieties whose qualities are too high or too low for consumer incomes there. Countries whose income distributions span a wider range can access more varieties with no particular welfare implications. Finally, an importer's expenditure shares on particular varieties depend on the income distribution and so vary across countries even when these countries face a common set of prices.

Most of the empirical work in the quality differentiation literature has focused on linking price variation to exporter characteristics. Some authors have provided correlations with importer characteristics, showing that within product categories, countries with high mean income per capita buy goods with higher mean prices (Hallak 2005, Hummels and Skiba 2004). Our work differs in that it provides an explicit structural linkage between non-homothetic preferences, income variation and product prices. We also examine the entire distribution of incomes and prices, rather than just the first moments.

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 the work in this literature 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 per capita incomes. An exception is Dalgin, Mitra and Trindade (2004), who show that the imports of luxury goods are increasing, (and imports of necessities are decreasing) in a measure of within-country income equality. We differ from this literature in two respects. First, we highlight 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 the income decile ratio) are neither theoretically appropriate nor empirically useful for cross-country differences in price distributions and numbers of export partners.

The paper proceeds as follows. Section 2 provides the theory linking a country's income and import price distributions. Section 3 discusses our empirical specification. Section 4 explains the construction of our income and price distribution data in detail. Section 5 presents the empirical results and Section 6 concludes.

2 The Model

Flam and Helpman (1987) provide a model in which heterogeneity in household income is mapped into heterogeneity in optimal quality choice. We extend their model to a multi-country, multi-good setting, with an analysis motivated by and focused on empirical feasibility. That is, in an international context we are unable to empirically observe household incomes and the qualities and prices of goods consumed at the household level. However, we can observe a country's income distribution, as well as the distribution of prices for imported goods. We provide propositions linking differences in importers' income distributions to differences in import price distributions.

We start with a model with one differentiated good and identical technologies across countries, and then extend our analyses to multiple goods and different technologies. We develop a key empirical prediction in this section, and then discuss complications to the stylized model that will subtly alter this prediction in section 3.

2.1 Identical Technologies and One Differentiated Good

There are two goods, a homogeneous numeraire good and a vertically differentiated good. There are C countries. Each country c has population N_c, with income I distributed exogenously[2] according to the probability distribution function (pdf) $g_c(.)$ with support G_c.

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

(1) $$u(y,z) = ye^{\alpha z} \quad s.t. \; y + p(z) \le I,$$

where $\alpha > 0$, αz 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 is sufficiently high so that every consumer consumes the differentiated good.

We initially assume that all countries produce with an identical technology. The

[2] This assumption allows us to focus on the role of national and world income distributions in determining quality demand, but we abstract from the feedback channels through which trade affects income, as in Flam and Helpman (1987)'s seminal work.

marginal cost of producing quality is

(2) $MC(z) = e^{\gamma z} w$.

w represents the 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. γz is the elasticity of the marginal cost with respect to quality. We assume that there are no trade costs, and that there are perfectly competitive markets at each quality level so that consumers in all countries face the same vector of prices $p(z) = MC(z)$.

Figure 1 shows the utility maximization problem. The budget constraint DD is concave because by equation (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(.) is tangent to DD,

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

(4) $p(z) = aI$, where $a = \dfrac{\alpha}{\alpha+\gamma}$.

Equation (4) indicates that a consumer with income I spends a fixed fraction $a = \dfrac{\alpha}{\alpha+\gamma}$ of his income on (one unit of) quality z.

These equations enable us to write a country's distribution of prices consumed in terms of its distribution of household incomes, $g_c(.)$. Equations (3) and (4) indicate that optimal qualities (and therefore prices paid for the differentiated good) are monotonically increasing in income. That is, 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.[3]

Further, the number of people in country c consuming z* is equal to the number of persons with income $I(z^*)$. As a consequence, the price distribution is a 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. For example, suppose income is distributed log normally $L(\mu,\sigma^2)$. Then the observed price distribution is also distributed log normally $L(\mu a,\sigma^2 a^2)$ and its mean and variance are directly proportional to the mean and variance of the income distribution, respectively. As consumer gains from quality (α) rise, or the cost of producing quality (γ) falls, the mean and variance of the price distribution also rise.

For the more general income distribution $g_c(.)$ with support G_c, let $f_p(I) = aI$ be the function mapping incomes to prices. Since prices are strictly increasing in income we can use (4) to rewrite income as an inverse function of prices, or $I = p/a$. Then we have a price distribution

(5) $$h_c(p(z)) = g_c\left(\frac{p(z)}{a}\right)\cdot\frac{1}{a} \qquad \text{with support } H_c = f_p(G_c).$$ [4]

Next, we describe cross-country differences in the price and income distributions. Because all consumers world-wide face the same vector of prices and the concave budget

---

[3] This is an implication of assuming no fixed costs of production and perfectly competitive markets.

[4] For example, if $G_c = [0, b]$, then $H_c = [0, ab]$. On the other hand, if $f_p(I)$ takes a more general form than aI, then equation (5) becomes $h(p(z)) = g\left(f_p^{-1}(p(z))\right)\cdot[f_p^{-1}(p(z))]'$ with support $H_c = f_p(G_c)$, provided that $f_p(I)$ is strictly increasing in income and its inverse exists and is differentiable.

set shown in Figure 1, equation (5) holds for every country c. This means that the price distribution of the differentiated good consumed in country c is mapped to its income distribution and 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.[5]

Definition 1 Dis-similarity Index (DSI): The DSI for the pair of distributions with pdf's $f_1(.)$ and $f_2(.)$ and 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 $\underline{S} = S_1 \cup S_2$, $f_1(.)$ is defined to be 0 for $\underline{S} - S_1$ and $f_2(.)$ defined to be 0 for $\underline{S} - S_2$.

The DSI quantifies the difference between $f_1(.)$ and $f_2(.)$ by calculating the vertical distance between them at every point x and then aggregating these vertical distances. If $f_1(.)$ and $f_2(.)$ are dis-similar, i.e. 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 $f_1(.)$ and $f_2(.)$ 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

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

$g_1(I)$ 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

[5] The DSI is half the L_1 distance between the pdf's $f_1(.)$ and $f_2(.)$. Another commonly used distance metric is the L_2 $\int_{\underline{S}} [f_1(x) - f_2(x)]^2 \, dx$. We have chosen the L_1 metric because it enables our DSI index to fall between 0 and 1.

measures the difference in population shares at each income level and then sums the difference over the support of the income distribution.

Writing out the price dissimilarity index explicitly, we have

(7) $$PDSI(h_1,h_2) \equiv \frac{1}{2}\int_{\underline{H}} \left| g_1\left(\frac{p(z)}{a}\right)\cdot\frac{1}{a} - g_2\left(\frac{p(z)}{a}\right)\cdot\frac{1}{a} \right| dp(z) \quad \text{where } \underline{H} = f_p(\underline{G}).$$

Proposition 1 $PDSI(h_c,h_{c'}) = IDSI(g_c,g_{c'})$ where c and c' represent any country pair.

Proof: See Appendix 1.

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

For an example to illustrate Proposition 1, suppose that the income distributions of countries 1 and 2 have the same support $G_1 = G_2 = [0,b]$ so that $\underline{G} = [0,b]$. Then

$$\begin{aligned} PDSI(.) &= \frac{1}{2}\int_0^{ab} \left| g_1\left(\frac{p(z)}{a}\right) - g_2\left(\frac{p(z)}{a}\right) \right| \frac{1}{a}\, dp(z) \\ &= \frac{1}{2}\int_0^{b} | g_1(.) - g_2(.) |\, dI \\ &= IDSI(.) \end{aligned}.$$

We also find it useful to compare the income and price distributions of a country c with the world distributions. The number of people with income I from country c equals $N_c g_c(I)dI$, while the 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, ... 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

(8) $\quad g_w(.) = \sum_c \lambda_c g_c(.)$ with support G_w.

Because every consumer consumes one unit of the differentiated good, the world price distribution has the pdf

(9) $\quad h_w(.) = \sum_c \lambda_c h_c(.)$ with support $H_w = \cup_c H_c$,

where H_c is as defined in equation (5).

Corollary 1 $PDSI(h_c, h_w) = IDSI(g_c, g_w)$ for every country c, where $g_w(.)$ and $h_w(.)$ are defined in equations (8) and (9), respectively.

Hereafter, we refer to the comparisons involving two countries c and c' as the bilateral comparisons and those involving a country c and the world as the multilateral comparisons.

2.2 Multiple Differentiated Goods

We extend the model to a multiple differentiated good setting to show that Proposition 1 holds for each differentiated good. Let $k = 1...K$ index the differentiated goods and z_k denote the quality of good k. Utility over the numeraire y and the K differentiated goods is given by

(10) $u = y\exp(\sum_k \alpha_k z_k)$.

The marginal cost of each differentiated good is given by equation (2) though γ may differ across goods, i.e. $MC(z_k) = \exp(\gamma_k z)w$.

As in equation (4), consumers spend a fixed fraction of their income on each differentiated good, and the remaining implications go through. To be specific, the mapping from prices to income for good k is given by

(11) $$p_k(z) = \frac{\alpha_k / \gamma_k}{1 + \sum_{l=1}^{K} \alpha_l / \gamma_l} I .$$

Clearly, the slopes of the price-income relationship will differ across goods k.[6] The goods with high values of α_k / γ_k, that is, those for which there is a high ratio of marginal utility to marginal cost of quality, will have a steep price-income slope. For a given distribution of income, goods with high values for α_k / γ_k will also have price distributions that have higher means and variances, just like in section 2.1. These price distributions are then a measure of the endogenous degree of vertical differentiation that the economy supports in equilibrium, as a function of technology, preferences, and the distribution of income.

However, the price dissimilarity index measures the difference between two countries' price distributions and the differencing removes the variation in the price-income slopes across goods. Put another way, since prices for each good k are a one to one mapping from income, the difference between the consumption shares for $p(z_k *)$ in

[6] Bils and Klenow (2001) call these slopes "Quality Engel Curves" and use US household data to estimate how they differ across a set of consumer durable goods.

countries 1 and 2 is just the difference in their population shares at $I(z_k*)$. Thus:

Corollary 2 When there are many vertically differentiated goods k = 1...K, $PDSI^k(h_c^k, h_{c'}^k) = IDSI(g_c, g_{c'})$ for every good k.

2.3 Multiple Countries with Different Technologies

Now we allow technologies to differ across supplying countries. This creates two difficulties related to the continuity of the price distribution: the price-income slope varies discretely depending on which country is supplying the differentiated good, and in the price distribution there are "holes", i.e. interior qualities for which there is no demand. However, as was the case with multiple goods described in section 2.2, we show that both problems are eliminated by differencing two consuming countries' price distributions.

