DISCUSSION PAPER SERIES

No. 3368

THE GEOGRAPHY OF INTRA-INDUSTRY TRADE: EMPIRICS

Patricia Rice, Martin Stewart
and Anthony J Venables

INTERNATIONAL TRADE

Centre for Economic Policy Research

www.cepr.org

Available online at: www.cepr.org/pubs/dps/DP3368.asp

ISSN 0265-8003

THE GEOGRAPHY OF INTRA-INDUSTRY TRADE: EMPIRICS

Patricia Rice, Southampton University
Martin Stewart, Centre for Economic Performance (CEP), LSE
Anthony J Venables, London School of Economics (LSE) and CEPR

Discussion Paper No. 3368
May 2002

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 3368

May 2002

ABSTRACT

The Geography of Intra-Industry Trade: Empirics*

This Paper uses bilateral trade data for OECD countries at the 3-digit industry level to investigate the geography of intra-industry trade (IIT). IIT diminishes with distance and much of the existing empirical literature suggests that this is an inherent characteristic of such trade, arguing that trade in sectors intensive in IIT is choked off rapidly by distance. We show that the dependence of IIT on geography arises not because of any inherent feature of the effects of distance on such trade, but because of the spatial structure of countries' supply and demand characteristics; close countries do a lot of IIT because they have similar economic structures.

JEL Classification: F10, O10 and R10
Keywords: geography, intra-industry trade, trade structure and transport costs

Patricia Rice
Department of Economics
University of Southampton
Highfield
SOUTHAMPTON
SO17 1BJ
Tel: (44 23) 8059 3847
Fax: (44 23) 8059 3858
Email: econ@soton.ac.uk

Martin Stewart
Centre for Economic Performance
LSE
Houghton St
London
WC2A 2AE
Email: m.stewart@lse.ac.uk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=101743

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=157642

Anthony J Venables
Department of Economics
London School of Economics
Houghton Street
WC2A 2AE
LONDON
Tel: (44 20) 7955 7522
Fax: (44 20) 7831 1840
Email: a.j.venables@lse.ac.uk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=101393

* This Paper was produced as part of the Globalisation Programme of the UK Economic and Social Research Council funded Centre for Economic Performance, London School of Economics. This Paper is produced as part of a CEPR research network on ‘The Economic Geography of Europe: Measurement, Testing and Policy Simulations’, funded by the European Commission under the Research Training Network Programme (Contract No: HPRN-CT-2000-00069).

Submitted 25 April 2002

1. Introduction:

One of the most robust empirical finds in the literature on intra-industry trade (IIT) is that measures of intra-industry trade relative to inter-industry trade decline steeply with the distance between the trading partners.[1] There are two distinct reasons why this could occur. One is that there are differences in the nature of the trading process that make IIT attenuate more rapidly with distance than other forms of trade; for example, products in which IIT is important may also be products for which distance chokes trade off rapidly, so IIT measures fall with distance. The other is that industry is located such that neighbouring countries produce and demand a similar mix of products, in which case IIT will be a larger share of trade between these countries than of trade between more distant countries. Then, even if the trading process is the same for all products, the data may show a systematic relationship between IIT and distance.[2]

Most of the empirical literature presumes that the former interpretation is correct. For example, Balassa (1986a) recognises that all trade volumes diminish with distance, but argues that "there is need for information on the characteristics of differentiated products such as machinery, transport equipment, and consumer goods, which are subject to intra-industry trade.... it can be assumed that the availability of information decreases, and its costs increase, with distance. Correspondingly, it may be hypothesised that the extent of intra-industry trade between any two countries will be negatively correlated with the distance between them" (p111). While this argument seems plausible, it is easy to construct counter arguments; for example, differentiated products have (almost by definition) lower price elasticities of demand, so any distance cost will choke off trade *less* fast for these products, suggesting that intra-industry trade might be positively correlated with distance.

The second explanation, that IIT is determined by country characteristics (comparative advantage and demand differences), is suggested by simple theoretical models of trade under monopolistic competition, such as Helpman and Krugman (1985). However, empirical work has not been very successful in relating IIT to cross-country differences in endowments and other country characteristics. Hummels and Levinsohn (1995) are reluctantly driven to conclude that "instead of factor differences explaining the share of intra-industry trade, much intra-industry appears to be specific to country pairs.... distance is especially important to this relationship. ....these results leave us pessimistic. If much intra-industry trade is specific to country pairs, we can only be skeptical about the prospects for developing any general theory to explain it" (p828).

In this paper we revisit this issue to establish the extent to which the observed pattern of intra-industry trade between pairs of countries can be explained by the characteristics of the countries, as compared to pair specific (or "between-country") characteristics such as distance or sharing a border. We argue that the pessimism of Hummels and Levinsohn is misplaced and that the spatial structure of IIT can be largely accounted for by the spatial structure of country characteristics.

