

Diversification Cones, Trade Costs and Factor Market Linkages

Chong Xiang*

Purdue University

Abstract: This paper finds non-uniform differences in the distribution functions of factor usage intensities among 10 rich OECD countries. The 10 countries form three distinct groups such that the between-group differences are more pronounced than within-group differences and capital-abundant countries are in capital-abundant groups. The estimation works even if the same industry codes represent different goods across countries in the data. The finding is consistent with the multiple-cone factor proportions theory with zero trade costs, with each group being one cone. An alternative interpretation is non-zero trade costs. Both interpretations imply weak factor market linkages between the countries in different groups.

JEL Classification: F1

Key words: diversification cone; cumulative distribution function (cdf); stochastic dominance test; trade cost; Kolmogorov-Smirnov (KS) statistics.

* Purdue University, Economics Department, KCRN, 403 West State Street, West Lafayette, IN 47907-2056, USA. Tel.: +1-765-494-4499; fax: +1-765-494-9658. Email address: cxiang@purdue.edu (C. Xiang). The author thanks Jason Abrevaya, Alan Deardorff, Jonathan Eaton, Juan Carlos Hallak, Gordon Hanson, David Hummels and two anonymous referees, as well as seminar participants at Purdue University, University of Texas at Austin, the Spring 2003 Midwest International Economics Meeting and 2003 EIIT for their very helpful comments and suggestions.

1 Introduction

The factor proportions theory, which includes both the Heckscher-Ohlin and monopolistic competition models, is an important element of international trade. Yet the one-cone version of the theory, which assumes worldwide factor price equalization (FPE) and that all countries produce the same set of goods, finds poor empirical support (e.g. Leamer and Levinsohn 1997). Not surprisingly, the literature has shifted its attention to the multiple-cone version of the theory (e.g. Schott 2003), in which countries produce the same set of goods and achieve FPE only if they are in the same diversification cone.

The study of cones is important. Within a cone, the Heckscher-Ohlin-Vanek theorem generates sharp predictions about the factor content of trade (Davis and Weinstein 2001), the Rybczynski mechanism allows countries to absorb immigration by changing their output mix without changing their factor prices (Hanson and Slaughter 2002), and a small open economy can grow and accumulate factors while the commodity prices remain constant so that diminishing returns to capital investment do not set in (Ventura 1997). Within a cone, factor prices are equalized and the equilibrium is as efficient as if factors could move freely, so that there is no incentive to exploit cross-country differences in factor prices via outsourcing (Feenstra and Hanson 1997) or cross-border FDI (Debaere 2002). Furthermore, as countries in different cones produce different goods, shocks from inside the cone might affect a country's factor prices through the Stolper-Samuelson mechanism while shocks from outside the cone might have little impact on the factor prices. Such shocks include trade liberalization (Leamer 1996), technological progress (Collins 1985) and factor accumulation.

However, empirical studies on diversification cones (e.g. Debaere and Demiroglu 2003,

Schott 2003)¹ face three challenges. First, the literature often assumes identical technologies across countries, but there is strong evidence that technology differences are an important determinant of international trade (e.g. Harrigan 1997, Eaton and Kortum 2002). Second, the literature uses a cross-country panel of industry data, but the same industry codes might represent different goods for different countries (Maskus 1991). This is particularly important for countries in different cones because trade theory predicts that they produce different goods (Schott 2003). For example, while “textile and apparel” made in the U.S. include high-end garments and outdoors-wear incorporating the latest technological advances in synthetic fiber, those made in Bangladesh are likely to be low-end T-shirts and linen sheets. Finally, the literature interprets the findings in the data under the zero-trade-cost assumption, but trade frictions abound (e.g. Anderson and van Wincoop 2004). How do non-zero trade costs affect the interpretation of the data?

To tackle the first two challenges,² this paper performs bilateral comparisons of the cumulative distribution functions (cdf’s) of factor usage intensities using the non-parametric Kolmogorov-Smirnov statistics. The use of factor usage intensities accommodates Hicks-neutral cross-industry productivity differences across countries and the cdf’s can be calculated from readily available data on factor uses by industry. On the other hand, the cdf’s are not affected by industry identities and so the estimation works if the same industry codes represent different goods across countries, or even if the industry codes of different countries come from different industry classification systems. Furthermore, the cdf’s assign equal

¹ See also Repetto and Ventura (1997), Bernard and Schott (2002), Hanson and Slaughter (2002) and Bernard, Redding, Schott and Simpson (2003).

² Schott (2003) addresses the second issue by constructing “HO Aggregates” from the data, but cautions that it “stretches industry-level data about as far as it can go”.

frequencies to all industries regardless of their outputs. This removes the Rybczynski effect and implies that the countries in the same cone have the same cdf, even though capital-abundant countries produce larger quantities of capital-intensive goods. The countries in different cones, however, have different cdf's since they have different factor usage intensities. The bilateral comparisons then reveal whether the cdf's differ for every country pair.

The finding is that among 10 rich OECD countries, the differences are non-uniform. The 10 countries form three distinct groups (with, e.g. U.K., France and U.S. in different groups) such that the differences are more pronounced across groups than within the same group and capital abundant countries are in capital abundant groups. This pattern is confirmed by the regression analysis. If trade costs are assumed to be zero, this finding is consistent with the multiple cone factor proportions theory and each of the three groups can be interpreted as one cone.

To tackle the challenge of non-zero trade costs, this paper uses simulations to show that both multiple cones and trade costs induce different factor prices and different factor usage intensities across countries. Thus trade costs are an alternative interpretation of the differences in the cdf's of factor usage intensities. What's more, as compared with the single-cone case in which FPE prevails and the countries' factor markets are fully integrated, both multiple cones and trade costs imply weak factor market linkages, in the sense that differences in factor endowments across countries lead to differences in factor prices. Therefore, whether the interpretation of the findings is multiple cones or trade costs, the message is the same: factor market linkages are weak between the countries in different groups.

In what follows, section 2 discusses the theory of diversification cones and sections 3 ~ 4

identify them empirically. Trade costs are assumed to be zero in sections 2 ~ 4. Section 5 uses simulations to show that both multiple cones and non-zero trade costs imply weak factor market integration and section 6 concludes.

2 The Theory of Diversification Cones

Assume that trade costs are zero. This assumption is maintained throughout sections 2 ~ 4 and is relaxed in section 5. Under this assumption, Figure 1 illustrates the intuition of diversification cones using a Heckscher-Ohlin model with 2 factors, 4 homogeneous goods and 3 countries. The curves 1 ~ 4 are the isoquants of one dollar's worth of outputs (unit value isoquants) for goods 1 ~ 4 and they are tangent to the unit cost lines HH and FF due to cost minimization. The slopes of the rays connecting the tangency points and the origin, O1 ~ O4, equal the capital intensities (ratios of capital uses to labor uses) of goods 1 ~ 4. The slopes of the rays OA ~ OC equal the capital abundance (ratios of capital endowments to labor endowments) of countries A ~ C. There are 2 diversification cones, one formed by O1 and O2, and the other by O3 and O4. Countries A and C are in the same cone and have the same wage-rental ratio (represented by the slope of the unit cost line FF). Thus they produce goods 3 and 4 with the same factor usage intensities.

(T1) The countries in the same cone have the same factor usage intensities.

On the other hand, countries A and B are in different cones and the capital-abundant country B has a higher wage-rental ratio (as represented by the steeper unit cost line HH). Since goods 1 ~ 4 are homogeneous, country B produces only the capital-intensive goods 1 and 2. If, however, goods 1 ~ 4 are differentiated (e.g. by their country of origin due to Armington preferences), country B might also produce good 3. Figure 2 illustrates the differentiated-good case. It is similar to Figure 1 except for the new unit-value isoquant 3'. 3'

is tangent to country B's unit cost line HH and represents the varieties of good 3 produced in country B. Because the substitution elasticity between the varieties of good 3 is finite, although country B's varieties use more capital per worker and sell at higher prices than country A's varieties (which are represented by the unit value isoquant 3), the demand for country B's varieties remains non-zero. In both the homogeneous- and differentiated-good cases,

(T2) The countries in a capital-abundant cone use more capital per worker in every good than the countries in a capital-scarce cone.

When there are many goods and many countries, (T1) and (T2) still hold for both the Heckscher-Ohlin (e.g. Deardorff 1979) and monopolistic competition (e.g. Helpman and Krugman 1986) models. To formalize (T1) and (T2), consider the following many-country-many-good setup that incorporates the production side of both the Heckscher-Ohlin and monopolistic-competition models. Production uses two factors, capital (K) and labor (L), and can have constant returns to scale or homothetic scale economies.³ An industry's product can be homogeneous or differentiated. Firms minimize costs and do not interact strategically. Within a country, all firms are symmetric and face the same set of factor prices. Across countries, there can be Hicks-neutral cross-industry productivity differences. The total cost of a firm in industry j and country c is:

$$\pi_c \pi_j c(\omega_c; q_{jc}; \theta_j)$$

π_c and π_j are country- and industry-specific Hicks-neutral productivities and they take the

³ The 2-factor setup is dictated by data constraints and the estimation in sections 3 and 4 could also be done for the multi-factor case. The monopolistic competition model exhibits homothetic scale economies when the fixed cost and the variable cost have the same factor usage intensities.

multiplicative form. This specification precludes productivity differences that are by country by industry. ω_c is the vector of factor prices that prevail in country c . q_{jc} is the industry output and represents the effect of scale economies. θ_j is the set of production technique parameters of industry j (e.g. the substitution elasticity in production) and does not vary across countries by assumption. Assume that $c(\cdot)$ is differentiable. Then the capital and labor uses, k and l , are the partial derivatives of the cost function with respect to factor prices, and with homothetic scale economies, the capital-labor ratio of industry j in country c does not depend on q_{jc} :

$$\log(k_{jc}/l_{jc}) = f(\omega_c; \theta_j) \quad (1)$$

Note that “capital-labor ratios” refer to $\log(k/l)$ and the Hicks-neutral productivities π_j and π_c do not affect the capital-labor ratios.