For notational simplicity, we return to the one good case, though it is easily shown that Corollary 2 holds when technologies differ across supplying countries. The marginal cost of producing quality z in country j is

(12) $MC_j(z) = \exp(\gamma_j z) w_j$.

w_j represents the cost differences (due to factor price or Ricardian technology differences) that are common to all quality levels. $\exp(\gamma_j z)$ expresses the degree to which country j has a comparative advantage in high or low quality levels. We continue to assume that there are no trade costs, so that the consumers desiring quality z buy it from the lowest marginal cost provider.

The assumption on technology allows different suppliers to have a comparative advantage in different ranges of quality and so creates kinks in the budget set and a discontinuous relationship between prices and income. Despite this complication, equations (3) and (4) hold with small adjustments. To illustrate this point, we use the two-country setting of Flam and Helpman (1987), where the North and the South have technologies $MC_N(z) = \exp(\gamma_N z)w_N$ and $MC_S(z) = \exp(\gamma_S z)w_S$ with $\gamma_N < \gamma_S$ and $w_S < w_N$. The North has the comparative advantage in high qualities.

Figure 2 shows the utility maximization problem for a consumer with income I_d and u(.) is the indifference curve. Figure 2 is similar to Figure 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 $D_S T$). When quality is higher than at point T, the budget constraint is determined by the Northern marginal cost (along the curve TD_N). The indifference curve is tangent to both segments of the budget constraint; 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.

First, note that there is no demand, in the North or the South, for the qualities between z_1 and z_2. Second, *for the qualities that are actually supplied to the market*, $[0, z_1] \cup [z_2, 1]$, equations (3) and (4) hold with the following adjustments: for incomes below I_d, γ_S and w_S replace γ and *w*, and for incomes above I_d, γ_N and w_N replace γ and *w*. That is, when the differentiated good is purchased from the South (North), the price-income slope in (4) is determined by Southern (Northern) technology. However, two consumers living in different countries but with the same income face the same price-income slope.

More generally, let j = 1…J index supplying countries, with each j being the lowest cost supplier of some set of qualities Z_j. Let G_j be the set of incomes for which some qualities in the set Z_j are the optimal quality choice so that consumers with $I \in G_j$ buy the differentiated good from exporter 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 (4) becomes

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

Rather than a single line of constant slope mapping incomes into prices, we have a set of lines whose slopes are determined by the technology of the lowest cost producer for the corresponding quality segment.

To get the price distribution in consuming country c we must combine the more complex expression for the price-income slopes (13) with country c's income distribution. Let $f_{pj}(I) = a_j I$ and let $G_{cj} = G_c \cap G_j$. G_{cj} describes the set of incomes for which consumers in country c buy the differentiated good from exporter j.[7] Then the price distribution of country c is still a transformation of its income distribution, but it is discontinuous and must be evaluated separately over segments of the income distribution.

[7] Supplier j produces the range of qualities that appeal to consumers with income range G_j, but incomes in country c may span part, but not all, of the range.

Equation (5) becomes[8]

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

Unlike in the identical-technology case, the support of the price distribution now consists of disjoint intervals that correspond to the ranges of qualities *actually supplied in the world*. The definition of the dis-similarity index (DSI), and equation (6), the expression for IDSI, do not change, but the expression for PDSI does change. Since the support of the price distribution consists of disjoint intervals, PDSI equals 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 country 1 or country 2 buys the differentiated good from exporter j. Then the expression for PDSI becomes[9]

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

where $\underline{H}_j = f_{pj}(\underline{G}_j)$.

Although the expression for PDSI becomes more complex, Proposition 1 remains intact. Consider first the "holes" in the price distribution, those qualities that are not demanded by any income level and so not demanded by any country. Since these qualities have zero consumption shares for both countries 1 and 2, they carry zero weights in the price distribution for both countries and simply drop out when differencing. Next consider the qualities actually supplied in the world. These segments

---

[8] Like (5), equation (14) still holds when $f_{pj}(I)$ takes a more general form than $f_{pj}(I) = a_j I$, provided that $f_{pj}(I)$ is strictly increasing in G_j, its inverse exists and is differentiable in G_j.

[9] Note that $\cup_j \underline{G}_j = G_1 \cup G_2$ and so $\cup_j \underline{H}_j = H_1 \cup H_2$, where H_1 and H_2 are as defined in equation (14).

have different price-income slopes given by (13), but along each segment prices are still a one to one mapping from income. Just as in the multi-good case, measuring cross country differences in price distributions removes the differences in the price-income slope. The difference between the consumption shares for the good with price $p(z^*)$ in countries 1 and 2 still equals the difference in their population shares at income $I(z^*)$.

To illustrate equations (14) and (15), consider the Flam and Helpman (1987) two-country example again. Suppose the support of the South's income distribution is $[0, b_S]$ and the support of the North's income distribution is $[0, b_N]$, $b_N > b_S$. Consumers in both countries with income $[0, I_d]$ buy the differentiated good from the South and those with income $(I_d, b_N]$ buy it from the North. The price distribution of the South is

$$g_S\left(\frac{p(z)}{a_S}\right)\cdot\frac{1}{a_S} \text{ for } p(z)\in[0,a_S I_d] \text{ with } a_S=\frac{\alpha}{\alpha+\gamma_S} \text{ and}$$

$$g_S\left(\frac{p(z)}{a_N}\right)\cdot\frac{1}{a_N} \text{ for } p(z)\in[a_N I_d, a_N b_S] \text{ with } a_N=\frac{\alpha}{\alpha+\gamma_N}.$$

and similarly for the North. Calculating the PDSI gives us

$$\begin{aligned} PDSI(.) &= \frac{1}{2}\int_{a_N b_S}^{a_N b_N} g_N\left(\frac{p(z)}{a_N}\right)\frac{1}{a_N}dp(z) \\ &+\frac{1}{2}\int_0^{a_S I_d}\left| g_N\left(\frac{p(z)}{a_S}\right)-g_S\left(\frac{p(z)}{a_S}\right)\right|\frac{1}{a_S}dp(z) \\ &+\frac{1}{2}\int_{a_N I_d}^{a_N b_S}\left| g_N\left(\frac{p(z)}{a_N}\right)-g_S\left(\frac{p(z)}{a_N}\right)\right|\frac{1}{a_N}dp(z) \end{aligned} \tag{16}$$

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

3 Empirical Specification

Our model derives a direct linkage between a consuming country's income distribution and that country's distribution of consumer prices for a particular quality differentiated good. Proposition 1 shows that the difference in two countries' price distributions for a particular good is equal to the difference in their income distributions. As we show in sections 2.2 and 2.3, examining cross-country differences in price distributions also allows us to control for differences in the price-income slopes across products and across suppliers.

Unfortunately, it is not possible to get cross-country data on all sales prices within some narrowly defined consumer good. However, it is possible to use import price data to approximate the price distributions. Given the technology assumed in equation (12) each exporting country j will specialize in a range of qualities, with a corresponding range of export prices. By knowing the prices charged by each exporter j, as well as the share of exporter j in importer c's purchases, we can calculate the import price distributions for importers c and c' and the difference between them.

We can then test two implications of the theory. The first directly follows from Proposition 1, implemented across country pairs c-c'

(17) $$PDSI(h_c, h_{c'}) = \alpha_0 + \beta_P IDSI(g_c, g_{c'}) + e_{cc'} .$$

The price distributions for two importers can differ either because they buy from different sets of exporters, or because they buy from similar sets of exporters but with different shares. For reasons we will describe below, international prices are subject to measurement error and so we employ a second test that looks only at whether two importers buy from a common set of exporters.

(18) $\ln(N_{cc'}) = \alpha + \beta_N \ln(IDSI_{cc'}) + e_{cc'}$

where $N_{cc'}$ is the number of exporters countries c and c' have in common. Following the theory, let some exporter j produce quality z^*. Then two importers will buy from exporter j only if both have some population with income $I(z^*)$. The smaller the overlap between two importer's income distribution (i.e. the larger is their income dis-similarity index), the fewer the exporters they will have in common.[10]

In implementing regressions (17) and (18) we make three modifications. First, we follow Corollary 2 and pool the observations across all the products. Second, Corollary 1 enables us to also compare an importer to the rest of the world. We do so by replacing the bilateral price and income dis-similarity indices in (17) with the multilateral price and income dis-similarity indices and replacing the number of common exporters in (18) with the number of exporters. The multilateral comparisons have a more direct analogue to the work that examines the level and growth of product variety across importers.

Finally, models with fixed costs of trade as in Melitz (2003) suggest that countries will import a larger set of varieties when the market is large and when trade costs are low. We augment regressions (17) and (18) with measures of market size[11] and trade costs for the multilateral regressions or relative market size and distance between the importer pairs for the bilateral regressions. This gives us four estimating equations. For the bilateral comparisons we have

---

[10] Note that two exporters (e.g. Mexico and Poland) might have identical technology and so specialize in an identical spectrum of quality. In this case, our theory does not say from which exporter two importers (e.g. US and Germany) will buy quality z*, only that they will buy from someone. If the US buys only from Mexico and Germany only from Poland, regression (18) would fail (i.e. $\beta_N = 0$) but (17) would still hold (i.e. $\beta_P > 0$).

[11] We use GDP to measure market size. GDP is the product of population and GDP per capita, and GDP per capita may be correlated with the mean of the income distribution. We also use population to control for market size instead and find similar results..

(19) $$\ln(N^k_{cc't}) = \alpha_F + \beta_N \ln(IDSI_{cc't}) + \beta_2 \left|\ln(\frac{GDP_{ct}}{GDP_{c't}})\right| + \beta_3 \ln DIST_{cc'} + \varepsilon^k_{cc't}$$

(20) $$PDSI^k_{cc't} = \alpha_F + \beta_P IDSI_{cc't} + \beta_2 \left|\log(\frac{GDP_{ct}}{GDP_{c't}})\right| + \beta_3 \ln DIST_{cc'} + \varepsilon^k_{cc't},$$

where k indexes products, t indexes time and α_F is a set of fixed effects we will explain below. For the multilateral comparisons we have

(21) $$\ln(N^k_{ct}) = \alpha_F + \beta_N \ln(IDSI_{cwt}) + \beta_2 \ln(GDP_{ct}) + \beta_3 \ln MP_{ct} + \varepsilon^k_{ct}$$

(22) $$PDSI^k_{cwt} = \alpha_F + \beta_P IDSI_{cwt} + \beta_2 \ln(GDP_{ct}) + \beta_3 \ln MP_{ct} + \varepsilon^k_{ct},$$

where w represents the world and $MP_{ct} = \sum_{l=1,l\neq c}^{C} GDP_{lt} \bullet d_{lc}^{-1}$ is the market potential of country c (e.g. Hanson and Xiang 2004) and measures the "remoteness" of c with respect to the world. We implement each of these regressions in cross-sections for each wave (i.e. *t* does not vary) and include product fixed effects $\left(\alpha_F = \alpha^k\right)$.[12] We also discuss the panel versions of these regressions in section 5.2.