We analyse bilateral trade data for the major OECD countries, disaggregating to the level of 219 3-digit industries. The argument proceeds in three stages. First, in section 2, we show how aggregate (over industries) IIT measures depend on geography, declining strongly with distance as has been found elsewhere in the literature. Decomposing these aggregate measures into within-industry and between-industry (compositional) effects, we show that most of the variation is within-industry. Thus – at least at this level of aggregation – the dependence of IIT on distance is not primarily because trade in IIT intensive industries is rapidly choked off by distance.

We therefore turn, in section 3, to modelling trade flows at the industry level. These depend on exporter and importer country characteristics, and on between-country characteristics such as distance. We estimate gravity-type equations to identify these effects, using country dummy variables to capture each country's 'supply capacity' and 'market capacity' for each industry.[3] In most industries the industry level IIT measures decline strongly with distance if these country characteristics are not controlled for, while controlling for them makes the effect of distance insignificant in up to 96% of industries. Thus, the spatial pattern of IIT is merely reflecting the spatial distribution of country characteristics.

Country characteristics are captured through the dummy variables in the trade flow equations and these provide measures of a country's supply capacity and market capacity in each industry. In section 4, we look directly at the spatial dependence of these country characteristics. We show that countries that are geographically close have more similar supply and demand structures. It is this that generates the dependence of intra-industry trade on geography, not anything particular to specific country pairs, or anything inherent in this type of trade. Section 5 repeats part of the analysis on the sub-sample of European countries, and section 6 concludes.

2. Intra-industry trade and distance:

Let the exports of commodity *k* from country *i* to country *j* be denoted x_{ij}^{k}. Intra-industry trade in this commodity between countries *i* and *j* can be measured in different ways, the most commonly employed measure being the Grubel-Lloyd index (Grubel and Lloyd 1975) which takes the form

$$g_{ij}^{k} \equiv 1 - abs\,(x_{ij}^{k} - x_{ji}^{k}) / (x_{ij}^{k} + x_{ji}^{k}). \tag{1}$$

This has the advantage of being an index that lies between zero and unity and of being symmetric, $g_{ij}^{k} = g_{ji}^{k}$. However, g_{ij}^{k} is a non-linear function of the bilateral trade flows between a pair of countries, and hence the error structure in a model of x_{ij}^{k} does not translate into a natural error structure on g_{ij}^{k}. Additionally, the dependent variable in (1) is truncated between zero and one, a problem that is rectified in some of the literature by a transformation of the variable (Balassa 1986a, b, Balassa and Bauwens 1987, 1988). We shall therefore conduct our analysis using both the Grubel-Lloyd index and an alternative simple measure of IIT first used by Verdoorn (1960), the ratio between flows in the two directions,

$$r_{ij}^{k} \equiv x_{ij}^{k} / x_{ji}^{k}. \tag{2}$$

We discuss the relationship between these measures in an appendix. Here we simply note that while $g_{ij}^{k} = g_{ji}^{k}$ this symmetry does not hold for r_{ij}^{k} and r_{ji}^{k}, as one is the reciprocal of the other. For the remainder of this section we simply select from the pair the value that is less than unity. This approach raises some sample selection issues which are addressed in section 3 once we have specified a model of bilateral trade flows.

An aggregate measure across commodities typically weights by the share of the commodity in bilateral trade, $s_{ij}^{k} \equiv (x_{ij}^{k} + x_{ji}^{k}) / \sum_k (x_{ij}^{k} + x_{ji}^{k})$, so the aggregate versions of the Grubel-Lloyd and the ratio measures are,

$$G_{ij} = \sum_k s_{ij}^{k} g_{ij}^{k}, \qquad R_{ij} = \sum_k s_{ij}^{k} \min[r_{ij}^{k}, r_{ji}^{k}]. \tag{3}$$

Our data is bilateral trade flows for 22 OECD countries ($N = 22$) and 219 3-digit industries ($K = 219$) averaged over the period 1995-97, obtained from the World Trade Analyzer (Feenstra 2000). These data are discussed more fully in the appendix.

The point of departure is to describe the dependence of the aggregate IIT measures on geography. The first two columns of each of tables 1 and 2 report the results of regressing these measures (Grubel-Lloyd in table 1, ratio measure in table 2) on the log of distance between countries and on a variable that takes value unity if the countries share a common border. We see that around 40% the variation in the IIT measure across trading pairs is accounted for by these geography variables. The estimated effect of distance is large, with a doubling of distance implying a reduction in the measure of intra-industry trade of around one-quarter in each case.