Pick two countries, A and B. Country B is relatively capital abundant. First, arrange the capital-labor ratios, $\log(k/l)$, of the population of country A’s industries 1...J in ascending order. Assign cumulative frequency $1/J$ to the least capital-intensive industry, $2/J$ to the second least capital-intensive industry...and $J/J = 1$ to the most capital-intensive industry. The relation between the cumulative frequencies and capital-labor ratios is country A’s cumulative distribution function (cdf) of $\log(k/l)$. Denote it by $F_A(x)$, where x represents $\log(k/l)$. Now generate $F_B(\cdot)$, country B’s cdf of capital-labor ratios. If A and B are in different cones, (T2) implies that at every cumulative frequency $1/J, 2/J \dots J/J$, the capital-labor ratio for country A is lower than for country B. If A and B are in the same cone, (T1) implies that at every cumulative frequency $1/J, 2/J \dots J/J$, the capital-labor ratios for countries A and B are the same. Thus:

Proposition 1 If country B is in a more capital-abundant cone than country A, $F_B(x) < F_A(x) \forall x$; i.e. F_B first-order stochastically dominates (FOSD) F_A for all the values of capital-labor

ratios. If countries A and B are in the same cone, F_A and F_B are the same distribution function.

Figure 3 illustrates Proposition 1. The horizontal axis is the capital-labor ratio and the vertical axis is the cumulative frequency. When countries A and B are in different cones, F_B lies to the right of F_A .

Proposition 1 might seem surprising. When countries A and B are in the same cone, the capital-abundant country B produces larger quantities of capital-intensive goods due to the Rybczynski effect. Why does F_B not lie to the right of F_A ? Because all industries are assigned equal frequencies and so industry outputs do not affect the distribution functions. To see this, suppose the output of the most capital-intensive industry in country A increases, the output of the least capital-intensive industry decreases and the capital-labor ratios of both industries are unchanged. Then the positions of these two industries on F_A do not change, and neither does the distribution function F_A itself. Therefore, F_A and F_B are the same distribution function as long as countries A and B are in the same cone, even if country B produces larger quantities of capital-intensive goods than country A.

To take Proposition 1 to the data, add an error term to equation (1):

$$\log(k_{jc}/l_{jc}) = f(\omega_c; \theta_j) + \varepsilon_{jc} \quad (2)$$

The error term ε_{jc} includes random measurement errors and idiosyncratic shocks that might prevent firms from minimizing costs exactly according to (1).⁴

3 The Non-parametric Estimation

3.1 The Kolmogorov-Smirnov (KS) Statistics

Proposition 1 suggests the following hypothesis testing to identify cones:

H_0 : F_A and F_B are the same distribution function (countries A and B are in the same cone)

⁴ This motivation is similar to that in cross section studies of cost functions (e.g. Berndt 1990).

H_1 : F_B FOSD F_A (country B is in a more capital-abundant cone)

As is usually the case with empirical work, the data does not reveal the population distribution functions F_A and F_B but their consistent estimates, the empirical distribution functions (empirical cdf's) G_A and G_B .⁵ The testing can be done by the following pair of one-sided Kolmogorov-Smirnov (KS) statistics:⁶

$$KS_{BA} = \sqrt{\frac{nm}{m+n}} \sup_x [G_A(x) - G_B(x)], \quad KS_{AB} = \sqrt{\frac{nm}{m+n}} \sup_x [G_B(x) - G_A(x)]$$

where n and m are the numbers of observations for countries A and B. The KS statistics look at the maximal distance between G_A and G_B . If KS_{BA} is large and statistically significant (i.e. it exceeds the 5% critical value), then there is strong evidence that F_B lies to the right of F_A (for some range of capital-labor ratios). If, in addition, KS_{AB} is statistically *insignificant*, then there is *no* strong evidence that F_A also lies to the right of F_B (for some other range of capital-labor ratios) and we can conclude that there is strong evidence that F_B FOSD F_A . Thus:

(T3) H_0 is rejected in favor of H_1 if KS_{BA} is statistically significant and KS_{AB} is statistically *insignificant*.

Since $G_A(\cdot) = G_B(\cdot) = 1$ eventually, the minimal value of the KS statistics is 0.

We need to look at both KS_{BA} and KS_{AB} because it is possible that both KS_{BA} and KS_{AB}

⁵ The consistency is by the Fundamental Theorem of Statistics (Davidson and MacKinnon 1993).

⁶ See Neave and Worthington (1988), Hollander and Wolfe (1999), McFadden (1989), Barrett and Donald (2000) and Abadie (2001). Another approach to testing differences in distribution functions uses analogues of the Pearson Goodness of Fit tests (PAT's) (e.g. Anderson 1996, Davidson and Duclos 2000). This approach partitions the range of the variable of interest into mutually exclusive and exhaustive categories (quantiles) and then exploits the joint asymptotic distribution of these quantiles. The partitioning needs to be fine enough to be representative of the underlying distribution. Since the sample size in this paper can be as small as 34 observations per distribution, the PAT approach is not used.

are statistically significant yet there is no first-order stochastic dominance (Proposition 1 requires that F_B FOSD F_A for all the values of capital-labor ratios). Figure 4 shows such a case, where F_A and F_B cross each other and F_B lies to the right of F_A for low values of capital-labor ratios but F_A lies to the right of F_B for high values of capital-labor ratios. In this case, if we look at KS_{BA} only we would draw the wrong conclusion that F_B FOSD F_A .

One advantage of the non-parametric estimation is that the empirical cdf's can be generated from readily available data on factor uses by industry. Another advantage is that the KS test compares two empirical cdf's and so remains valid when the same industry codes represent different goods for different countries in the data. To see this, suppose the least and most capital-intensive industries in country A swap their capital-labor ratios. Then their positions on G_A are swapped, but the empirical cdf G_A itself is unaffected. Since industry identities are irrelevant for the empirical cdf's, the industries used to generate G_A and G_B can be totally different industries and the estimation remains valid even if the industry codes of countries A and B are from different industry classification systems. Thus the non-parametric estimation could overcome a major obstacle in obtaining dis-aggregated internationally comparable data because industry-level, firm-level or even plant-level data from different countries could be used without first concurring them to the same classification system.

However, this advantage comes at a price. Suppose the same industry codes do represent the same goods across countries. Then because the KS test does not use this information, it tends to have less power than the tests that utilize this information, such as a regression analysis (e.g. Neave and Worthington 1988, Anderson 2001). That is to say, in this case, the non-parametric estimation might mistake the distributions of capital-labor ratios of "different sectors" that are not dissimilar enough as evidence of the same cone (i.e. it might fail to reject

H_0 when H_1 is true) and so tend to find too few cones. To address this issue I present the results of the regression analysis in section 4 as supplements of the non-parametric estimation. On the other hand, this possible bias towards finding too few cones implies that the KS test tends to be conservative and so when it does reject H_0 in favor of H_1 , this is really strong evidence for multiple cones.

The other advantages of the non-parametric estimation are as follows. First, the empirical cdf's assign equal frequencies to all industries regardless of their outputs and this removes the Rybczynski effect. To see this, suppose country A becomes more capital abundant but stays in the same cone and the Rybczynski effect kicks in: the capital intensive industries expand and the labor intensive industries contract while the capital-labor ratios of these industries remain unchanged. Then the positions of these industries on the empirical cdf G_A do not change and neither do their frequencies. Thus G_A does not change, either. In contrast, an estimation that gives more weight to larger industries might be biased towards finding too many cones due to the Rybczynski effect. Second, the exact distribution of the KS statistics in finite samples can be calculated, and this addresses the concern about applying asymptotic properties to finite samples. Finally, the KS statistics are non-parametric and so the underlying distributions do not need to be parameterized.

3.2 Data and Results

The data comes from Davis and Weinstein (2001) (DW 2001 henceforth) and has the outputs and direct and total (direct plus indirect) requirements for capital and labor of 34 sectors in 10 OECD countries. This dataset is constructed from the OECD's Input-Output, STAN and ISDB databases for 1985. The 34 sectors are listed in Table 1. See their paper for more detailed descriptions of the data.

Column 2 of Table 2 lists the capital abundance --- ratio of capital endowments to labor endowments --- of the 10 OECD countries in the data. Each country's empirical cdf can be constructed using the data on the total requirements for capital and labor. For the data to be consistent with multiple cones (with each cone having multiple countries), the empirical cdf's need to show non-uniform differences across country pairs: across cones, the empirical cdf's of capital-abundant countries lie to the right of capital-scarce countries while within a cone, the empirical cdf's exhibit no systematic pattern of FOSD. Figure 5 shows the empirical cdf's of France and Japan. The two curves cross each other and do not exhibit any pattern of FOSD. Figure 6 shows the empirical cdf's of France and Canada. Even though the difference in capital abundance is smaller between France and Canada than between France and Japan, the empirical cdf of Canada, the capital-abundant country, clearly lies to the right of France, the capital-scarce country. This suggests that the differences in empirical cdf's are non-uniform across country pairs.