Representative agent models with homothetic preferences imply that a country's income distribution should have no effect on the number of goods it imports or their price distribution, i.e. $\beta_N = 0$ and $\beta_P = 0$. We take $\beta_N < 0$ and $\beta_P > 0$ as evidence for our theory with quality differentiation and non-homothetic preferences. Even though Proposition 1 implies a more stringent test for the price distribution regressions, $\alpha_0 = 0, \beta_P = 1$, the model of section 2 is stylized. When confronting the data we face difficulties due to missing domestic sales data, trade costs, and more general forms of

[12] Since IDSI does not vary across products, we have cluster samples and the observations within a cluster might be correlated. We employ Wooldridge (2002)'s test for random effects and fail to reject the hypothesis that the within cluster correlations are zero. We also employ robust standard errors (the "cluster" command in STATA) and obtain similar results.

quantity choice.

We use import price data to calculate price distributions, and so we will miss domestic sales data, the portion of the price distribution that a country supplies to itself. We cannot sign the biases in our estimates of β_P in all cases, but we provide two results in Appendix 1. First, recall the setup in Flam-Helpman (1987), where the higher income country also produces higher quality goods.[13] In this case omitting domestic sales will truncate the higher income country's price distribution from above, and truncate the lower income country's price distribution from below. This reduces the measured variation in PDSI relative to IDSI and biases β_P downward away from 1. Second, the errors to the measured price distribution approach zero as the number of countries, C, becomes large and the range of qualities each country supplies shrinks. Having relative size controls in our estimating equations (19)-(22) also helps address this bias, and we experiment with additional robustness checks in section 5.

Our theory assumed away trade costs so that all importers would face the same lowest cost supplier for each quality. With trade costs, a country might source a smaller range of qualities from abroad and a larger range from itself, and this raises the same issues as missing domestic sales data. Trade costs also imply that different countries might import the same quality from different exporters, as discussed in footnote 10. We control for trade costs using distance in the bilateral comparisons and market potential in the multilateral comparisons.

Finally, our theory assumed that every consumer would purchase one unit of the differentiated good. In a more general setting, consumers might use marginal income to

[13] In the fully specified general equilibrium, higher incomes and the ability to produce higher quality goods both stem from a larger endowment of skilled labor.

expand consumption both on a quantity and a quality dimension (Bils and Klenow 2001). We derive the expressions for the price distribution and the price dis-similarity index, PDSI, for this more general form of quantity choice in Appendix 1. However, the exact relationship between PDSI and IDSI is complicated by interactions between the income distribution and the functions mapping income into optimal quantity choice, and so we cannot sign the bias to our estimates of β_P.

4 Data

4.1 Income Data

To facilitate cross-country and inter-temporal comparisons of income distributions, we employ the Luxembourg Income Study (LIS) data. The LIS data are a compilation of the income survey data files of 30 countries, made comparable by rearranging or reclassifying the income measures from national household budget surveys.

Another widely used dataset on cross-country income distribution is Deininger and Squire (1996) and its extensions by the World Bank (the DSWB).[14] We have chosen the LIS data for three reasons. One, the LIS is more consistent and better suited for cross-country and inter-temporal comparisons of income distributions (Atkinson and Brandolini 2001, Deaton 2003). It provides disposable household income (monetary income after direct taxes and transfer payments), and allows us to make adjustments to account for differences in family size. Two, the LIS allows us to calculate household income at single *percentile* increments while the DSWB provides *quintile* level income shares. Three, even though the LIS covers a smaller number of countries than the

[14] For other income distribution data see Chen and Ravallion (2001), Bourguignon and Morrisson (2002) and Milanovic (2002).

DSWB, our theory makes predictions that involve pair-wise comparisons of countries and so we can still generate considerable cross-section variation in income distributions.

As shown in Table 1, we have up to 30 countries at roughly 5 year intervals for the period 1979-2001 (Wave 1 - Wave 5). The starred countries in Table 1 have data for only one or two years. In some other cases, data are missing for a wave and have been estimated following a procedure detailed in Appendix 2. In Table 1 we mark such cases with an "(e)".

Our theory requires us to construct and then compare income distributions across countries. To construct a continuous income distribution from the discrete household income data we perform a non-parametric kernel estimation using the "kdensity" command in STATA (Deaton 1997). We use STATA's default kernel, the Epanechnikov, and STATA's default bandwidth,[15] and evaluate the densities of the distributions of all the countries at the same income levels of $100, $200 ... $150,000.[16] We then calculate the differences in income distributions, both pair wise and relative to the "world" following equations (6) and (8). For our purposes the "world" consists of all the countries in the LIS data in a given year.[17] We then multiply the IDSI by 100.

Figure 3 shows several measures of income dispersion for the LIS countries in Wave 5 (2000). The 3rd column reports the multilateral income dissimilarity index (IDSI) given by equation (6). The 4th column reports a more conventional measure of within-country income dispersion, the ratio of the 90th percentile income to the 10th percentile income, or

[15] 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.
[16] It is vital that the evaluations be at the same income levels; otherwise we would calculate many IDSI's to be 1. To see this, suppose we record country 1's distribution as $10 and $20 with probabilities 0.5 and country 2's distribution as $10.1 and $19.9 with probabilities 0.5. Then we will calculate their IDSI as 1.
[17] Note that this builds in a correlation between the world income distribution and the individual countries' distributions in our sample. We experiment with excluding country c from the construction of the world distribution for comparison to country c and obtain similar results.

decile ratio. Finally, we normalize each country's income relative to the US median income ($24,094 = 100) and plot the range of income starting at the 10^{th} percentile (P10) and ending at the 90^{th} percentile (P90). We arrange the countries in ascending order of their decile ratios.

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 provides no information about how much of the world income distribution a country's income spans. This can be seen by comparing the US and Mexico in Figure 3. The decile ratio for Mexico (10.4) is nearly twice that of the US (5.4), but as the mean income levels for Mexico are lower than those for the US, Mexico's income distribution spans a much smaller range than the US.

Further, the decile ratio contains no information about how two countries' income distributions compare within the P10-P90 range. In contrast, our IDSI measure employs data from all points in the income distribution and explicitly compares both the level and distribution of two countries' income. When the income distributions of two countries lie far away from each other (e.g. the US and Mexico in Figure 3), the vertical distances between them are large at each point in the distribution and IDSI is large. IDSI achieves its maximal value of 100 if two countries have completely disjoint distributions and achieves its minimal value of 0 if two countries have identical distributions.

4.2 Data on Import Prices and their Distributions

The trade data to implement regressions (19)-(22) come from the world trade flows database (the WTF) (Feenstra et. al., 2005) and the United Nations (UN) trade database.

These data report bilateral import value and quantity at the SITC 4 digit level (roughly 1000 goods). To match the LIS income data wave 1 – 5, we use import data for 1980, 1985, 1990, 1995, and 1999.[18]

These data allow us to count the number of exporters from whom an importer has purchased a product in a given year (for multilateral comparisons), or the number of common exporters from whom two importers have purchased a good (for bilateral comparisons). These counts are used as the dependent variables in estimating equations (19) and (21).

The price data are more problematic and subject to measurement errors of three sorts. First, we construct import prices using import unit values (value/quantity), but the quantity units are unknown and are likely to be importer specific.[19] Controlling for the measurement error of this sort is critical for our application because we can not properly calculate the difference between two importers' price distributions unless we have accurate data on the level of prices. Second, exporters may produce a range of qualities within a product category, but we observe average prices rather than the entire range. Third, quantity (but not value) data are missing for some of the importers

Accordingly, we extract exporter-specific signals from the noise of the raw data in the following way. For each SITC 4 digit product k, we observe the raw prices (unit values) for some subset of importer c – exporter j pairs. We regress the log price on importer-product and exporter-product fixed effects plus bilateral distance.

(23) $$\ln p_{cjt}^{k} = \alpha_{ct}^{k} + \alpha_{jt}^{k} + \beta_{t}^{k}\text{Distance}_{cj} + \varepsilon_{cjt}^{k}$$

Bilateral distance sweeps out Alchian-Allen effects in pricing (Hummels and Skiba 2005)

---

[18] The WTF 2000 data has some technical problems (Feenstra et al. 2005), so we use 1999 data.

[19] Some importers might report quantities in weight terms, either kilograms or pounds, while others use counts. See also Hummels and Klenow (2005).

and the importer-product fixed effects absorb the variation in unit values that arises from differences in units.

We use the exporter-product fixed effects as our measure of export prices, $\widehat{\ln p_{jt}^k} = \hat{\alpha}_{jt}^k$. This implies that any importer c buying product k from exporter j faces this common export price. In addition to removing importer-specific measurement errors, regression (23) provides us with (estimated) exporter prices even for those importers for whom no quantity data are reported.

A drawback of this approach is that we may lose useful variation across importers in the price charged by a particular exporter. For example, suppose that Germany produces a range of car qualities (and prices). While Mexico imports cars from the lower end of the range, the US may import the entire quality range. Unfortunately, we are unable to separate this true variation in German export prices from importer-specific measurement errors, and so we err on the side of removing all the importer-specific variation.

Does this approach yield sensible export price data? We know from previous work using different data that prices are highly correlated with exporter's per capita income (Schott 2003, Hummels and Klenow 2005). We regress log prices on exporter per capita GDP and product fixed effects, using both the raw price data, $\ln p_{cjt}^k$, and our estimated prices, $\widehat{\ln p_{jt}^k}$.

$$\ln p_{cjt}^k = -0.97 + .246 \ln(Y_{jt} / L_{jt}) + \alpha_t^k + e_{cjt}^k \qquad R^2 = 0.48$$
$$\widehat{\ln p_{jt}^k} = -2.95 + .233 \ln(Y_{jt} / L_{jt}) + \alpha_t^k + e_{cjt}^k \qquad R^2 = 0.94$$

There is no statistically significant difference in the exporter per capita GDP coefficient. However, the regression using our estimated prices has almost twice the R^2 of the regression using the raw data. This suggests that our estimated prices are much cleaner measures than the raw prices.