Why should this effect be present? We first present a decomposition to distinguish within- and between-industry effects. Between-industry effects arise because the intensity of intra-industry trade varies across industries and the share of industries in trade changes with distance; thus, industries with high IIT may trade more intensively over short distances than long distances. Within-industry effects occur as each industry's IIT measure varies with distance.

The decomposition involves writing (for G_{ij}, and analogously for R_{ij})

$$G_{ij} = \sum_k s_{ij}^k g_{ij}^k = \sum_k \left[-s^k g^k + s_{ij}^k g^k + s^k g_{ij}^k + (s_{ij}^k - s^k)(g_{ij}^k - g^k) \right] \qquad (4)$$

where g^k is the trade weighted average of the bilateral indices for industry *k* and s^k is the share of industry *k* in total trade,

$$g^k \equiv \sum_i \sum_j s_{ij}^k g_{ij}^k$$
$$s^k \equiv \sum_i \sum_j (x_{ij}^k + x_{ji}^k) / \sum_i \sum_j \sum_k (x_{ij}^k + x_{ji}^k) \qquad (5)$$

The final three columns of tables 1 and 2 report the result of regressing each element of this decomposition on geography. Notice that, since this an exact decomposition, the coefficients on distance and border effects sum to the corresponding coefficient for the aggregate (unlogged) variable (coefficients in columns 3-5 sum to those in column 2).

Table 1: Grubel-Lloyd index and distance

	$ln(G_{ij})$	G_{ij}	$\sum_k g^k s_{ij}^k$ (between)	$\sum_k g_{ij}^k s^k$ (within)	$\sum_k cov(s_{ij}^k g_{ij}^k)$
Constant	1.674 (4.76)	0.933 (12.83)	0.412 (17.66)	0.908 (14.40)	-0.060 (-2.09)
ln distance	-0.397 (-9.28)	-0.081 (-9.16)	-0.014 (-4.98)	-0.072 (-9.33)	0.005 (1.36)
common border	0.233 (1.34)	0.144 (4.01)	0.001 (-0.12)	0.112 (3.59)	0.034 (2.39)
Adjusted R^2	0.35	0.42	0.11	0.41	0.02

No. obs: 231 = $N(N - 1)/2$
t-statistics in parentheses

Table 2: Ratio measure and distance

	$ln(R_{ij})$	R_{ij}	$\sum_k r^k s_{ij}^k$ (between)	$\sum_k r_{ij}^k s^k$ (within)	$\sum_k cov(s_{ij}^k r_{ij}^k)$
Constant	1.539 (4.11)	0.726 (12.14)	0.312 (17.01)	0.717 (13.69)	-0.054 (-2.23)
ln distance	-0.420 (-9.22)	-0.064 (-8.81)	-0.011 (-4.79)	-0.058 (-9.08)	0.004 (1.52)
common border	0.279 (1.51)	0.130 (4.38)	0.0001 (0.01)	0.102 (3.93)	0.028 (2.31)
Adjusted R^2	0.35	0.42	0.1	0.41	0.02

No. obs: 231

This decomposition shows that the spatial dependency of IIT measures arises overwhelmingly from the within-industry variation in IIT measures (g_{ij}^k, r_{ij}^k) across trading partners, i, j, rather than from between-industry effects due to the composition of trade (s_{ij}^k) varying with distance. For example, the term $\sum_k g^k s_{ij}^k$ captures changes in the composition of trade, as the weights, s_{ij}^k, applied to each industry's IIT measure, g^k, change with distance. This varies negatively and significantly with distance, but little of the variation of G_{ij} with respect to distance is accounted for by variation of this term (around 17%, 0.014/0.081). The terms $\sum_k g_{ij}^k s^k$ use fixed industry weights, s^k, and vary across country pairs in so far as within-industry IIT measures, g_{ij}^k, vary across country pairs. Spatial variation of these within- industry measures accounts for 89% (0.072/0.081) of the spatial variation of the aggregate measure. Thus, it is not the changing commodity composition of trade with distance (between-industry effects), but predominantly within-industry variation in g_{ij}^k that creates the dependence of aggregate IIT on distance. Similarly, looking at the ratio measures, 0.058/0.064 = 91% of the distance effect is accounted for by within-industry spatial variation of r_{ij}^k. Furthermore, 78% of the common border effects are also accounted for by within-industry spatial variation.

This decomposition demonstrates that the negative relationship between the aggregate IIT measures and distance is not, primarily, due to IIT intensive sectors' trade being choked off by distance. However, it does not inform us as to why within-industry IIT should decline with distance. To answer this we have to look within each industry.