To illustrate this pattern of the data, Table 3 reports the KS statistics for all the 90 country pairs (by T3 the order of the two countries that form a country pair matters). The countries are arranged down the rows and along the columns in ascending order of their capital abundance and each cell reports the KS statistic for the column country-row country pair. For two countries A and B, strong evidence that B FOSD A involves a large KS statistic in the cell of which B is the column country and A is the row country and a small KS statistic in the cell of which A is the column country and B is the row country. The "5%" and "1%" critical values are 1.3339 ($p = 2.82\%$) and 1.5765 ($p = 0.66\%$),⁷ and the KS statistics ≥ 1.3339 are in boldface. For example, $KS = 1.6977$ for the Canada-France pair and $KS = 0.1213$ for the France-Canada pair, suggesting that Canada FOSD France. In Table 3 there is not a single

⁷ In the brackets are the exact p-values of KS calculated based on Gnedenko and Koroyuk (1961).

case where both KS statistics for two countries are statistically significant (i.e. the case shown in Figure 4 is not observed in the data).

Guided by the boldfaced KS statistics, we see that the 10 countries fall into three groups: (1) the most capital-scarce country, U.K., is dominated by every other country except Denmark; (2) two of the three most capital-abundant countries, U.S. and Canada, dominate almost every other country; (3) the countries with medium capital abundance, such as Japan, Germany and France, dominate capital-scarce countries such as U.K., but are dominated by capital-abundant countries such as U.S. Australia is an exception. It is the second most capital-abundant country but behaves like a country with medium capital abundance, dominating only U.K. and being dominated by both U.S. and Canada. To summarize, the differences in empirical cdf's are non-uniform: they are more pronounced across different groups than within the same group and capital abundant countries are in capital abundant groups.

This pattern is consistent with the multiple-cone factor proportions theory with zero trade cost and by Proposition 1, each group can be interpreted as one cone. To be specific, the countries in the same cone do not first-order stochastically dominate one another, and every country in a relatively capital-abundant cone first-order stochastically dominates at least one country in every other cone with less capital abundance. Cone 1 is the most capital-scarce and has U.K. and Denmark. Cone 2 has medium capital abundance and includes the continental European countries of Italy, Germany, France and Netherlands plus Japan and Australia. Cone 3 is the most capital-abundant and has the North American countries of Canada and U.S. This designation of cones is listed in Table 4.

To visualize these three cones, Figure 7 plots their empirical cdf's side by side. Each

cone's empirical cdf is constructed from the pooled data of all the countries in this cone. From the left to the right are the empirical cdf's of cones 1, 2 and 3. To verify this pattern of first-order stochastic dominance I calculate the KS statistics for all the six pairs of cones. For the pairs cone 2-cone 1 and cone 1-cone 2, $KS = 2.4505$ and 0.1400 and the (asymptotic) 5% critical value is 1.2239 . For the pairs cone 3-cone 2 and cone 2-cone 3, $KS = 2.0304$ and 0 and the 5% critical value is also 1.2239 . For the pairs cone 3-cone 1 and cone 1-cone 3, $KS = 3.087$ and 0 and the "5%" critical value is 1.2862 ($p = 3.65\%$).

3.3 Issues

First, when two countries are in the same cone and there are more goods than factors, there might exist multiple equilibria, in some of which there is complete specialization at the industry level. For example, suppose there are at least 3 commonly produced (tradable) goods. Then it is possible that the less capital-abundant country produces the most capital-intensive good while the more capital-abundant country does not. This could be a concern if many industries have zero outputs and zero factor uses in the data. However, this is not the case. All the observations in my data have positive capital and labor uses and only 5 observations have zero gross output: industries 21 (motor vehicles) and 22 (aircraft) for Denmark and industry 35 (other producers) for Denmark, U.K. and France. Dropping these 5 observations produces almost identical results to those in Table 3. Furthermore, even if industry-level specialization were a concern, the countries in the same cone would still have the same distribution function for non-tradable goods since these goods are produced in every country. Thus cones could still be identified by comparing the distribution functions for the non-tradable goods. Doing so produces similar results (following DW 2001 I classify the service sectors and agriculture as non-tradable): there is strong evidence (5% level or better) that cone 3 FOSD both cone 1

and cone 2; the evidence that cone 2 FSOD cone 1 is significant at the 10% level.⁸

Second, the data satisfies the constraints that the factor uses across sectors add up to the factor endowments of each country, and so exhibits the Rybczynski feature: capital-abundant countries tend to have higher outputs of capital-intensive goods. As discussed in sections 2 and 3.1, the empirical cdf's assign equal frequencies to all industries and so are unaffected by industry outputs. Therefore the Rybczynski feature of the data does not bias the results in favor of multiple cones.

Third, I have used total factor requirements (TFRs) to measure capital and labor uses so far. The TFRs are the correct measure when the intermediate goods are non-tradable. When the intermediate goods are freely traded and available to all countries at the same prices, if factor price equalization (FPE) holds, the correct measure is the direct factor requirements (DFRs) instead (Staiger 1986). When I use the DFRs to calculate the capital-labor ratios, the results are similar to those reported in Table 3: Australia moves into the same cone as U.K. and Denmark while Netherlands moves into the same cone as U.S. and Canada. On the other hand, dealing with traded intermediate goods in the case of non-FPE is beyond the scope of this paper. This is because doing so requires data on how much each industry of every country uses the product of each industry of every other country (Davis and Weinstein 2003); such bilateral input-output-relation data is hard to come by (Trefler and Zhu 2005) and the literature has not reached a consensus on how to deal with traded intermediate goods in the non-FPE case using available data.⁹

⁸ This loss of statistical significance is not surprising since non-tradable goods account for about 30% of the observations.

⁹ See the discussions in Trefler and Zhu (2000, 2005), Davis and Weinstein (2003) and Feenstra (2004). Trefler and Zhu (2005) argue that most sources of input-output tables (e.g. the OECD) that separate imported and

Fourth, so far factor-augmenting productivity differences have not been accounted for and all the 10 OECD countries might be in the same cone when measured in efficiency equivalency units. To address this issue I use data on international test scores in math and science (Hanushek and Kimko 2000) to measure labor quality,¹⁰ and the last two columns of Table 2 report the scores. When I multiply the labor uses in each country by its test score to obtain the efficiency equivalency units of labor and then repeat the analysis in section 3.2 using the adjusted data, I obtain similar results. Due to its high test scores, Japan has large efficiency equivalency units and moves into the same cone as U.K. and Denmark. Italy shows the opposite pattern and moves into the same cone as U.S. and Canada due to its low test scores.

Finally, an advantage of the non-parametric estimation is that the KS test works even if the same industry codes represent different goods across countries. Is this the case in the data? If the same industry codes are the same goods across countries, the sectors' rankings in capital-labor ratios should be the same; i.e. their rank correlations should be 1. The data

domestically produced intermediate goods impute the data using the shares of imports divided by domestic absorption (the sum of gross outputs and net imports). Davis and Weinstein (2003) argue that this imputation might overstate the importance of traded intermediate goods.

¹⁰ Bernard and Schott (2002) and Bernard et al. (2003) address this issue by looking at the ratios of factor payments (e.g. rk/wl) in a regression framework. Factor-payment data is not available in the DW 2001 dataset. Furthermore, since w/r varies inversely with k/l across cones, the factor-payment ratios would take away useful variations in k/l across cones. This problem becomes more serious the closer the substitution elasticity between capital and labor approaches 1 from above. If this elasticity is less than 1, factor-payment ratios produce the reverse rankings across cones: the rk/wl of countries in capital-abundant cones are lower than in capital-scarce cones. The empirical estimates of this substitution elasticity are below 1 or close to 1 for many industries (e.g. Deardorff and Stern 1990).

suggests otherwise. “Ship building and repairing” is the 6th least capital-intensive industry in U.K. but the 16th in Japan; “Drugs and Medicines” is the 27th least capital-intensive industry in U.S. but the 13th in Canada. Among the rank correlations for the 45 country pairs, 80% are between 0.68 and 0.88. This suggests that the same industry codes might represent different goods across countries.¹¹ But how can this be? Perhaps countries in different cones produce different goods. If multiple cones are the only explanation, the rank correlations should be 1 for the countries within the same cone. Again, the data suggests otherwise. For example, the rank correlation is 0.7855 between U.K. and Denmark but 0.8304 between U.K. and Canada; it is 0.7616 between Germany and Japan but 0.886 between Germany and U.S. What’s more, the average rank correlation for all the within-cone country pairs, 0.8299, is not very different from that for all the cross-cone country pairs, 0.7698. Thus the comparability problems of cross-country data that Maskus (1991) documents are likely to be at work. This highlights the advantage of the non-parametric estimation: the same industry codes could be different goods across countries in the KS test.

Still, the rank correlations are all positive and pretty high, suggesting that the economic activities represented by a given industry code may be similar across countries. When the degree of similarity is high, the KS test may tend to be conservative and find too few cones (see section 3.1). To address this issue I use the regression analysis in section 4. On the other hand, the working paper version of this paper shows that the results in Table 3 are also robust to the issues of different production functions of non-tradable goods, multiple factors, heteroscedasticity and sector aggregation.