We use these estimated prices to construct a price distribution for each importer c, product k and wave t in the following way. We take the estimated prices $\widehat{\ln p_{jt}^{k}}$ of every exporter j, and weigh them by the share of j in importer c's purchases of k (this share is zero if j does not ship to c).[20] The resulting distribution is discrete and we again employ the nonparametric kernel estimation to obtain a smooth price distribution. Finally, we take the differences in the price distributions (either relative to another importer c' or to the world) to obtain the price dis-similarity indices following Definition 1.[21] We then multiply the PDSI by 100.

5. Results

We calculate the coefficient of variation (standard deviation/mean) for the key variables in regressions (19)-(22) using the data of all countries and all waves and report the results in Table 2. Not surprisingly, the variables for the bilateral comparisons have more variation (relative to the means) than those for the multilateral comparisons. Figure 4 plots the simple average of the multilateral PDSI across all products against the multilateral IDSI for wave 2 and shows a clear positive correlation between these two variables. Their correlation coefficient is a significant 0.42. Figure 5 is the

[20] We use the value shares, even though quantity shares are more consistent with theory, because quantity data are missing for a significant number of observations (e.g. 34% of the observations for wave 1). We experiment with dropping the observations with missing quantity data, or imputing the missing quantities as values divided by the prices estimated using equation (23), and then constructing price distributions using quantity shares. The results are almost identical to Tables 3-6

[21] The "world" consists of those countries for whom we have income measures from LIS, not the entire sample of world trade flows.

corresponding scatter plot for the bilateral PDSI and IDSI; again, the positive correlation is unmistakable, with a statistically significant correlation coefficient of 0.51.

5.1 Main Regression Results

Table 3 reports the results of estimating the multilateral PDSI regression (22) by wave. The dependent variable is the multilateral price dis-similarity index (PDSI). In the top panel we estimate the base regression with the multilateral income dis-similarity index (IDSI) as the only regressor, and in the bottom panel we include the additional control variables of log GDP and log market potential. For the base regression, the coefficients on the multilateral IDSI are positive and precisely estimated for all waves, ranging from 0.2 (wave 5) to 0.45 (wave 2). These coefficients are somewhat smaller when we include the additional controls, but the positive correlation between IDSI and PDSI remains significant. That is to say, the countries whose income distributions are most different from the world also have import price distributions that are most different from the world, consistent with our theory. On the other hand, our control variables have the expected signs: small domestic markets and those farthest from suppliers have price distributions less similar to the world.

Table 4 is organized similarly to Table 3 and reports the bilateral PDSI regression (20). Like in Table 3, the coefficients on (the bilateral) IDSI are positive and significant for all waves, with or without the additional controls of log distance and the absolute value of the log difference in GDP. Again our control variables have the expected signs: the countries of different sizes have dissimilar price distributions, as do the countries far away from one another. The latter is important as Proposition 1 rests on the assumption that trade costs are negligible (so that all importers face a similar price vector) whereas trade costs are likely to matter in the data. Thus it is reassuring that introducing the additional control variables does not substantially alter the coefficients on IDSI.

Table 5 reports the results of the number-of-export-partner regression (21). Unlike for the PDSI regressions (20) and (22), theory does not provide a specific functional form for regression (21) and so we report the log linear specification in the top panel and the linear specification in the bottom panel. The linear specification has more observations because it includes the zero values and also has lower adjusted R^2. In both specifications the coefficients on (the multilateral) IDSI are negative and precisely estimated, and they range from -0.11 to -0.47 for the log linear specification.[22] Put another way, the countries whose income distributions are more similar to the world source from a larger set of export partners, consistent with our theory. On the other hand, country size is positively correlated with the number of export partners for most waves but market potential has limited impacts.

Table 6 is organized similarly to Table 5 and reports the results of the number-of-common-export-partner regression (19), again estimated with both log linear and linear specifications. Again, consistent with our theory, the coefficients on IDSI are negative and significant and they range from -0.17 to -0.25 for the log linear specification: the countries with dissimilar income distributions have fewer export partners in common, consistent with our theory. Distant countries also have fewer export partners in common.

5.2 Robustness Exercises

[22] It is misleading to directly compare the coefficients on IDSI for the log linear specification with the linear specification because the former are elasticities but the latter are derivatives of levels on levels.

We also explore whether the changes in income distribution within a country over time induce changes in this country's import price distribution. We do so by pooling over waves 1-5 and estimating regressions (19)-(22) in panels. Since the set of countries with available income data varies by wave and the multilateral PDSI and IDSI change with the country composition of the data, we restrict our sample to the set of countries (those without a "*" in Table 1) whose data are available for all waves. We also include country-pair by product or country by product fixed effects to focus on inter-temporal variation. The results for the panel regressions are weaker than those for the cross-section regressions in Tables 3-6: the IDSI coefficients are much smaller in magnitude and sometimes not significant. This is likely because most of the variation in the PDSI and IDSI is cross-sectional. For example, an analysis-of-variance shows that the cross-country variation of the multilateral PDSI is over 28 times the inter-temporal variation.

So far we have pooled over all product codes available in the data. One may be concerned that our theory maps household incomes into consumer product prices and so may be less appropriate for intermediate or capital goods.[23] We use the UN Broad Economic Classification system to separate the SITC product codes into consumption, intermediate and capital goods and then re-run regressions (19)-(22) for each of the three categories, employing all controls.[24] To save space we only report the IDSI coefficients and their standard errors for the consumption goods regressions in Table 7. They are significant and have the expected signs and are similar in magnitude to those in Tables 3-

[23] If high quality consumption goods are made using high quality intermediate and capital goods in the country of consumption then the idea behind the theory goes through as before. It is more problematic in cases, like Mexico, where a lower income country imports high quality intermediates in order to produce high quality consumption goods for export to the US.

[24] We also experimented with separating the SITC products into manufacturing goods (SITC 5-8) and commodities (SITC 0-4), or into differentiated, reference priced and homogeneous goods using the Rauch (1999) classification. The results for regressions (19)-(22) are similar between these categories.

6. The IDSI coefficients for the intermediate and capital goods regressions are also significant and have the expected signs in most cases, although they tend to be smaller in magnitude than those for the consumption goods regressions.

Our measure of income dissimilarity, IDSI, is derived from theory and it uses both within- and cross-country information on income distribution. Does a more conventional and less theoretically appropriate measure also work? We replace the multilateral IDSI in regressions (21)-(22) with the decile ratio (the ratio of the 90^{th} percentile income to the 10^{th} percentile income), a measure that reflects only within-country inequality. Suppose that all countries had identical median incomes so that the decile ratio were a sufficient statistic for an importer's income range relative to other countries'. Then according to our theory, the decile ratio should be positively correlated with the number of exporters and negatively correlated with our multilateral PDSI measure. Yet the opposite is true as shown in Table 8: the coefficients on the decile ratio are significant and of the wrong signs in all cases. One reason is that the countries with especially large decile ratios (e.g. Mexico, Russia) also have very low median incomes and span but a small portion of the world income distribution. We conclude that employing the theoretically appropriate income measure is critical to our results.

Previous studies (e.g. Hallak 2005, Hummels and Skiba 2004) have shown that rich countries tend to import high priced goods. Does a positive correlation between the first moments of the income distribution and price distribution drive our results? If so, dropping the poorest countries should weaken our results. We drop the 8 countries with the lowest per capita income[25] from our sample, re-calculate the PDSI and IDSI and obtain very similar results. The IDSI coefficients are sometimes larger and sometimes

[25] They are Russia, Romania, Estonia, Slovak Republic, Poland, Hungary, Mexico and Czech Republic.

smaller than in Tables 3-6. Thus our results are not driven by the first moments of the income and price distributions.

Our theory refers to the entire consumption price distribution but our data include only import prices and exclude domestic sales. We have partially dealt with this by including market size and trade cost controls in regressions (19)-(22), and we also try the following for the bilateral regressions (19)-(20). For importer c', we have all the sales data except domestic sales originating in c'. When comparing importer c to c' we want a common reference set of countries. Accordingly, we eliminate country c exports to c' so that country c sales are missing from the sales data for both countries. Re-calculating the bilateral PDSI in this manner and re-running regressions (19)-(20), we obtain results almost identical to Tables 4 and 6.

6. Conclusion

In this paper, we investigate how the distribution of income shapes patterns of consumption and international trade in quality differentiated varieties within narrow product categories. We extend Flam and Helpman (1987) to the case of multiple differentiated goods and multiple countries with different technologies. We show that cross-country differences in the distribution of income lead to differences in variety consumed and in the distribution of product prices. Our extension provides two critical empirical benefits. One, by deriving results in terms of national income and price distributions we are able to evaluate a model that predicts heterogeneity in household consumption decisions without needing household consumption data. Two, the relationship between income, product quality and product may vary widely across products and this can confound efforts to use price data as a signal of quality. The differencing removes cross-product variation in price-income slopes and allows us to make a clean evaluation of income distribution effects.

To test these predictions we employ microdata on income from household surveys for 30 countries over 20 years to construct income distributions within and across countries. We provide an easy to implement methodology for extracting useful information on export prices from the noise of raw trade data and construct price distributions. We find strong support for the predictions of our model. The pairs of importers whose income distributions look more similar have more export partners in common and more similar import price distributions. Importers whose income distributions look more like the world buy from more exporters and have import price distributions that look more like the world.

Our findings, based on a structural model with quality differentiation, show that a country's income distribution shapes its import demand in important ways. This view of trade patterns lies in stark contrast to the dominant models of horizontal product differentiation in the trade literature, which provide no role for heterogeneous consumers or income differences in explaining trade patterns. Further, our findings lend support to Murphy and Shleifer's (1997) insight that developing countries may have limited access to developed countries' markets because the goods they produce lack the high qualities that high-income consumers demand.

Finally, there is a rich theoretical literature on quality differentiation in trade in which authors combine 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, Shleifer 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.

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, Nonhomethetic 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.

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.

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., 2005, Product Quality and the Direction of Trade, *Journal of International Economics,* forthcoming.

Hanson, Gordon H. and Chong Xiang. "The Home Market Effect and Bilateral Trade Patterns." *American Economic Review*, September 2004, 94(4), 1108 – 1129.

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, 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.

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.

Murphy, K. M. and A. Shleifer, 1997, Quality and Trade, *Journal of Development Economics*, 53, pp. 1-15.

Murphy, K. M., A. Shleifer and R. Vishny, 1989, Income Distribution, Market Size, and Industrialization, *Quarterly Journal of Economics*, 104 (3), pp. 537-564.

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.

Stokey, N. L., 1991, The Volume and Composition of Trade Between Rich and Poor Countries, *Review of Economic Studies*, 58(1), pp. 63-80.