3. Industry analysis:

A general model of industry level trade expresses trade in good *k* from country *i* to country *j*, x_{ij}^k, as a function of supplier country characteristics, s_i, importer country characteristics, m_j, and between-country factors such as distance, t_{ij} (= t_{ji}),

$$x_{ij}^k = F^k(s_i, m_j, t_{ij}). \tag{6}$$

Using this in the definition of an industry level IIT measure (either Grubel-Lloyd or ratio) produces a relationship that, in general, depends both on country characteristics and between-country characteristics. We express this as,

$$IIT_{ij}^k = I(x_{ij}^k, x_{ji}^k) = I^k(s_i, s_j, m_i, m_j, t_{ij}). \tag{7}$$

The null hypothesis is that country characteristics determine IIT, and that there is no spatial dependency of IIT beyond that incorporated in country characteristics, i.e. $\partial I^k/\partial t_{ij} = 0$. The objective of the empirical analysis is to test this hypothesis.

The direct approach to doing this is to specify a statistical version of (7) and to test for the significance of the between-country terms, t_{ij}. The difficulty with this is controlling adequately for the supplier and importer characteristics of each country in the pair, as is demonstrated in the work of Hummels and Levinsohn. An alternative is to use country dummies. However, it is not possible to identify supplier and importer effects separately for each country (i.e. both s_i and m_i) using dummy variables in a statistical version of (7).

The approach adopted in this paper is to exploit the additional information contained in the individual trade flow equations. Under the null hypothesis that $\partial I^k / \partial t_{ij} = 0$ it must be the case that the trade equations are separable in t_{ij}, as follows[4]

$$x_{ij}^k = F^k(s_i, m_j, t_{ij}) = f^k(s_i, m_j)\, h^k(t_{ij}) \tag{8}$$

This can be seen by using (8) in the definitions of the IIT measures, equations (1) and (2). Terms in t_{ij} cancel out for all values of s_i and m_i if, and only if, the separability holds. The trade flow equations (8) are estimated on all $N(N\text{-}1)$ trade flows allowing the supplier and importer effects to be estimated separately for each country using country dummies. This is in contrast to the relationship for IIT (equation (7)) which makes use of only $N(N\text{-}1)/2$ observations of IIT, in effect throwing away the information on the direction of trade.[5]

In what follows we estimate the trade flow equations with separability imposed, (8), and use the results to compute predicted IIT measures. We then examine the behaviour of the difference between the actual and predicted IIT measures. If our null hypothesis – implying separability – holds then all relevant information is contained in the predicted values, and the differences between the actual and predicted values do not depend on between-country characteristics.

The trade equations:

We estimate for each industry k, a trade equation of the form

$$\ln(x_{ij}^k) = \alpha^k + \beta_i^k s_i + \gamma_j^k m_j + \theta^k \ln(d_{ij}) + \eta^k b_{ij} + u_{ij} \tag{9}$$

where s_i and m_j are exporter and partner country dummy variables, d_{ij} is distance and b_{ij} is a dummy variable which is unity if countries share a common border. Equation (9) is a gravity equation with all the factors that affect a country's exports to all destinations, or its imports from all sources, summarised by the exporter and partner country dummy variables respectively. Thus, the coefficients on the exporter dummies, β_i^k, summarise (for the sector under study) everything that affects country i's 'supply capacity' in its exports to *all* destinations. Similarly, on the demand side, the vector of coefficients, γ_j^k, summarises everything that effects country j's 'market capacity' in creating import demand for the sector.

This (like standard gravity equations) has the separability structure of (8). The geographical destination of exports and sourcing of imports depends on the between-country variables, distance and a common border, and the separability assumption says that the elasticity of a trade flow with respect to a change in a country characteristic is the same at all distances. It is important to note that if equation (9) is misspecified, and in particular if the assumption of separability is not valid, then the interactions between country characteristics and distance will be captured in the error term u_{ij}^{k}.

Equation (9) is estimated on the bilateral trade data. As one would expect at this level of disaggregation, the outcome $x_{ij}^{k} = 0$ is frequently observed, and we use Tobit methods to allow for this. Given the log specification in (9), this involves adding a small amount of trade ($1000) to all flows and setting the truncation value at the log of this value.[6] As usual, the estimated gravity models fit the data with high explanatory power. Estimates of the parameters of (9) generate predicted values of the log of the trade flows, $\widehat{lnx}_{ij}^{k}$. The predicted value of the log of the ratio IIT measure is computed directly from these as $\widehat{lnr}_{ij}^{k} = \widehat{lnx}_{ij}^{k} - \widehat{lnx}_{ji}^{k}$. Since the Grubel-Lloyd measure is not linear in $\widehat{lnx}_{ij}^{k}$, things are less straightforward; we define $\hat{g}_{ij}^{k}$ as the value of g_{ij}^{k} evaluated at the predicted values of x_{ij}^{k}.