4 The Regression Analysis

When the same industry codes represent the same goods across countries, the following

¹¹ This finding is consistent with Schott (2003).

regression can be estimated:

$$\log(k_{jc}/l_{jc}) = \alpha_0 + \alpha_j + \beta_c + \varepsilon_{jc} \quad (3)$$

where j indexes sectors and c indexes countries. α_0 is the constant. α_j represents sector fixed effects and they capture the effects of production technique parameters that are identical across countries (θ_j in equation 2). β_c represents country fixed effects and they capture the effects of factor prices that are identical across sectors within a country (ω_c in equation 2). The specification that the country and sector fixed effects are additive is predicated on the assumption that the marginal effects of factor prices are the same across sectors within a country regardless of the production technique parameters of these sectors. This assumption holds, for example, when all industries have the same constant substitution elasticity between capital and labor.

In estimating regression (3) I set U.K.'s fixed effect to 0 by not including β_{UK} in the regression. Then β_c corresponds to the extent to which country c 's empirical cdf shifts to the right relative to the empirical cdf of U.K. Table 5 shows the β_c estimates of running (3) and Figure 8 plots them against the countries' capital abundance. In Table 5 the countries are arranged down the rows in ascending order of their capital abundance. For these results to be consistent with section 3, the country fixed effects β_c should show non-uniform differences: the countries found earlier to be in the same cone should have similar fixed effects while the countries in capital-abundant cones should have larger fixed effects than the countries in capital-scarce cones. It is reassuring that this pattern holds. U.K.'s fixed effect is constructed to be 0, and Denmark, the other country in cone 1, has the smallest fixed effect of 0.188 (not significantly different from 0). The countries in cone 2 --- continental European countries plus Japan and Australia --- have similar fixed effects ranging from 0.303 to 0.477, and the

countries in cone 3 --- U.S. and Canada --- have the largest fixed effects of 0.670 and 0.655.

We can also examine the relation between factor usage intensities and factor endowments by replacing the country fixed effects in regression (3) with the log of capital abundance:

$$\log(k_{jc}/l_{jc}) = \alpha_0 + \alpha_j + \beta_F \log(K_c/L_c) + \varepsilon_{jc} \quad (4)$$

If there is strong evidence for multiple cones among the countries in regression (4), $\beta_F > 0$ since variations in factor endowments change factor usage intensities across cones. This is what we see when (4) is run for all the 10 countries.¹² As shown in Table 5, the estimate of β_F is positive and significant. On the other hand, if the countries in (4) are in the same cone, variations in factor endowments have no effect on factor usage intensities and so $\beta_F = 0$. This is also what we see when (4) is run cone by cone. As shown in Table 5, the estimates of β_F are now insignificant in most cases. Therefore, consistent with section 3, the correlation between factor endowments and factor usage intensities is more pronounced for countries in different cones than for countries within the same cone.

So far we have maintained the zero-trade-cost assumption, which allows us to interpret the non-uniform differences in the empirical cdf's of factor usage intensities as evidence for multiple cones. What if trade costs are non-zero?

5 Non-zero Trade Costs

Non-zero trade costs pertain to the degree of factor market integration, just like multiple

¹² The estimation is OLS and I get very similar results when I introduce country random effects. Another concern is that there might be a built-in positive correlation between $\log(k_{jc}/l_{jc})$ and $\log(K_c/L_c)$ in the data since the data is constructed for the factor uses to add up to the factor endowments for each country. To address this issue, I use non-residential capital per capita and per capita GDP from the Penn World Tables to instrument for $\log(K_c/L_c)$ and get very similar results.

cones. To illustrate this point, simulate the following 2x2x2 monopolistic competition model. Countries A and B use capital (K) and labor (L) to produce two differentiated goods, 1 and 2. Country B is capital-abundant and good 1 is capital-intensive. Both goods are subject to the iceberg trade cost $\tau \geq 1$ (when $\tau = 1$ trade costs are zero). Figure 9 plots how the log of the wage-rental ratio, $\log(w/r)$, changes with the log of capital abundance, $\log(K/L)$, for country B.¹³

Figure 9 shows two benchmark cases and they clarify the meaning of factor market integration. The first benchmark is autarky and country B's path is OH. In this case $\log(w/r)$ changes fairly quickly with $\log(K/L)$ and there is no linkage between countries A and B's factor markets. The second benchmark is zero-trade-cost ($\tau = 1$) and country B's path is OF. In this case, countries A and B are in the same cone, $\log(w/r)$ does not change with $\log(K/L)$ due to factor price equalization¹⁴ and countries A and B's factor markets are fully integrated. Now consider the path FS. In this case countries A and B are in different cones with country B specializing in good 1 and country A specializing in good 2. Similar to the autarky path OH, $\log(w/r)$ changes fairly quickly with $\log(K/L)$. This suggests that multiple cones imply weak factor market integration.

Figure 9 also shows the paths of country B with τ values ranging from 1.1 to 2. In all these cases, $\log(w/r)$ changes less with $\log(K/L)$ as compared with the autarky path OH and changes more with $\log(K/L)$ as compared with the zero-trade-cost path OF. When τ is 1.1, country B's

¹³ The model is similar to Romalis (2004). In the simulation I set both goods' consumption expenditure shares to 0.5 and the consumption substitution elasticity to 4. I also set the cost share of capital in production to 2/3 for good 1 and 1/3 for good 2. See the working paper version of this paper for more details of the simulation.

¹⁴ In the simulations I change the factor endowments of both countries A and B so that the integrated world economy (e.g. Dixit and Norman 1980) remains unchanged.

path lies very close to OF, indicating strong integration between countries A and B's factor markets. When τ is 2, country B's path lies close to OH, indicating weak integration between countries A and B's factor markets. These results suggest that as trade costs get higher, the relation between factor prices and factor endowments resembles the zero-trade-cost benchmark less and so there is weaker factor market integration.

The simulation also shows that non-zero trade costs induce different factor prices across countries (see also Romalis 2004). As a result, trade costs are an alternative interpretation of the differences in the empirical cdf's of factor usage intensities. Figure 10 illustrates how trade costs affect the multiple-cone interpretation. G_A and G_B are countries A and B's empirical cdf's and G_B FOSD G_A . In the zero-trade-cost case the difference between G_A and G_B can be interpreted as evidence for multiple cones as in section 3. When trade costs are non-zero, suppose G_B is also country B's theoretical distribution function in a model with the trade cost $\tau > 1$ and this model predicts G_c as country A's theoretical distribution function. Then the distance between G_B and G_c represents the portion of the difference between G_A and G_B that trade costs account for. As trade costs get higher, they account for a larger portion of this difference. We can visualize this by imagining that G_c moves to the left towards G_A in Figure 10. When G_c is close enough to G_A so that G_c does not FOSD G_A , countries A and B can no longer be inferred to be in different cones.

How high does the trade cost τ need to be to invalidate the multiple-cone interpretation? The definitive answer must wait for future research because it requires the calibration of a rich general equilibrium model with technology differences and in which the factor uses of a given country are affected by its trade costs with all its trading partners. As a *suggestive* answer I simulate a two-country model for each cross-cone country pair in the working paper version

of this paper and obtain the range of $1.4 \sim 1.7$ for τ .¹⁵ This is comparable to the trade cost estimates in the literature. For example, the survey by Anderson and van Wincoop (2004) summarizes that the estimates of cross-border trade costs among industrialized countries roughly fall in the range of $1.4 \sim 1.8$.¹⁶

However, even though multiple cones and non-zero trade costs might be observationally equivalent in terms of their effects on the distributions of factor usage intensities, they both imply weak factor market integration. Thus although the multiple-cone interpretation might not survive non-zero trade costs, its key message of weak factor market integration does.

6 Conclusion and Discussion

Despite the theoretical importance of diversification cones, the empirical study of them can be uninteresting if either (1) there is only one cone so that all countries have the same factor usage intensities or (2) each country is in a separate cone and cone identity coincides with country identity so that factor usage intensities differ significantly for every country pair. This paper performs bilateral comparisons of the cumulative distribution functions (cdf's) of factor usage intensities for 10 rich OECD countries and shows that the data is consistent with neither (1) nor (2); rather, the cdf's differ in a non-uniform way. The 10 countries form 3 distinct groups (with, e.g. UK, France and US in different groups) such that the differences are more pronounced across groups than within the same group and capital abundant countries are in capital abundant groups. The use of factor usage intensities accommodates Hicks-neutral cross-industry productivity differences across countries and the cdf's can be generated

¹⁵In the simulation I assume identical trade costs across goods, no technology differences across countries and Cobb-Douglas production.

¹⁶For the literature that estimates trade costs see also Engle and Rogers (1996), Eaton and Kortum (2002), Evans (2003) and Anderson and van Wincoop (2003).

using readily available data. Since the cdf's are not affected by industry outputs or industry identities, the estimation works even if the same industry codes represent different goods across countries.

The finding of non-uniform differences in factor usage intensities is consistent with the multiple-cone factor proportions theory with zero trade costs, with each group being one cone. An alternative interpretation of the data is non-zero trade costs because they also induce different factor prices and so different factor usage intensities across countries. Both the multiple-cone and trade-cost interpretations share one implication: factor market linkages are weak between the countries in different groups.