Appendix 1 Theory

1. Proof of Proposition 1

Consider the setting of section 2.3, in which technology differs across supplier countries. Since $\cup_j \underline{G}_j = G_1 \cup G_2 = \underline{G}$ (see footnote 9), we can re-write equation (6) as IDSI(.) = $\frac{1}{2}\sum_{j=1}^{C}\int_{\underline{G}_j} | g_1(.) - g_2(.) | dI$. Since $f_{pj}(I) = a_j I$ is strictly increasing in $\underline{G}_j$ (see equation (13)), $\int_{\underline{G}_j} | g_1(.) - g_2(.) | dI = \int_{\underline{H}_j} | g_1(\frac{p(z)}{a_j}) - g_2(\frac{p(z)}{a_j}) | dp(z)$ for all j and so IDSI(.) = $\frac{1}{2}\sum_{j=1}^{C}\int_{\underline{H}_j} | g_1(\frac{p(z)}{a_j}) - g_2(\frac{p(z)}{a_j}) | \frac{1}{a_j} dp(z)$ = PDSI(.), where the last equality is by equation (15).

Suppose that $f_{pj}(.)$ takes a more general form than $a_j I$, but $f_{pj}(.)$ is strictly increasing in G_j, its inverse exists and is differentiable in G_j. Then 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 steps as above go through and PDSI(.) = IDSI(.). Q.E.D.

2. Missing Domestic Sales Data I

Consider the setting of section 2.3 again and let Λ_c be the range of qualities that country c supplies. Assume that

Assumption A1 As C $\rightarrow +\infty$, $\int_{\Lambda_c} m dz \rightarrow 0$ for all c and any finite number *m*.

Below we show that

Proposition A1 As C $\rightarrow +\infty$, $\beta_P \rightarrow 1$ in regression (17) under Assumption A1.

Proof: Let h_1'(.) denote the price distribution of country 1 that *we observe in our data*. Suppose that country 1 supplies the quality range with the set of prices H_1^S. Let the probability mass of H_1^S be $P(H_1^S) = \int_{H_1^S} h_1(p(z)) dp(z)$, where $h_1(.)$ is as defined in equation (14). Then $h_1'(p(z)) = \frac{1}{1-P(H_1^S)} h_1(p(z))$ if $p(z) \in H_1 - H_1^S$, the set of country 1's import prices, and $h_1'(p(z)) = 0$ if $p(z) \in H_1^S$, the set of country 1's domestic sales prices. Likewise, for country 2, $h_2'(.) = \frac{1}{1-P(H_2^S)} h_2(.)$ if $p(z) \in H_2 - H_2^S$ and $h_2'(.) = 0$ if $p(z) \in H_2^S$, where $P(H_2^S) = \int_{H_2^S} h_2(p(z)) dp(z)$.

$H_1 - H_1^S$ is a subset of $\cup_d H_d^S$, where d = 2, … C, because when country 1 buys quality p(z) from abroad, it must buy it from some other country. For the set $B \equiv (H_1 - H_1^S) \cap H_2^S$, by Assumption 1,

(A1) $\int_B | h_1'(.) - h_2'(.) | dp(z) \rightarrow 0$ as C $\rightarrow +\infty$.

For the rest of the set $H_1 - H_1^S$, we observe both countries 1 and 2's price distributions and so $h_1'(.) = \frac{1}{1-P(H_1^S)} h_1(.)$ and $h_2'(.) = \frac{1}{1-P(H_2^S)} h_2(.)$. Thus by equation (A1),

(A2) $\int_{H_1 - H_1^S} | h_1'(.) - h_2'(.) | dp(z) =$

$$\int_{H_1 - H_1^S - B} \left| \frac{1}{1-P(H_1^S)} h_1(.) - \frac{1}{1-P(H_2^S)} h_2(.) \right| dp(z) + \int_B \left| h_1^{'}(.) - h_2^{'}(.) \right| dp(z)$$

$$\rightarrow \int_{H_1 - H_1^S} \left| \frac{1}{1-P(H_1^S)} h_1(.) - \frac{1}{1-P(H_2^S)} h_2(.) \right| dp(z) \quad \text{as } C \rightarrow +\infty.$$

By Assumption 1, as $C \rightarrow +\infty$, $P(H_1^S) \rightarrow 0$, $P(H_2^S) \rightarrow 0$ and

(A3) $\int_{H_1^S} | h_1^{'}(.) - h_2^{'}(.) | dp(z) \rightarrow 0$ as $C \rightarrow +\infty$.

Thus equation (A2) becomes

(A4) $\int_{H_1 - H_1^S} | h_1^{'}(.) - h_2^{'}(.) | dp(z) \rightarrow \int_{H_1 - H_1^S} | h_1(.) - h_2(.) | dp(z)$ as $C \rightarrow +\infty$.

By equations (A3) and (A4), for the set H_1,

$$\int_{H_1} | h_1^{'}(.) - h_2^{'}(.) | dp(z) = \int_{H_1 - H_1^S} | h_1^{'}(.) - h_2^{'}(.) | dp(z) + \int_{H_1^S} | h_1^{'}(.) - h_2^{'}(.) | dp(z)$$

$$\rightarrow \int_{H_1 - H_1^S} | h_1(.) - h_2(.) | dp(z) \rightarrow \int_{H_1} | h_1(.) - h_2(.) | dp(z) \text{ as } C \rightarrow +\infty.$$

Likewise, for the set H_2,

$$\int_{H_2} | h_1^{'}(.) - h_2^{'}(.) | dp(z) \rightarrow \int_{H_2} | h_1(.) - h_2(.) | dp(z) \text{ as } C \rightarrow +\infty.$$

Therefore, as $C \rightarrow +\infty$, the PDSI that we observe in the data, $\int_{H_1 \cup H_2} | h_1^{'}(.) - h_2^{'}(.) | dp(z)$, approaches the true PDSI, $\int_{H_1 \cup H_2} | h_1(.) - h_2(.) | dp(z)$. Since the true PDSI equals IDSI by Proposition 1, our β_P estimate approaches 1. Q.E.D.

3. Missing Domestic Sales Data II

Consider the setting of section 2.3 again. Suppose that country 1 is rich and country 2 is poor. Assume that

Assumption A2 Country 1 specializes in high qualities and they are not demanded by country 2; country 2 specializes in low qualities and they are not demanded by country 1.

Below we show that

Proposition A2 For countries 1 and 2, the PDSI observed in the data is no larger than IDSI.

Proof: Let $h_1'(.)$ denote the price distribution of country 1 observed in the data and let H_1^D denote the set of prices for the high qualities country 1 produces. Since country 1 supplies H_1^D and only H_1^D to itself and we do not observe domestic sales data, $h_1'(p(z)) = 0$ if $p(z) \in H_1^D$ and $h_1'(p(z)) = \frac{1}{1-P(H_1^D)} h_1(p(z))$ if $p(z) \in H_1 - H_1^D$, the set of country 1's import prices, where $P(H_1^D) = \int_{H_1^D} h_1(p(z)) dp(z)$ is the probability mass of H_1^D and $h_1(.)$ is as defined in equation (14). Likewise, for country 2, $h_2'(.) = \frac{1}{1-P(H_2^D)} h_2(.)$ if $p(z) \in H_2 - H_2^D$ and $h_2'(.) = 0$ if $p(z) \in H_2^D$, where $h_2'(.)$ is the observed price distribution of country 2, H_2^D is the set of prices for the low qualities that country 2 produces and $P(H_2^D) = \int_{H_2^D} h_2(p(z)) dp(z)$.

Over H_1^D, $h_2'(.) = h_2(.) = 0$ by Assumption A2. Thus

(A5) $\int_{H_1^D} | h_1^{'}(.) - h_2^{'}(.) | dp(z) = 0$, $\int_{H_1^D} | h_1(.) - h_2(.) | dp(z) = \int_{H_1^D} h_1(.) dp(z) = P(H_1^D)$.

Likewise, over H_2^D, $h_1'(.) = h_1(.) = 0$ by Assumption A2 and

(A6) $\int_{H_2^D} | h_1^{'}(.) - h_2^{'}(.) | \, dp(z) = 0, \ \int_{H_2^D} | h_1(.) - h_2(.) | \, dp(z) = \int_{H_2^D} h_2(.) dp(z) = \mathrm{P}(\mathrm{H}_2^{\mathrm{D}})$.

On the other hand, over the set $R \equiv H_1 \cup H_2 - H_1^D - H_2^D$,

$$\text{(A7)} \quad \int_R | h_1^{'}(.) - h_2^{'}(.) | \, dp(z) = \int_R \left| \frac{1}{1-P(H_1^D)} h_1(.) - \frac{1}{1-P(H_2^D)} h_2(.) \right| dp(z)$$

$$= \int_R \left| h_1(.) - h_2(.) + \frac{P(H_1^D)}{1-P(H_1^D)} h_1(.) - \frac{P(H_2^D)}{1-P(H_2^D)} h_2(.) \right| dp(z)$$

$$\leq \int_R | h_1(.) - h_2(.) | \, dp(z) + \frac{P(H_1^D)}{1-P(H_1^D)} \int_R h_1(.) dp(z) + \frac{P(H_2^D)}{1-P(H_2^D)} \int_R h_2(.) dp(z)$$

$$= \int_R | h_1(.) - h_2(.) | \, dp(z) + \mathrm{P}(\mathrm{H}_1^{\mathrm{D}}) + \mathrm{P}(\mathrm{H}_2^{\mathrm{D}}),$$

where the last equality is by Assumption A2. By (A5) ~ (A7)

$$\text{(A8)} \quad \int_{H_1 \cup H_2} | h_1^{'}(.) - h_2^{'}(.) | \, dp(z) \leq \int_{H_1 \cup H_2} | h_1(.) - h_2(.) | \, dp(z).$$

Since the true PDSI equals IDSI by Proposition 1, the observed PDSI is no larger than IDSI. Q.E.D.

For the country pairs consisting of one rich country and one poor country, Proposition A2 implies that missing domestic sales data is likely to compress the variation in PDSI relative to IDSI, resulting in $\beta_P < 1$ in regression (17).