The IIT equation:

The second stage involves an analysis of the relationship between the actual IIT measures and those predicted under the null hypothesis. The objective is to determine whether the difference between these measures is systematically related to between-country characteristics, in particular distance.

Looking first at the ratio measures, the analysis is based on the following relationship,

$$\ln(r_{ij}^{k}) - \widehat{lnr}_{ij}^{k} = a_0^{k} + a_2^{k}\ln(d_{ij}) + a_3^{k}b_{ij} + \nu_{ij}^{k}. \qquad (10)$$

Under the null hypothesis

$$E(\ln(r_{ij}^{k})) = E(\widehat{lnr}_{ij}^{k}) = \beta_i^{k}s_i + \gamma_j^{k}m_j - \beta_j^{k}s_j + \gamma_i^{k}m_i \qquad (11)$$

and hence the parameters a_0^{k}, a_2^{k}, a_3^{k} in (10) are equal to zero. We also consider a more general specification,

$$\ln(r_{ij}^{k}) = b_0^{k} + b_1^{k}\widehat{lnr}_{ij}^{k} + b_2^{k}\ln(d_{ij}) + b_3^{k}b_{ij} + u_{ij}^{k}, \qquad (12)$$

in which, under the null, $b_1^k = 1$ and $b_0^k = b_2^k = b_3^k = 0$.

One further point needs to be made about estimation of these equations. As we have already noted, the measures r_{ij}^k, r_{ji}^k (both actual and predicted) are not symmetric; one is the reciprocal of the other. Selecting observations on the basis of the actual values implies censoring the distribution of the error ν_{ij}^k in (10) and hence would introduce sample selection bias. To avoid such problems we select on the basis of the lesser of the predicted values, $ln\hat{r}_{ij}^k$, rather than the actual values.

We also study the analogous equations for the Grubel-Lloyd index,

$$\ln(g_{ij}^k) - \ln(\hat{g}_{ij}^k) = a_0^k + a_2^k \ln(d_{ij}) + a_3^k b_{ij} + \nu_{ij}^k, \tag{13}$$

and

$$\ln(g_{ij}^k) = b_0^k + b_1^k \ln(\hat{g}_{ij}^k) + b_2^k \ln(d_{ij}) + b_3^k b_{ij} + u_{ij}^k. \tag{14}$$

Unlike the ratio measure, the Grubel-Lloyd index is not a linear function of the (log) of the trade flows and hence $E(\ln(g_{ij}^k)) \neq E(\ln(\hat{g}_{ij}^k))$. Nevertheless, under the null $b_2^k = b_3^k = 0$, although b_1^k is not necessarily equal to unity. Sample selection is not an issue for this equation, since $g_{ij}^k = g_{ji}^k$ and either can be used.

Results for the passenger motor car industry[7] are given in table 3, and summarised for all industries in figures 1 and 2. The first three columns of table 3 use the ratio measure. In column 1, the predicted value, $ln\hat{r}_{ij}^k$, is excluded and the estimated coefficient on the distance variable is negative, statistically significant and quantitatively important. The second column reports results including the predicted value with its coefficient constrained to be unity, equation (10); this improves the fit of the equation and reduces the size and the significance level of distance as a determinant of r_{ij}^k. The third column undertakes the same regression with the coefficient on $ln\hat{r}_{ij}^k$ unconstrained, equation (11). The estimated coefficient turns out to be close to unity and, unsurprisingly, other effects are similar.

Columns 4-6 present the analogous results for the Grubel-Lloyd measure. As with the ratio measure, distance is found to have a significantly negative effect on intra-industry trade when the predicted value of the measure is excluded. Including the predicted value, $\hat{g}_{ij}^k$, both with and without the parameter restriction, reduces both the magnitude and the significance of the coefficient on distance. Notice, however, that the coefficient on $\hat{g}_{ij}^k$ in column 6 is significantly less than unity, as expected because of the non-linearity of the Grubel-Lloyd measure.