Finally, this paper suggests that the empirical study of diversification cones could benefit from the incorporation of non-zero trade costs. How much does each of the two alternatives, multiple cones and trade costs, explain the non-uniform differences in factor usage intensities? How do trade costs affect the way countries respond to and absorb international shocks? These, and many other questions, await future research.

Reference:

- Abadie, A., 2002. Bootstrap tests for distributional treatment effects in instrumental variable models. *Journal of the American Statistical Association* 97, 284--292.
- Anderson, G., 1996. Nonparametric tests of stochastic dominance in income distributions. *Econometrica* 64, 1183--1193.
- Anderson, G., 2001. The power and size of nonparametric tests for common distributional characteristics. *Econometric Review* 20, 1--30.
- Anderson, J. E., van Wincoop, E., 2003. Gravity with gravitas: a solution to the border puzzle. *American Economic Review* 93, 170--192.
- Anderson, J. E., van Wincoop, E., 2004. Trade costs. *Journal of Economic Literature* 42, 691--751.
- Barrett, G., Stephen G. D., 2002. Consistent tests for stochastic dominance. *Econometrica*, forthcoming.
- Bernard, A., Redding, S., Schott, P., Simpson H., 2003. Factor price equalization in the UK? A general test and evidence. Mimeo, Yale School of Management.
- Bernard, A., Schott, P., 2002. Factor price equality and the economies of the United States. Mimeo, Yale School of Management.
- Berndt, E. R., 1990. *The Practice of Econometrics, Classics and Contemporary*. Addison-Wesley, New York.
- Collins, S., 1985. Technical progress in a three-country Ricardian model with a continuum of goods. *Journal of International Economics* 19, 171--179.
- Davidson, R., Duclos, J.V., 2000. Statistical inference for stochastic dominance and for the measurement of poverty and inequality. *Econometrica* 68, 1435--1464.
- Davidson, R., MacKinnon, J.G., 1993. *Estimation and Inference in Econometrics*. Oxford University Press, Oxford.
- Davis, D., Weinstein, D., 2001. An account of global factor trade. *American Economic Review* 91, 1423--1453.
- Davis, D., Weinstein, D., 2003. The factor content of trade, in: Choi, E.K., Harrigan, J. (Eds.), *Handbook of International Trade*, Vol. 1. Blackwell, Malden, pp. 119 -- 145.
- Deardorff, A.V., 1979. Weak links in the chain of comparative advantage. *Journal of International Economics* 9, 197--209.
- Deardorff, A.V., Stern, R.M., 1990. *Computational Analysis of Global Trading Arrangements*. The University of

- Michigan Press, Ann Arbor.
- Debaere, P., 2002. Reversing the perspective: expansion activities of multinational corporations from middle income countries. Mimeo, University of Texas at Austin.
- Debaere, P., Demiroglu, U., 2003. On the similarity of country endowments and factor price equalization. *Journal of International Economics* 59, 101--136 .
- Dixit, A.K., Norman, V., 1980. *Theory of International Trade*. Cambridge University Press, Cambridge.
- Eaton, J., Kortum, S., 2002. Technology, geography and trade. *Econometrica* 70, 1741--1779.
- Engel, C., Rogers, J.H., 1996. How wide is the border? *American Economic Review* 86, 1112--1125.
- Evans, C.L., 2003. The economic significance of national border effects. *American Economic Review* 93, 1291--1312.
- Feenstra, R.C., 2004. *Advanced International Trade: Theory and Evidence*. Princeton University Press, Princeton.
- Feenstra, R.C., Hanson, G.H., 1997. Foreign direct investment and wage inequality: evidence from Mexico's maquiladoras. *Journal of International Economics* 42, 371--393.
- Gnedenko, B., Koroyuk, V., 1961. On the maximum discrepancy between two empirical distributions. *Selected Translations in Mathematical Statistics and Probability* 1, 13--16.
- Hanson, G.H., Slaughter, M.J., 2002. Labor-market adjustment in open economies: evidence from US states. *Journal of International Economics* 57, 3--29.
- Harrigan, J., 1997. Technology, factor supplies, and international specialization: estimating the neoclassical model. *American Economic Review* 87, 475--494.
- Helpman, E., Krugman, P., 1986. *Market Structure and Foreign Trade*. MIT Press, Cambridge.
- Hollander, M., Wolfe, D.A., 1999. *Nonparametric Statistical Methods*. John Wiley & Sons, New York.
- Leamer, E. E., 1996. In search of stolper-samuelson effects on U.S. wages. NBER Working Paper 5427.
- Leamer, E. E., Levinsohn, J., 1997. International trade theory: the evidence, in: Grossman, G., Rogoff, K. (Eds.), *Handbook of International Economics*, Vol. 3. North-Holland, Amsterdam, pp. 1339--1394.
- Maskus, K., 1991. Comparing international trade data and product and national characteristics data for the analysis of trade models, in: *NBER Studies in Income and Wealth* 55. University of Chicago Press, Chicago, pp. 17-- 56.

- McFadden, D., 1989. Testing for stochastic dominance, in: Fomby T., Seo T.K. (Eds.), *Studies in the Economics of Uncertainty* (in honor of J. Hadar). Springer-Verlag. Berlin, pp. 113--134.
- Neave, H.R., Worthington, P.L., 1988. *Distribution-Free Tests*. Routledge, London.
- Repetto, A., Ventura, J., 1997. The Leontief-Trefler hypothesis and factor price insensitivity. MIT Economics Department Working paper 97-13.
- Romalis, J., 2004. Factor proportions and the structure of commodity trade. *American Economic Review* 94, 67-97.
- Schott, P.K., 2003. One size fits all? Heckscher-Ohlin specialization in global production. *American Economic Review* 93, 687--708.
- Staiger, R.W., 1986. Measurement of the factor content of foreign trade with traded intermediate goods. *Journal of International Economics* 21, 361 -- 368.
- Trefler, D., Zhu, S.C., 2000. Beyond the algebra of explanation: HOV for the technology age. *American Economic Review* 90, 145 --149.
- Trefler, D., Zhu, S.C., 2005. The structure of factor content predictions. NBER working paper 11221.
- Ventura, R., 1997. Growth and interdependence. *Quarterly Journal of Economics* 112, 57--84.

Table 1 Short Descriptions of the Sectors

Sector	Description
1	Agriculture, forestry and fishery
2	Mining and quarrying
3	Food, beverages and tobacco
4	Textiles, apparel and leather
5	Wood products and furniture
6	Paper, paper products and printing
7	Industrial chemicals
8	Drugs and medicines
9	Petroleum and coal products
10	Rubber and plastic products
11	Non-metallic mineral products
12	Iron and Steel
13	Non-ferrous metals
14	Metal products
15	Non-electrical machinery
16	Office and computing machinery
17	Electric apparatus, nec (not elsewhere classified)
18	Radio, TV and communication equipment
19	Shipbuilding and repairing
20	Other transport
21	Motor vehicles
22	Aircraft
23	Professional goods
24	Other manufacturing
25	Electricity, gas and water
26	Construction
27	Wholesale and retail trade
28	Restaurants and hotels
29/30	Transport and storage and communication
31	Finance and insurance
32	Real estate and business services
33	Community, social and personal services
34	Producers of government services
35	Other producers

Table 2 Capital Abundance and Test Scores by Country

country	K/L	QL1	QL2
(1)	(2)	(3)	(4)
U.K.	18679	53.98	62.52
DENMARK	24188	53.48	61.76
ITALY	27725	44.59	49.41
JAPAN	30494	60.65	65.50
GERMANY	32526	59.03	48.68
FRANCE	34173	54.15	56.00
NETHERLANDS	35628	56.84	54.52
CANADA	35800	47.57	54.58
AUSTRALIA	36930	48.13	59.04
U.S.	38810	43.43	46.77

Notes:

QL1 and QL2 are indices constructed from the scores of six individual tests. QL1 is based on the world average score of 50 for all the six tests and QL2 is benchmarked to the U.S. performance on NAEP (National Assessment of Education Progress). The data for Denmark is imputed. See Hanushek and Kimko (2000) for more details. The results reported in section 3.3 are based on QL1. The results based on QL2 are similar.

Table 3 The Kolmogorov-Smirnov (KS) Statistics

	U.K.	DENMARK	ITALY	JAPAN	GERMANY	FRANCE	NETHER.	CANADA	AUL.	U.S.
U.K.		0.9701	1.8190	1.5765	1.9403	2.0616	2.1828	2.4254	1.3339	2.4254
DENMARK	0.0000		1.4552	0.9701	1.2127	1.5765	1.8190	2.3041	0.8489	1.9403
ITALY	0.1213	0.3638		0.2425	0.3638	0.3638	0.8489	1.0914	0.3638	1.3339
JAPAN	0.1213	0.3638	0.7276		0.7276	0.7276	1.0914	1.5765	0.4851	1.4552
GERMANY	0.2425	0.3638	0.8489	0.4851		0.7276	1.0914	1.4552	0.6063	1.4552
FRANCE	0.1213	0.2425	0.8489	0.4851	0.2425		1.2127	1.6977	0.3638	1.5765
NETHER.	0.1213	0.1213	0.1213	0.1213	0.2425	0.1213		0.7276	0.2425	0.8489
CANADA	0.0000	0.0000	0.1213	0.0000	0.2425	0.1213	0.1213		0.0000	0.6063
AUL.	0.0000	0.2425	0.8489	0.4851	0.7276	0.9701	1.0914	1.5765		1.4552
U.S.	0.0000	0.0000	0.0000	0.1213	0.1213	0.0000	0.0000	0.3638	0.1213	

Notes:

In the table are one-sided Kolmogorov-Smirnov (KS) statistics for first-order stochastic dominance. The critical value is 1.3339 ($p = 2.82\%$) and the KS statistics ≥ 1.3339 are in boldface. "NETHER." stands for Netherlands and "AUL." stands for Australia.