4. General Form of Quantity Choice

To minimize notation, consider the setup with identical technologies of section 2.1. The derivations below can be easily extended to the cases of different technologies across supplier countries or multiple differentiated goods. Let $f_q(I)$ be the quantity of the differentiated good consumed by *each* consumer with income I. In country 1, the number of people with income I_0 is $N_1 g_1(I_0) dI$ and they pay the price aI_0 (see equation (4)) for $N_1 f_q(I_0) g_1(I_0) dI$ units of the differentiated good. Since the total quantity of consumption by country 1 is $N_1 Q_1$ with $Q_1 \equiv \int_{G_1} f_q(I) g_1(I) dI$, the probability mass for country 1's price distribution at price aI_0 (i.e. the fraction of the differentiated good with price aI_0) is $h_1(aI_0)dp(z)$ where

$$\text{(A9)} \quad h_1(p(z)) = \frac{1}{Q_1} f_q\left(\frac{p(z)}{a}\right) g_1\left(\frac{p(z)}{a}\right) \cdot \frac{1}{a}.$$

Thus $h_1(.)$ is the pdf of country 1's price distribution and its support is $H_1 = f_p(G_1)$. Likewise, country 2 has the pdf $h_2(p(z)) = \frac{1}{Q_2} f_q\left(\frac{p(z)}{a}\right) g_2\left(\frac{p(z)}{a}\right) \cdot \frac{1}{a}$ with support $H_2 = f_p(G_2)$ and $Q_2 \equiv \int_{G_2} f_q(I) g_2(I) dI$.

Equation (A9) is equation (5), country 1's price distribution in section 2.1, augmented by $f_q(.)/Q_1$, which represents quantity weights. Weighting is necessary because quantity differs across prices and larger weights go to the prices with larger quantities. If $f_q(I) = 1$ (or a constant), $f_q(.) = Q_1$, the same quantity weight goes to all prices and equation (A9) is the same as equation (5). If $f_q(.)$ increases in income, so does $f_q(.)/Q_1$ (Q_1 is a constant); larger quantity weights go to higher prices and (A9) has a fatter right tail than (5).

On the other hand, the IDSI for countries 1 and 2 is still by equation (6) but the PDSI becomes

(A10) $PDSI(.) = \frac{1}{2}\int_{\underline{H}} \left| \frac{1}{Q_1} f_q(\frac{p(z)}{a}) g_1\left(\frac{p(z)}{a}\right) - \frac{1}{Q_2} f_q(\frac{p(z)}{a}) g_2\left(\frac{p(z)}{a}\right) \right| \frac{1}{a} dp(z)$,

where $\underline{H} = f_p(\underline{G})$. Unfortunately, it is hard to derive the exact relationship between IDSI and PDSI because it depends on how $g_1(.)$, $g_2(.)$ and $f_q(.)$ vary and how they interact with each other. We use two numerical examples to illustrate this point below.

Suppose that the probabilities of I_1 (low income) and I_2 (high income) are both 0.5 for country 1 and they are 0.6 and 0.4, respectively, for country 2. Then IDSI = 0.2. Further, suppose that $f_q(I_1) = 1$. When $f_q(I_2) = 1.4$, PDSI = 0.2011 > IDSI, but when $f_q(I_2) = 1.8$, PDSI = 0.1948 < IDSI. Thus the general form of quantity choice may either strengthen or weaken the correlation between IDSI and PDSI and imply $\beta_P > 1$ or $\beta_P < 1$ in regression (17).

Appendix 2 Data

1. Income Data

From the LIS we extract disposable household income (DPI), a commonly used measure in the analysis of income inequality. DPI includes monetary income after direct taxes and transfer payments. The data are in local currency values and we convert them to real US dollar values using the PPP data from Penn World Tables 6.1. DPI omits indirect taxes, benefits from public spending such as those from health care, education, or most housing subsidies, and wealth, except to the extent that it is represented by cash interest, rent, and dividends. The 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 that summarizes the wide range of scales in use:

$$\text{Adjusted Income} = \text{DPI} / \text{Household}\, Size^E .$$

The equivalence elasticity $E \in [0,1]$ represents economies of scale in household size. We employ the LIS Equivalence Scale (E = 0.5), a commonly used scale among researchers who study income inequality using the LIS data (e.g. Atkinson et al. 1995). An alternative popular approach explicitly employs data on the numbers of adults and children in the household. This approach is only feasible for a limited subset of our data.

Once we have calculated adjusted income from each household in the survey, we calculate the level of income for households at each percentile of the distribution.

Some of our countries have one (occasionally, two) wave of data missing. The literature has shown that quantile income levels within a country tend to follow smooth trends over time (e.g. Dollar and Kraay 2002, Sala-i-Martin 2005) and this finding is consistent with the patterns we see in the LIS data. Accordingly, for each un-starred country with missing data, we estimate linear income trends by percentile using the data of the available waves and then extrapolate the data for the missing wave(s).

Specifically, we fit linear trends for Australia, France, Italy, Mexico and Switzerland, and their average R^2 is 0.87. We fit log linear trends for Austria, Belgium, Finland, Ireland, Luxemburg, Netherlands, Poland and Spain, and their average R^2 is 0.91. We do not extrapolate Hungary's wave 1 data due to poor fit (and Hungary's wave 1 trade data do not exist anyway), nor do we extrapolate Russia's data before wave 3 because Russia was part of the Soviet Union.

2. Miscellaneous Issues

The raw income data for Russia 1992 have an unreliable scale (the median income is $2569329) and so do the raw data for Israel 1979, Poland 1986 and Russia 1995. We rescale these data in the following way. For each country, we find a wave with a reliable scale (wave t) and calculate the ratio of this country's median income (Y_{50}) to its real per capita GDP in PWT 6.1 (CGDP) for wave t. We then impute the median income for wave s as $\hat{Y}_{50,s} = CGDP_s \times (Y_{50,t}/CGDP_t)$ and impute the *b*th percentile income (Y_b) for wave s as $\hat{Y}_{b,s} = Y_{b,s} \times (\hat{Y}_{50,s}/Y_{50,s})$. For Russia we use 2000 as wave t and for Israel and Poland we use 1986 and 1992, respectively.

The WTF does not have trade quantity data for wave 1 (1980) and so we use the UN trade data instead. However, the WTF data use the 4-digit SITC Rev. 2 classification whereas the UN data use 5-digit SITC Rev. 1. We follow Feenstra et al. (2005) and concord the SITC Rev. 1 codes to SITC Rev. 2 in the following way (this concordance is available from us upon request). First, the 5-digit Rev. 1 codes are truncated to the 4-digit level. Second, the 4-digit Rev.1 codes are matched to the corresponding 4-digit Rev.2 codes using the maximum count of the Rev.2 frequency. In a tie, the maximum is given to the first 4-digit Rev.2 code listed numerically. Finally, if many Rev.1 codes are matched to a single Rev.2 code we aggregate these Rev.1 codes.

Figure 1 Quality Choice with Identical Technologies

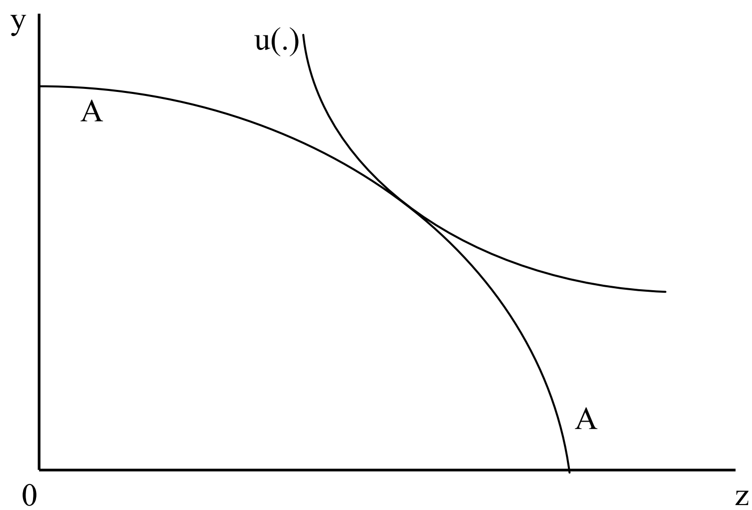


Figure 2 Quality Choice with Different Technologies

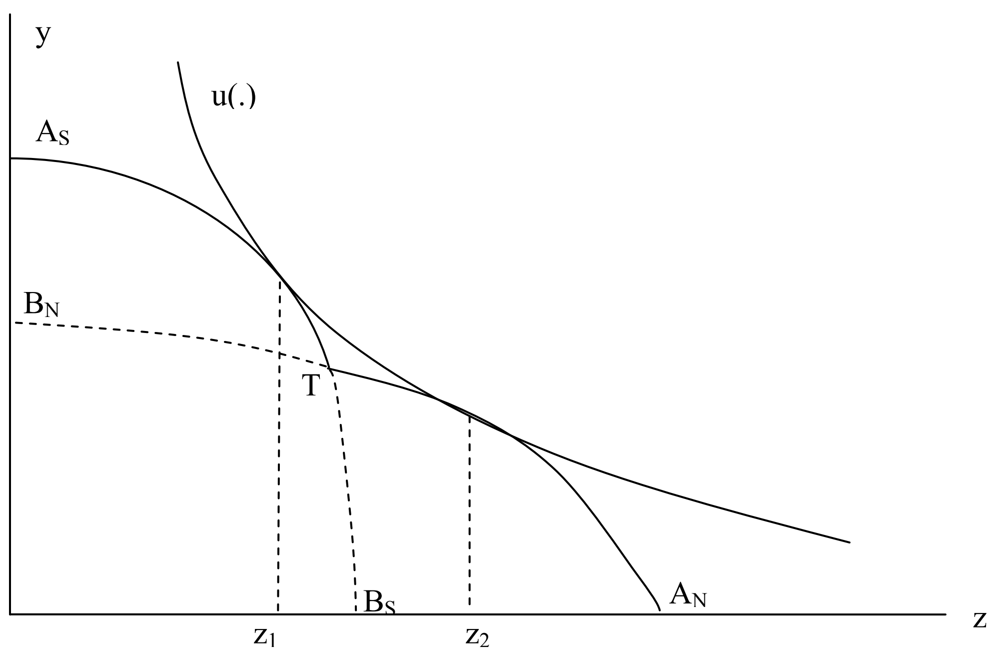
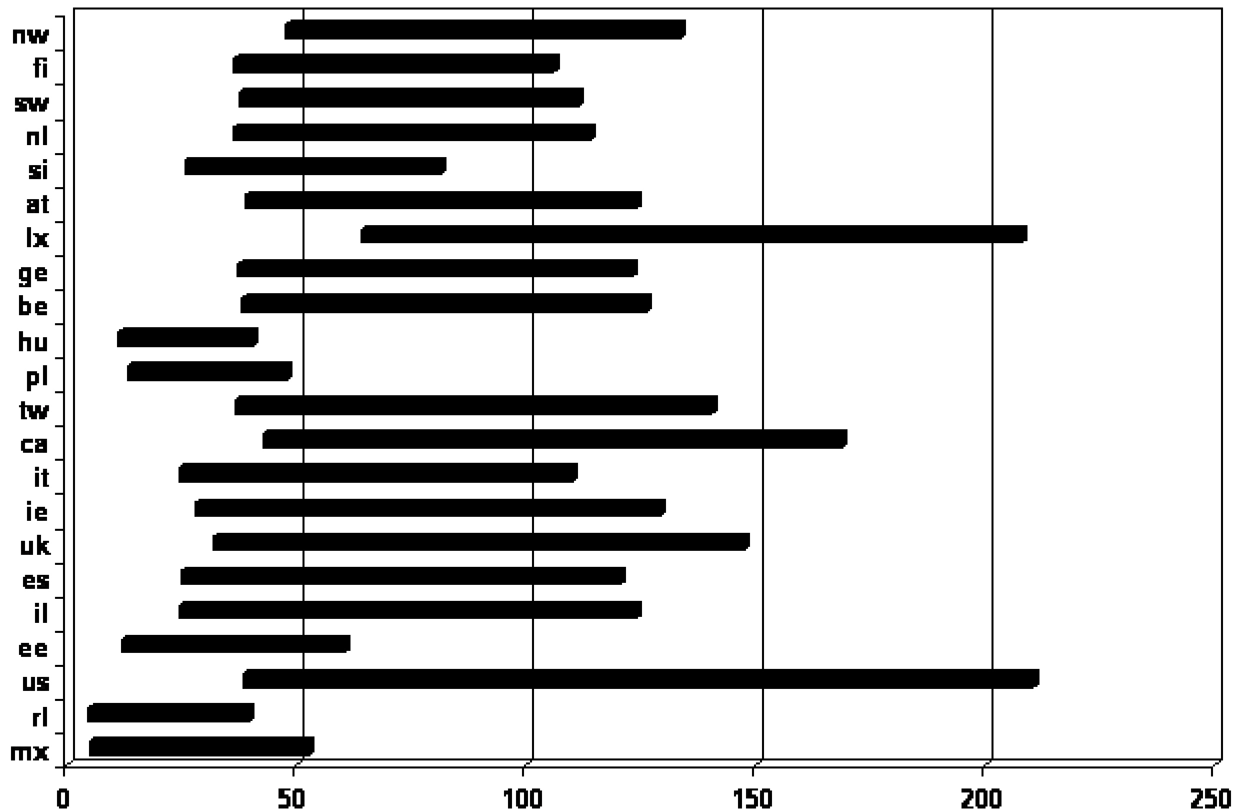