Table 3: Representative industry (SITC 781 Passenger motor cars)

	$\ln(r_{ij}^k)$	$\ln(r_{ij}^k) - ln\hat{r}_{ij}^k$	$\ln(r_{ij}^k)$	$\ln(g_{ij}^k)$	$\ln(g_{ij}^k) - \ln(\hat{g}_{ij}^k)$	$\ln(g_{ij}^k)$
Constant	0.054 (0.03)	0.159 (0.15)	0.160 (0.15)	0.391 (0.29)	0.121 (0.13)	0.171 (0.19)
$ln\hat{r}_{ij}^k$			1.014 (18.95)			
$\ln(\hat{g}_{ij}^k)$						0.816 (17.17)
ln distance	-0.417 (-1.95)	-0.004 (-0.03)	0.001 (0.01)	-0.453 (-2.73)	-0.057 (-0.51)	-0.130 (-1.18)
common border	1.121 (1.33)	0.717 (1.39)	0.711 (1.37)	1.019 (1.56)	0.663 (1.51)	0.729 (1.71)
Adjusted R^2	0.04	0.002	0.64	0.07	0.01	0.6

No. obs: 221. Observations where trade in both directions are zero have been dropped.

Figure 1 summarises results for all industries for the ratio measure, and contains four panels. The first two (1a and 1b) give the distributions of parameter estimates and *t*-statistics for the distance variable in the regression with $ln\hat{r}_{ij}^k$ excluded, as in the first column of table 3. We see that the estimated distance coefficients are negative for 89% of industries, and significantly negative for 60% [8]. The second two panels (1c and 1d) give these distributions when $ln\hat{r}_{ij}^k$ is included, as in the third column of table 3. The distributions are shifted to the right and centred around zero. 43% of estimated distance coefficients are negative, and just 4% are significantly so. Effects on border coefficients are similar although less pronounced: with $ln\hat{r}_{ij}^k$ excluded 83% are positive, and 16% significantly so, these proportions dropping to 77% and 4% respectively when $ln\hat{r}_{ij}^k$ is included.

Figures 2a-2d give the same summary information using the Grubel-Lloyd measures, and the same conclusion emerges. When the predicted values are excluded the estimated distance coefficients are negative for 92% of industries, and significantly negative for 68%. With predicted values included exactly 50% of estimated coefficients are negative, and just 11% are significantly so.

Thus we conclude that for the vast majority of industries (96% using the ratio measure) it is not possible to reject the null hypothesis that the spatial variables have no effect on IIT once country characteristics are properly controlled for. This suggests that it is not distance *per se* causing the negative correlation between industries' IIT measures and distance, but rather the spatial structure of country characteristics.

Aggregate measures:

We have shown that for the vast majority of 3-digit industries, separability of the trade flow equation as in (8) is not rejected by the data. What does this imply for the aggregate measures of intra-industry trade that we discussed in section 2? The estimates of the trade equations can be used to compute predicted values for the aggregate measures,

$$\hat{G}_{ij} = \sum_k \hat{s}_{ij}^k \hat{g}_{ij}^k, \qquad \hat{R}_{ij} = \sum_k \hat{s}_{ij}^k \min[\hat{r}_{ij}^k, \hat{r}_{ji}^k]. \tag{15}$$

These provide the basis for estimation of the aggregate analogues of equations (12) and (14).

The results of this exercise are reported in table 4. The first and fourth columns are drawn from tables 1 and 2. Looking at column 2 we see that the distance and border variables are statistically significant even after inclusion of the predicted value, $\ln\hat{R}_{ij}$. However, they add little explanatory power over and above the predicted values (compare columns 2 and 3). The continuing significance of these variables is in contrast to the industry level results, but not inconsistent with them. While separability of the underlying trade flow equation implies that the industry level IIT^k indices, g_{ij}^k and r_{ij}^k, are independent of distance, this does not necessarily follow for the industry aggregate IIT indices, G_{ij} and R_{ij}. These aggregate measures are functions of the trade shares, $s_{ij}^k \equiv (x_{ij}^k + x_{ji}^k) / \sum_k (x_{ij}^k + x_{ji}^k)$, and the shares depend on the functions $h^k(t_{ij})$, unless it happens to be the case that these functions are the same for all industries and so can be cancelled out of top and bottom of this expression. Thus, just as we saw in section 2, in general there are some 'between-industry' effects present in the aggregate measures, although they are quantitatively dominated by the 'within-industry' effects analysed in section 3. Remaining columns of table 4 repeat this exercise with the Grubel-Lloyd index and demonstrate similar findings.

This exercise sheds light also on the use of more disaggregate data, at a level finer than the 3-digit classification. The fact that separability is not rejected at the 3-digit level suggests that it holds at any finer level *and further* that for all sub-industries of a given 3-digit industry the functions $h^{\ell}(t_{ij})$ are similar.