Table 4 The Three Cones under the Zero-trade-cost Assumption

Cone 1	U.K., Denmark
Cone 2	Italy, Japan, Germany, France, Netherlands, Australia
Cone 3	Canada, U.S.

Table 5 The Regression Analysis

	Regression (3)	Regression (4)			
		All Countries	Cone 1	Cone 2	Cone 3
$\ln(K_c/L_c)$		0.749 (0.124)	0.724 (0.181)	0.164 (0.184)	0.177 (0.594)
Denmark	0.188 (0.119)				
Italy	0.377 (0.121)				
Japan	0.303 (0.116)				
Germany	0.303 (0.106)				
France	0.381 (0.115)				
Netherlands	0.477 (0.127)				
Canada	0.655 (0.114)				
Australia	0.335 (0.113)				
U.S.	0.670 (0.118)				
constant	10.985 (0.140)	3.614 (1.201)	4.299 (1.813)	9.425 (1.912)	10.017 (6.257)
No. of Obs.	340	340	68	204	68
R^2	0.828	0.799	0.951	0.816	0.949

Notes:

The dependent variable is $\log(k_{jc}/l_{jc})$, the capital-labor ratio, where j indexes industries and c indexes countries, and the White heteroscedasticity-consistent standard errors are in the parentheses. The countries are arranged in ascending order of their capital abundance and sector fixed effects are not reported to save space.