Figure 3 The Range of Incomes for Wave 5

Year	Country	Multilateral IDSI	Decile Ratio P90/P10
2000	Norway	34.8	2.8
2000	Finland	33.8	2.9
2000	Sweden	32.2	3.0
1999	Netherlands	30.6	3.1
1999	Slovenia	37.6	3.2
2000	Austria	29.9	3.2
2000	Luxembourg	42.8	3.2
2000	Germany	28.6	3.3
2000	Belgium	28.0	3.3
1999	Hungary	53.4	3.6
1999	Poland	49.6	3.6
2000	Taiwan	23.8	3.8
2000	Canada	28.3	3.9
2000	Italy	21.7	4.5
2000	Ireland	20.6	4.6
1999	United Kingdom	17.8	4.6
2000	Spain	19.0	4.8
2001	Israel	16.6	5.0
2000	Estonia	40.3	5.1
2000	United States	29.5	5.4
2000	Russia	53.6	8.4
2000	Mexico	47.3	10.4
	Average	**32.7**	**4.3**



Notes: The data are normalized using the U.S. median income ($24,094 = 100). P10 and P90 are the 10th and 90th percentile incomes, respectively and the lengths of the bars represent the gap between P10 and P90.

Figure 4 The Average Multilateral PDSI and Multilateral IDSI for Wave 2

Figure 5 The Average Bilateral PDSI and Bilateral IDSI for Wave 2

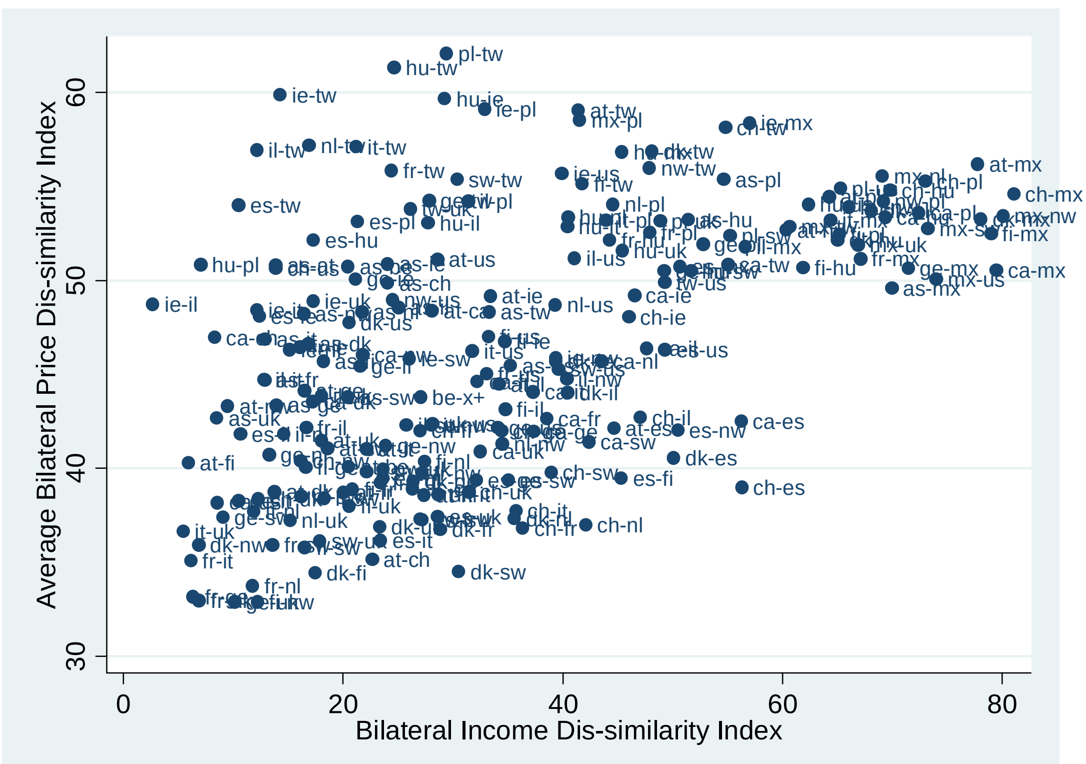


Table 1 The Coverage of the LIS Income Data

Country	Abbre-viation	Wave 1 around 1980	Wave 2 around 1985	Wave 3 around 1990	Wave 4 around 1995	Wave 5 around 2000
Australia	AS	1981	1985	1989	1994	2000(e)
Austria	AT	1981(e)	1987	1990(e)	1995	2000
Belgium	BE	1981(e)	1985	1992	1995	2000
Canada	CA	1981	1987	1991	1994	2000
Czech Republic*	CZ	.	.	1992	1996	.
Denmark*	DK	.	1987	1992	.	.
Estonia*	EE	.	.	.	.	2000
Finland	FI	1981(e)	1987	1991	1995	2000
France	FR	1981	1984	1989	1994	1999(e)
Germany	GE	1981	1984	1989	1994	2000
Greece*	GR	.	.	.	1995	2000
Hungary*	HU	.	1985(e)	1991	1994	1999
Ireland	IE	1981(e)	1987	1990(e)	1995	2000
Israel	IL	1979(a)	1986	1992	1997	2001(b)
Italy	IT	1981(e)	1986	1991	1995	2000
Luxembourg	LX	1981(e)	1985	1991	1994	2000
Mexico	MX	1981(e)	1984	1989	1996	2000
Netherlands	NL	1981(e)	1987	1991	1994	1999
Norway	NW	1979	1986	1991	1995	2000
Poland	PL	1981(e)	1986(a)	1992	1995	1999
Romania*	RO	.	.	.	1995	.
Russia*	RL	.	.	1992(a)	1995(a)	2000
Slovak Republic*	SK	.	.	1992	1996	.
Slovenia*	SI	.	.	.	1997	1999
Spain	ES	1980	1985(e)	1990	1995	2000
Sweden	SW	1981	1987	1992	1995	2000
Switzerland	CH	1982	1985(e)	1992	1995(e)	2000
Taiwan	TW	1981	1986	1991	1995	2000(b)
United Kingdom	UK	1979	1986	1991	1995	1999
United States	US	1979	1986	1991	1994	2000

Notes: In each cell is the actual year for which the LIS income data are available for a given country and a given wave. A "(e)" indicates that the data have been extrapolated using the available data of the same country. A "(a)" indicates that the data have been rescaled because the scale of the raw data is unreliable. A "(b)" indicates that an adjacent year's PPP data have been used to convert the raw data into US dollars. The countries without a "*" have income data for all the five waves.

Table 2 Summary Statistics

Variables	Mean/Std. Deviation	Number of Obs.
IDSI(c,W)	2.868	85,133
ln IDSI(c,W)	9.631	85,133
IDSI(c,c')	1.539	900,617
ln IDSI(c,c')	4.397	900,617
PDSI(c,W)	1.833	74,212
PDSI(c,c')	1.783	770,039
N(c)	1.047	85,133
ln N(c)	-0.084	76,622
N(c,c')	0.906	899,252
ln N(c,c')	0.101	741,813

Notes: "IDSI" means income dis-similarity index and "PDSI" means price dis-similarity index. For IDSI and PDSI, "(c, W)" denotes the multilateral comparison (i.e. country c relative to the world) and "(c, c')" denotes the bilateral comparison (i.e. country c relative to c'). "N(c)" means the number of export partners for country c and "N(c,c')" means the number of common export partners for c and c'. "ln" means logs.