Table 4: Aggregate merchandise trade

	$\ln(R_{ij})$	$\ln(R_{ij})$	$\ln(R_{ij})$	$\ln(G_{ij})$	$\ln(G_{ij})$	$\ln(G_{ij})$
Constant	1.539 (4.11)	0.492 (2.03)	-0.234 (-3.21)	1.674 (4.76)	0.575 (2.57)	-0.176 (-2.89)
$\ln\hat{R}_{ij}$		0.829 (18.68)	0.947 (23.73)			
$\ln\hat{G}_{ij}$					0.863 (19.41)	0.984 (24.34)
ln distance	-0.420 (-9.22)	-0.118 (-3.60)		-0.397 (-9.28)	-0.117 (-3.90)	
common border	0.279 (1.51)	0.291 (2.50)		0.233 (1.34)	0.251 (2.35)	
Adjusted R^2	0.35	0.74	0.71	0.35	0.75	0.72

No. obs: 231

4. Spatial correlations of country characteristics:

In section 3 we demonstrated that once country characteristics are properly controlled for, the hypothesis that between-country variables such as distance have no significant effect on intra-industry trade is not rejected by the data at the 3-digit industry level. The apparent strong negative correlation between IIT measures and distance found in simple regression models is a reflection of the particular spatial structure of country characteristics, suggesting that countries that are located relatively close to each other have relatively similar supply and market capacities. We now look at this directly.

Estimation of the trade flow equations (9) for each industry – at the 3-digit level – yields a KxN matrix of supply capacities, $B = [\hat{\beta}_i^k]$, and a similar matrix of market capacities, $\Gamma = [\hat{\gamma}_j^k]$. Thus, the ith column of the matrix B gives the supply capacity of country i in each industry. How similar are these columns, and how does the degree of similarity vary with distance? Recognising that countries and industries differ in size, we normalise elements of these matrices by expressing each relative to the sum of the elements for the country and relative to the industry share of the total (see appendix). The similarity of the supply capacities of a pair of countries is measured by the correlation coefficient between these normalized column vectors, and similarly for the market capacities. The correlation matrices are given in the appendix.

Table 5 reports the results of regressing the correlation coefficients on distance and on the presence of a common border. Thus, the results in the first column come from regressing the $N(N-1)/2$ correlation coefficients between countries' vectors of supply capacities on distance and border variables. These two geographical variables account for 52% of the variation in this measure of similarity, and the effect of distance is negative and highly significant, suggesting that more distant countries have more dissimilar production structures. The relationship is somewhat weaker for the correlations of market capacity, but the coefficient on distance is still negative and statistically significant; this does not necessarily imply different structures of consumer demand, as intermediate as well as final demands influence market capacities.

Table 5: Spatial correlations of supply and market capacities and geography

	Supply capacity correlations	Market capacity correlations
Constant	1.666 (17.41)	1.408 (13.77)
ln distance	-0.165 (-14.18)	-0.137 (-10.97)
common border	-0.011 (-0.23)	-0.003 (-0.06)
Adjusted R^2	0.52	0.39

No. obs: 231 = $N(N-1)/2$

5. The European subset:

Our base data set contains 16 European countries plus Australia, Canada, Japan, New Zealand, Turkey and the United States. This geography – a large number of relatively close countries and a few spatial outliers – might be influential in driving our results. For this reason, we re-estimated the model focusing on the 16 European countries alone.

Estimation of the trade equations, (9), was unchanged: supply and market capacities should reflect trade as a whole, not that with a subset of countries. However, the subsequent analysis was based on just the 120 (16x15/2) observations arising from intra-European trade. The histograms, reported in Figure 3a-d (ratio measures only), indicate that the results are essentially unchanged. When regressing the ratio measure on the geography variables only, the estimated distance coefficients are significantly negative in 57% of the industries. However, once the predicted ratio is included as an explanatory variable, the proportion of significantly negative distance coefficients falls to 5% .

The spatial correlations of supply and market capacities are given in table 6. We see that the correlations of supply and market capacities with distance are much weaker than on the full OECD data set. However, it is interesting that common border effects become significant in the supply capacity correlations, indicating similarity of supply characteristics across contiguous countries in Europe.

Table 6: Spatial correlations of supply and market capacities: Europe

	Supply capacity correlations	Market capacity correlations
Constant	0.481 (2.23)	0.083 (0.25)
ln distance	-0.041 (-1.38)	0.019 (0.41)
common border	0.104 (2.15)	0.046 (0.62)
Adjusted R^2	0.09	-0.01

No. obs: 120

6. Conclusions:

Measures of IIT between pairs of countries decline with the distance between the countries in the pair. Most of the existing literature suggests that this is because such trade is rapidly choked off by distance due, for example, to high information costs; thus, the composition of a country’s exports changes according to the distance they travel. This paper finds that this effect is present, but is quantitatively dominated by spatial variation of within-industry IIT. Around 60% of 3-digit industries have IIT measures declining significantly with distance. However, this within-industry variation can be almost entirely accounted for by country characteristics; controlling for these, only 4% of industries have IIT falling significantly with distance. These country characteristics are, in turn, spatially correlated. Closer countries tend to have more similar structures of underlying export supply and import demand and, as a consequence, IIT tends to relatively high between close (and hence similar) countries.