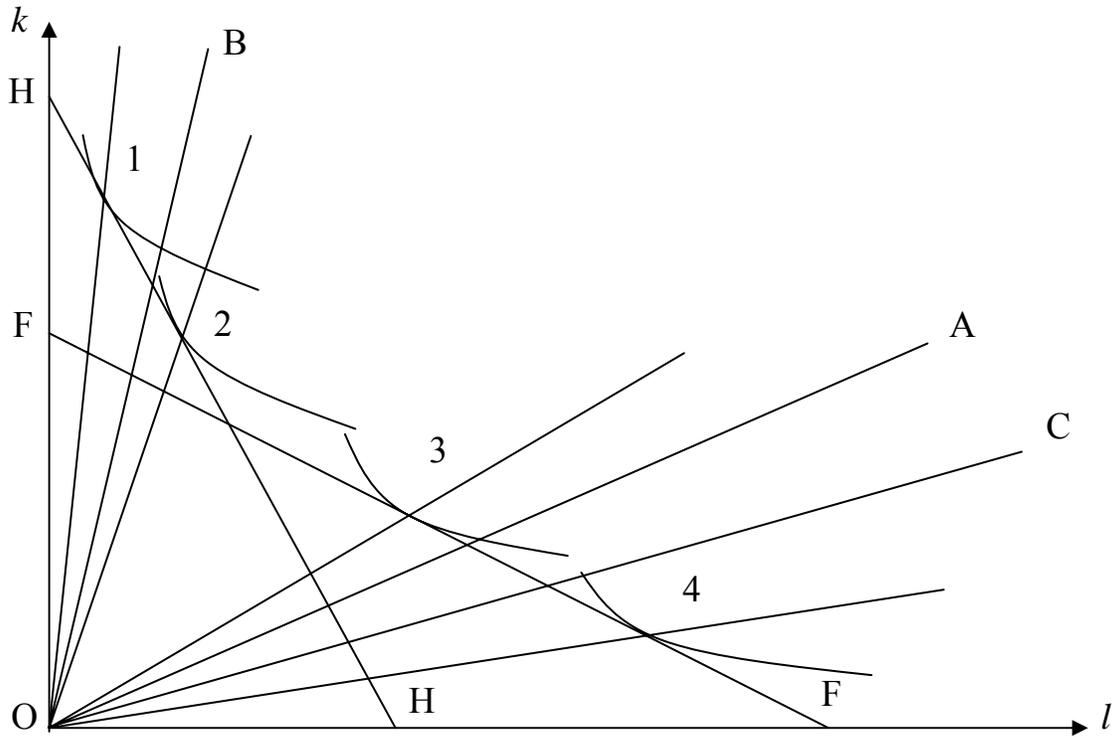


Figure 1 Cones with Zero Trade Cost and Homogeneous Goods

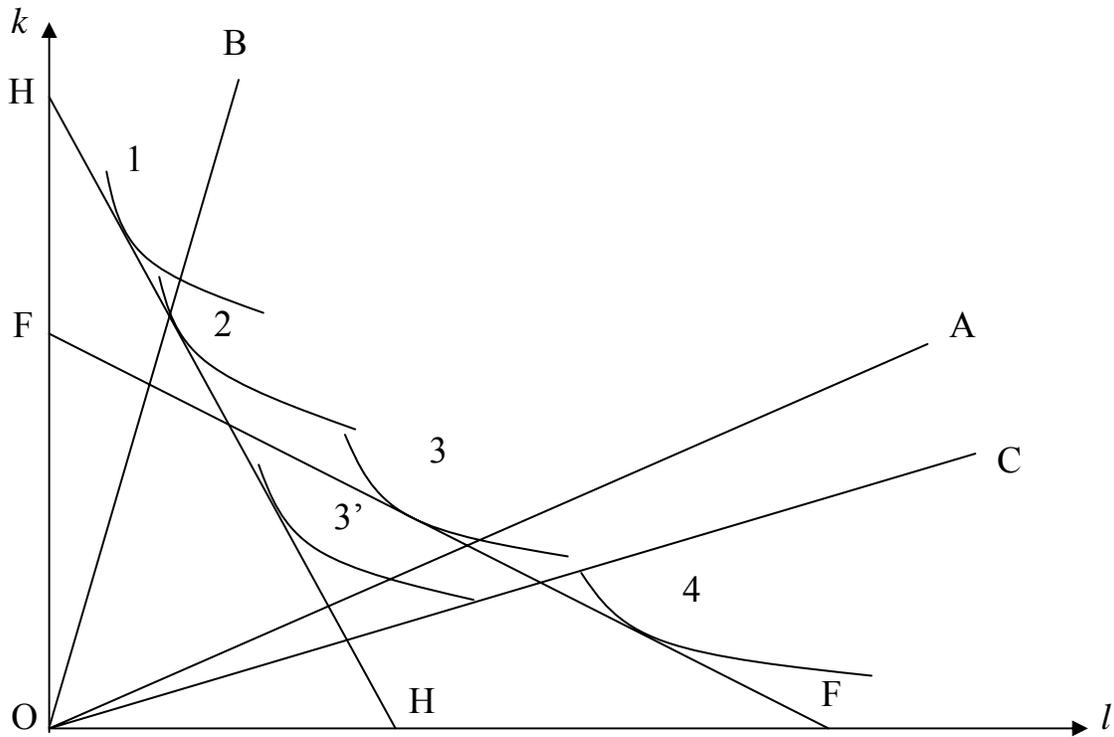


Figure 2 Cones with Zero Trade Cost and Differentiated Goods

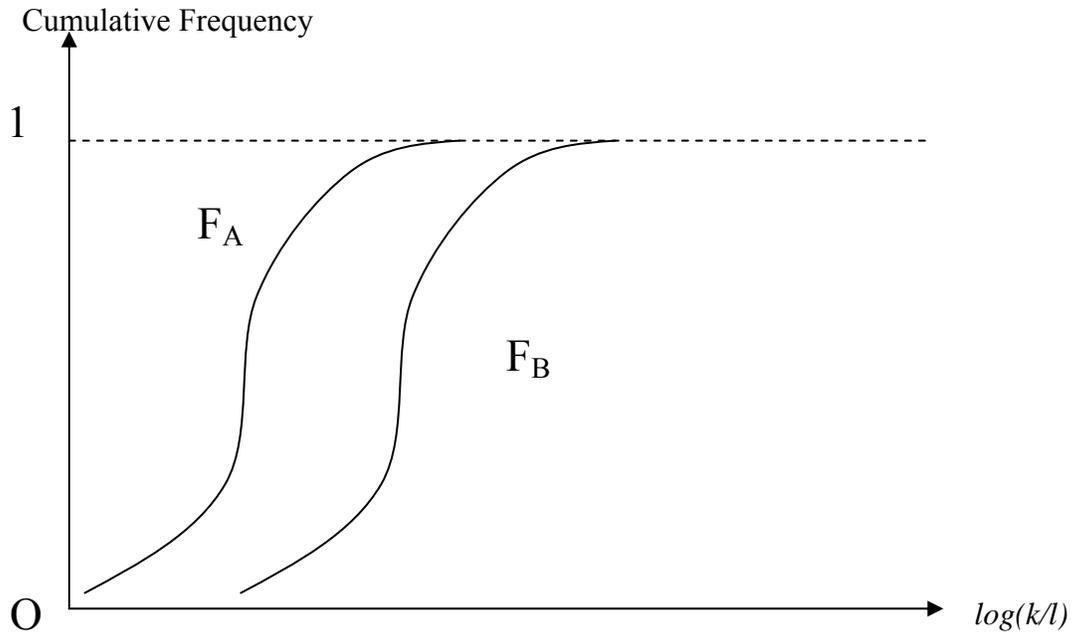


Figure 3 First Order Stochastic Dominance: I

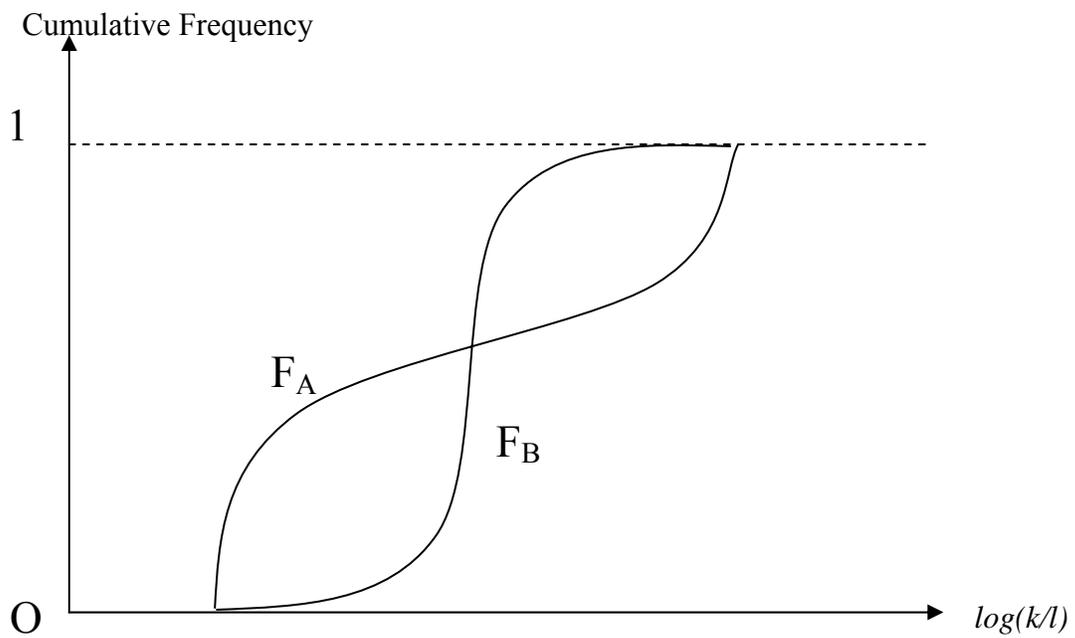


Figure 4 First Order Stochastic Dominance: II

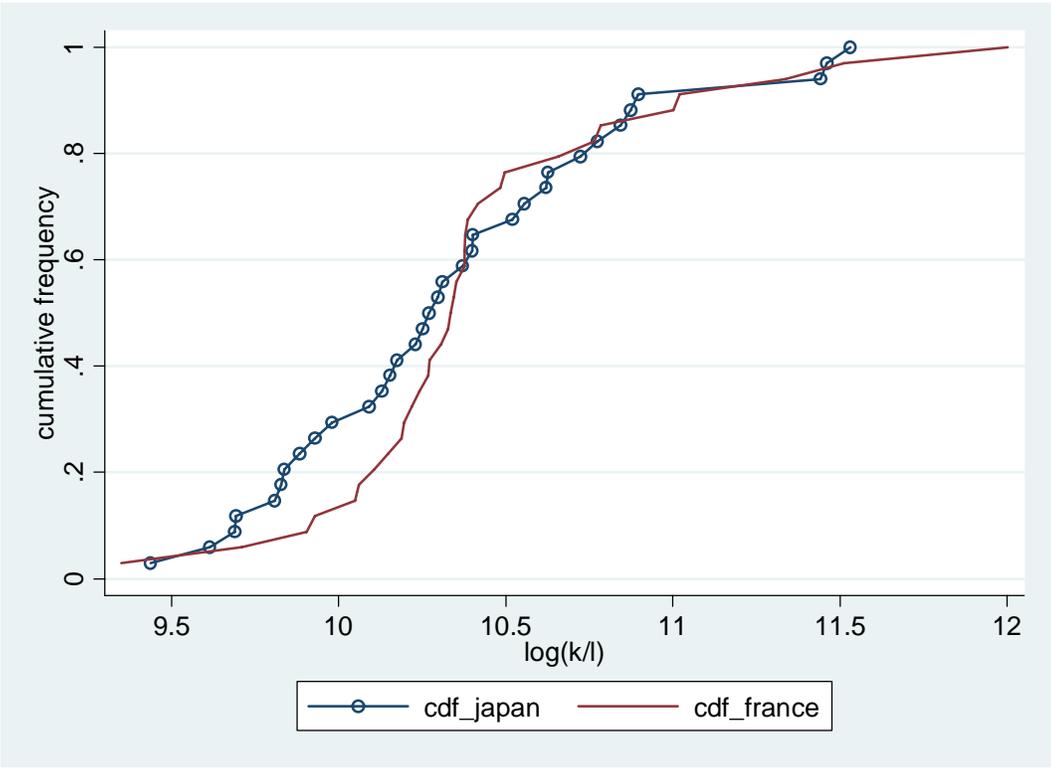


Figure 5 The Empirical cdf's of France and Japan