Table 3 The Multilateral Price Dis-similarity Index Regression

	Dependent Var: PDSI(c,W)				
	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
IDSI(c,W)	0.341***	0.452 ***	0.310 ***	0.271 ***	0.194 ***
	0.023	*0.015*	*0.017*	*0.012*	*0.013*
Constant	28.649 ***	27.082 ***	28.023 ***	28.589 ***	31.513 ***
	0.583	*0.426*	*0.536*	*0.426*	*0.412*
Number of Obs.	9,288	14,840	15,225 ***	17,789	17,070
Adjusted R²	0.25	0.28	0.28	0.28	0.28
IDSI(c,W)	0.130 ***	0.371 ***	0.147 ***	0.075 ***	0.023 *
	0.029	*0.016*	*0.018*	*0.014*	*0.014*
ln GDP(c)	-2.379 ***	-2.5114 ***	-3.244 ***	-2.894 ***	-3.360 ***
	0.205	*0.162*	*0.119*	*1.070*	*0.099*
ln MP(c)	0.317 ***	-0.250 ***	0.076 *	0.049 ***	0.177 ***
	0.067	*0.066*	*0.044*	*1.183*	*0.034*
Constant	58.956 ***	61.846 ***	72.811 ***	70.469 ***	77.028 ***
	2.696	*1.727*	*1.735*	*2.240*	*1.374*
Number of Obs.	9,288	14,840	15,225	17,789 ***	17,070
Adjusted R²	0.26	0.31	0.32	0.31	0.33

Notes: This table reports the results of regression (22), with (the bottom panel) and without (the top panel) the additional explanatory variables that control for market size and trade costs. We do not report the product fixed effects to save space. Standard errors are in italics. "***", "**" and "*" indicate significance levels of 1%, 5% and 10%, respectively.

Table 4 The Bilateral Price Dis-similarity Index Regression

	Dependent Var: PDSI(c,c')				
	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
IDSI(c,c')	0.085 ***	0.229 ***	0.151 ***	0.091 ***	0.075 ***
	0.005	*0.003*	*0.003*	*0.002*	*0.002*
Constant	41.963 ***	38.279 ***	39.344 ***	40.538 ***	41.461 ***
	0.182	*0.125*	*0.098*	*0.086*	*0.087*
Number of Obs.	73,977	142,038	151,483 ***	209,547	192,994
Adjusted R^2	0.21	0.20	0.22	0.21	0.22
IDSI(c,c')	0.069 ***	0.188 ***	0.123 ***	0.077 ***	0.051 ***
	0.005	*0.003*	*0.003*	*0.002*	*0.002*
GDPgap(c,c')	1.201 ***	0.5526 ***	1.206 ***	0.920 ***	1.809 ***
	0.076	*0.056*	*0.052*	*0.040*	*0.038*
ln Dist(c,c')	2.964 ***	2.869 ***	2.867 ***	2.306 ***	2.356 ***
	0.072	*0.055*	*0.049*	*0.042*	*0.045*
Constant	17.809 ***	16.364 ***	15.941 ***	21.706 ***	20.878 ***
	0.574	*0.424*	*0.390*	*0.334*	*0.361*
Number of Obs.	73,977	142,038	151,483	209,547	192,994
Adjusted R^2	0.23	0.21	0.24	0.23	0.24

Notes: This table reports the results of regression (20), with (the bottom panel) and without (the top panel) the additional explanatory variables that control for market size and trade costs. "GDPgap(c, c')" is the absolute value of the log difference in GDP for countries c and c'. We do not report the product fixed effects to save space. Standard errors are in italics. "***", "**" and "*" indicate significance levels of 1%, 5% and 10%, respectively.

Table 5 The Number of Export Partners Regression

	Dependent Var: ln N (c)				
	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
ln IDSI(c,W)	-0.280 ***	-0.466 ***	-0.225 ***	-0.114 ***	-0.116 ***
	0.020	*0.014*	*0.016*	*0.014*	*0.014*
ln GDP(c)	0.171 ***	0.177 ***	0.275 ***	-0.020 ***	0.282 ***
	0.006	*0.005*	*0.003*	*0.033*	*0.003*
ln MP(c)	0.004 *	0.035 ***	-0.001	0.338 ***	0.002 **
	0.002	*0.002*	*0.002*	*0.036*	*0.001*
Constant	-1.230 ***	-1.042 ***	-2.730 ***	-3.703 ***	-3.243 ***
	0.113	*0.077*	*0.077*	*0.089*	*0.067*
Number of Obs.	9,596	15,350	15,699	18,333	17,644
Adjusted R^2	0.26	0.35	0.37	0.42	0.43
	Dependent Var: N (c)				
IDSI(c,W)	-0.017 ***	-0.016 ***	-0.008 ***	-0.004 ***	-0.004 ***
	0.001	*0.001*	*0.001*	*0.001*	*0.001*
ln GDP(c)	0.200 ***	0.368 ***	0.312 ***	-0.772 ***	0.297 ***
	0.006	*0.010*	*0.006*	*0.043*	*0.004*
ln MP(c)	-0.004 *	0.012 ***	0.001	1.228 ***	0.004 **
	0.002	*0.004*	*0.002*	*0.048*	*0.002*
Constant	-0.856 ***	-3.017 ***	-2.604 ***	-4.664 ***	-2.555 ***
	0.085	*0.106*	*0.085*	*0.092*	*0.060*
Number of Obs.	10,314	18,414	17,160	20,020	19,225
Adjusted R^2	0.25	0.13	0.15	0.30	0.23

Notes: This table reports the results of regression (21) for the log linear (the top panel) and linear (the bottom panel) specifications. We do not report the product fixed effects to save space. Standard errors are in italics. "***", "**" and "*" indicate significance levels of 1%, 5% and 10%, respectively.

Table 6 The Number of Common Export Partners Regression

	Dependent Var: ln N(c,c')				
	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
ln IDSI(c,c')	-0.173 ***	-0.280 ***	-0.224 ***	-0.253 ***	-0.176 ***
	0.003	*0.003*	*0.002*	*0.002*	*0.002*
GDPgap(c,c')	-0.024 ***	0.0430 ***	0.018 *	-0.023 ***	-0.083 ***
	0.002	*0.001*	*0.001*	*0.001*	*0.001*
ln Dist(c,c')	-0.106 ***	-0.104 ***	-0.100 ***	-0.014 ***	-0.005 ***
	0.002	*0.001*	*0.001*	*0.001*	*0.001*
Constant	1.554 ***	1.771 ***	1.522 ***	1.020 ***	0.794 ***
	0.016	*0.012*	*0.011*	*0.010*	*0.010*
Number of Obs.	73,825	134,582	146,247	201,172	185,987
Adjusted R^2	0.20	0.30	0.28	0.26	0.25
	Dependent Var: N(c,c')				
IDSI(c,c')	-0.013 ***	-0.015 ***	-0.011 ***	-0.012 ***	-0.008 ***
	0.000	*0.000*	*0.000*	*0.000*	*0.000*
GDPgap(c,c')	-0.034 ***	0.0263 ***	-0.002	-0.024 ***	-0.077 ***
	0.003	*0.003*	*0.002*	*0.002*	*0.002*
ln Dist(c,c')	-0.136 ***	-0.197 ***	-0.177 ***	-0.077 ***	-0.082 ***
	0.003	*0.003*	*0.002*	*0.002*	*0.002*
Constant	2.615 ***	3.109 ***	2.844 ***	2.136 ***	2.163 ***
	0.022	*0.020*	*0.018*	*0.016*	*0.020*
Number of Obs.	86,658	174,016	172,808	242,187	223,583
Adjusted R^2	0.08	0.10	0.09	0.07	0.04

Notes: This table reports the results of regression (19) for the log linear (the top panel) and linear (the bottom panel) specifications. “GDPgap(c, c’)” is the absolute value of the log difference in GDP for countries c and c’. We do not report the product fixed effects to save space. Standard errors are in italics. “***”, “**” and “*” indicate significance levels of 1%, 5% and 10%, respectively.

Table 7 The Results for Consumption Goods

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
Multilateral PDSI, Reg. (22)					
	-0.008	0.420 ***	0.121 ***	0.100 ***	0.093 ***
	0.063	*0.036*	*0.039*	*0.031*	*0.028*
Bilateral PDSI, Reg. (20)					
	0.068 ***	0.268 ***	0.179 ***	0.122 ***	0.094 ***
	0.012	*0.008*	*0.006*	*0.004*	*0.004*
Multilateral N, Reg. (21)					
	-0.241 ***	-0.546 ***	-0.153 ***	-0.242 ***	-0.261 ***
	0.043	*0.033*	*0.035*	*0.032*	*0.029*
Bilateral N, Reg. (19)					
	-0.205 ***	-0.385 ***	-0.260 ***	-0.350 ***	-0.244 ***
	0.008	*0.007*	*0.005*	*0.004*	*0.004*

Notes: This table reports the coefficients on the income dis-similarity index, IDSI, and their standard errors when regressions (19)-(22) are run for consumption goods only. We use the UN Broad Economic Classification system to designate which products are consumption goods. Standard errors are in italics. "***", "**" and "*" indicate significance levels of 1%, 5% and 10%, respectively.

Table 8 The Results Using the 90-10 Decile Ratio

		Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
Multilateral PDSI, Reg. (22)						
	P90/P10(c)	5.455 ***	3.324 ***	1.776 ***	1.038 ***	1.132 ***
		0.276	*0.141*	*0.106*	*0.101*	*0.076*
	ln GDP(c)	-4.046 ***	-5.368 ***	-4.001 ***	-6.170 ***	-3.646 ***
		0.167	*0.165*	*0.113*	*1.063*	*0.091*
	ln MP(c)	0.714 ***	0.423 ***	0.091 ***	3.112	0.222 ***
		0.065	*0.065*	*0.043*	*1.182*	*0.032*
	Constant	59.202 ***	85.276 ***	79.913 ***	70.947	75.928 ***
		1.777	*1.473*	*1.427*	*2.105*	*1.119*
	Number of Obs.	9,288	14,840	15,225	17,789	17,070
	Adjusted R^2	0.29	0.31	0.33	0.32	0.34
Multilateral N, Reg. (21)						
	P90/P10(c)	-0.186 ***	-0.150 ***	-0.094 ***	-0.083 ***	-0.065 ***
		0.009	*0.008*	*0.005*	*0.004*	*0.003*
	ln GDP(c)	0.317 ***	0.494 ***	0.352 ***	-0.530 ***	0.321 ***
		0.005	*0.010*	*0.006*	*0.043*	*0.004*
	ln MP(c)	-0.026 ***	-0.018 ***	0.000	0.997 ***	-0.001
		0.002	*0.004*	*0.002*	*0.048*	*0.001*
	Constant	-1.814 ***	-4.024 ***	-2.994 ***	-4.566 ***	-2.654 ***
		0.056	*0.092*	*0.071*	*0.086*	*0.049*
	Number of Obs.	10,314	18,414	17,160	20,020	19,225
	Adjusted R^2	0.25	0.14	0.16	0.31	0.25

Notes: This table reports the results of regressions (21) and (22) with the income dissimilarity index, IDSI, replaced by the decile ratio. Standard errors are in italics. "***", "**" and "*" indicate significance levels of 1%, 5% and 10%, respectively.