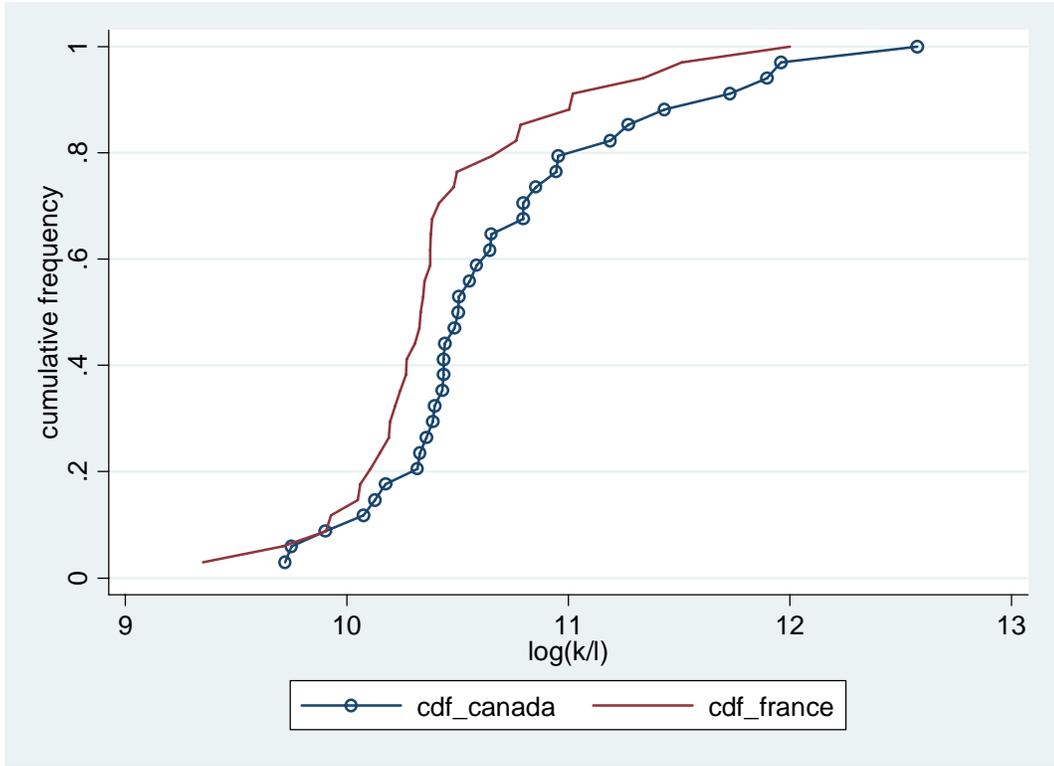


Figure 6 The Empirical cdf's of France and Canada

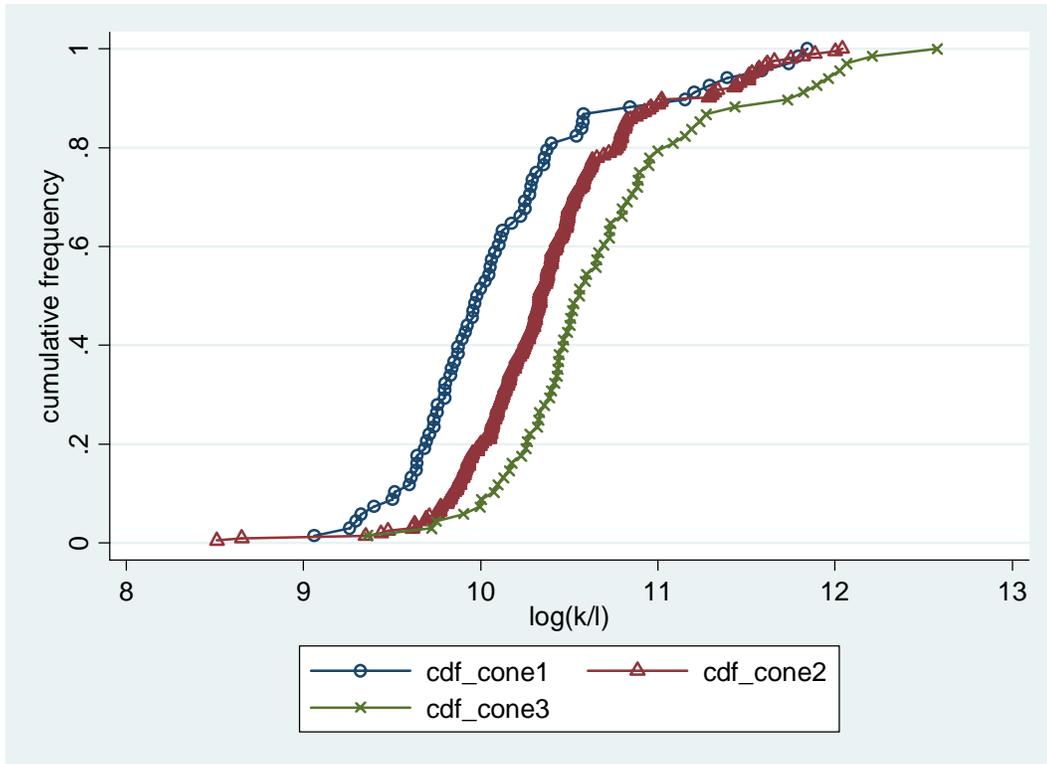


Figure 7 The Empirical cdf's of the Three Cones

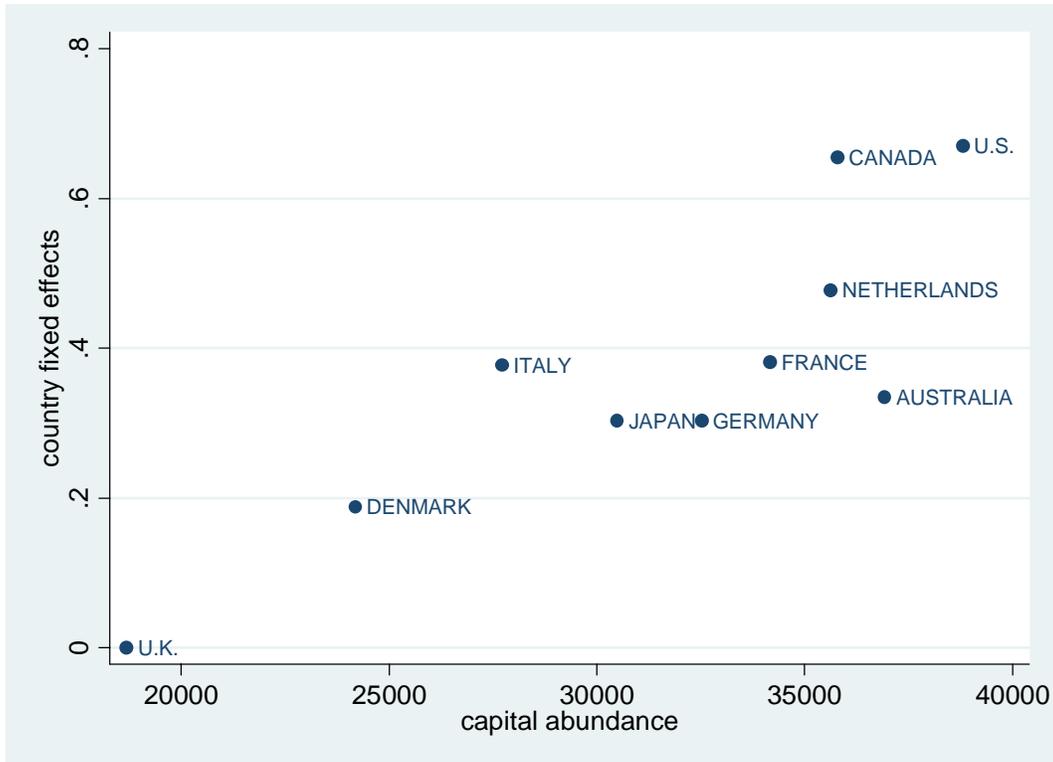


Figure 8 Country Fixed Effects and Capital Abundance

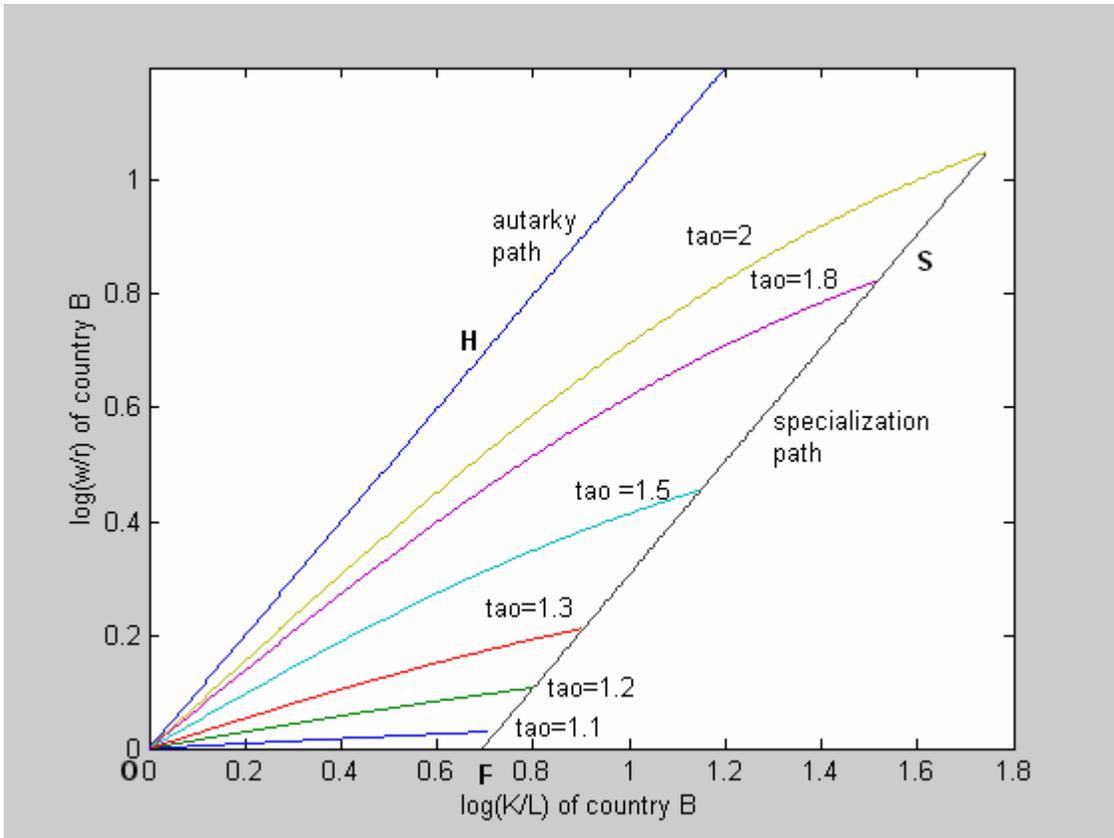


Figure 9 Cones, Trade Costs and Factor Market Integration

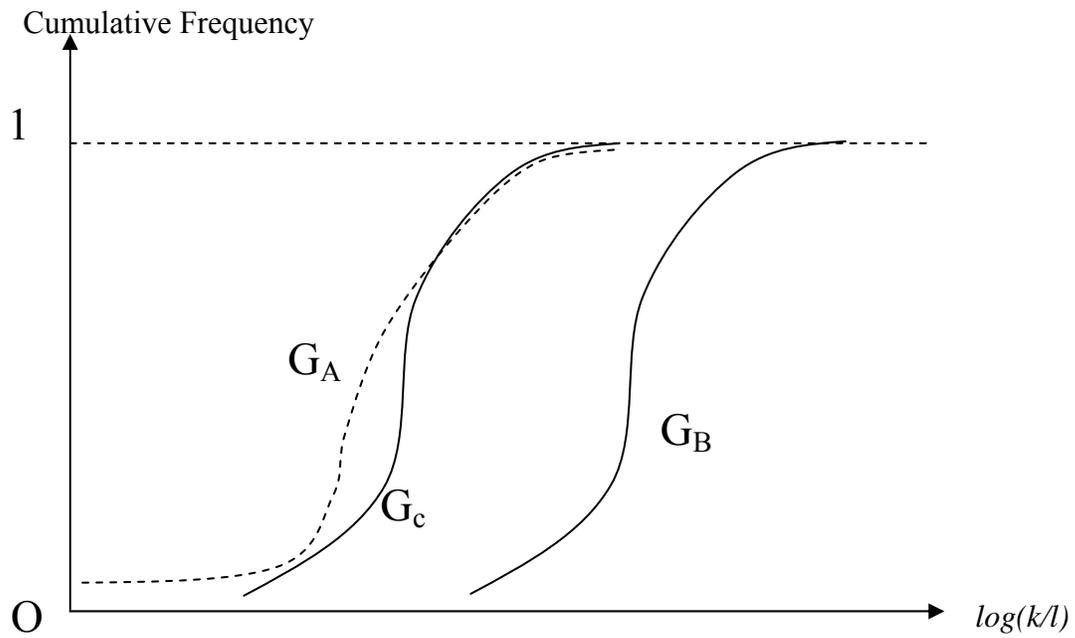


Figure 10 Trade Costs and the Multiple-cone Interpretation