In this paper we remain agnostic about the reasons for the spatial structure of countries' supply and demand characteristics. It will be the subject of future research.

Appendix:

Data:

We use average trade for the period 1995-97 for 219 3-digit industries among 22 OECD countries (Australia, Austria, Belgium-Luxembourg, Canada, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, Turkey, United Kingdom and the United States; we exclude Iceland because of its small size). The data is in thousands of constant US dollars, deflated using the US GDP deflator from the World Bank's World Development Indicators CD-Rom, 2001.

Trade flows are from the World Trade Analyzer, a consistently recompiled version of the bilateral trade data published by the United Nations (Feenstra 2000). The database is made available as a single array of positive trade flows and, due to the preparatory work made by its compilers, we can be reasonably confident that non-reported trades are actually zeroes or negligible amounts. Our three-year average further ensures that any problems due to unreported or misallocated trades have been smoothed out.

The World Trade Analyzer reports for all 4-digit SITC (Revision 2) commodities, but a number of the trades are actually reported at the 3- and 2-digit levels. We have aggregated all trades to the 3-digit level and apportioned any 2-digit trades to their constituent 3-digit commodities. Furthermore, we aggregated the 5 commodities in which almost all trades were zeroes (i.e. over 75 percent zeroes) into other commodity classifications. Thus, for our analysis we have 219 3-digit industries.

The distance data is the great circle distance in kilometres between the economic centres of gravity of each country, i.e. the city with the largest population (as per the UN Demographic Yearbook 1995). The common border data is a dummy equal to one if the country-pair share a common land border, and zero otherwise.

The relationship between Grubel-Lloyd and ratio measures:

We work with both g_{ij}^k and r_{ij}^k. The relationship between these variables is illustrated in figure 4, and takes the form:

$$if \quad r_{ij}^k < 1, \quad g_{ij}^k = 2r_{ij}^k/(r_{ij}^k + 1), \qquad r_{ij}^k = g_{ij}^k/(2 - g_{ij}^k),$$

$$if \quad r_{ij}^k > 1, \quad g_{ij}^k = 2/(r_{ij}^k + 1), \qquad r_{ij}^k = (2 - g_{ij}^k)/g_{ij}^k.$$

Working with r_{ij}^k has the advantage that it can be directly related to the first stage trade estimates. However, implementation of an estimating equation for r_{ij}^k has the problem that, for each country pair, r_{ij}^k takes two values – r_{ij}^k and its reciprocal, r_{ji}^k. We only want to take elements on either the left or the right hand parts of figure 4, thus ensuring one value for each country pair. However, simply selecting on the observed value of r_{ij}^k introduces sample selection bias as it involves censoring the distribution of the error in a regression of r_{ij}^k. We resolve this by selecting on the basis of predicted values, $\hat{r}_{ij}^k$, from our first stage analysis.

Correlations of market capacity and supply capacity

The normalization takes the form (for β_i^k and analogously for γ_j^k):

$$\overline{\beta}_i^k = \frac{\beta_i^k / \sum_i \beta_i^k}{\sum_k \beta_i^k / \sum_i \sum_k \beta_i^k}$$

The correlation matrices are given below, where the below diagonal elements are correlations of supply capacity and the above diagonal correlations of market capacity.

	AUS	AUT	BLX	CAN	CHE	DEU	DNK	ESP	FIN	FRA	GBR	GRC	IRL	ITA	JPN	NLD	NOR	NZL	PRT	SWE	TUR	USA
AUS	1.000	0.405	0.081	0.485	0.251	0.027	0.240	0.022	0.408	0.163	0.038	0.235	0.403	-0.165	-0.076	0.194	0.358	0.618	0.287	0.381	0.154	0.199
AUT	-0.235	1.000	0.512	0.459	0.665	0.569	0.731	0.449	0.712	0.642	0.468	0.488	0.703	0.436	-0.091	0.517	0.604	0.417	0.640	0.658	0.391	0.073
BLX	-0.052	0.641	1.000	0.137	0.373	0.617	0.477	0.544	0.406	0.567	0.547	0.266	0.466	0.476	-0.077	0.497	0.263	0.067	0.414	0.337	0.316	-0.083
CAN	0.179	-0.021	-0.070	1.000	0.359	0.089	0.369	0.041	0.462	0.275	0.210	0.141	0.363	-0.071	-0.189	0.173	0.381	0.339	0.160	0.366	0.130	0.506
CHE	-0.290	0.728	0.575	-0.110	1.000	0.482	0.511	0.167	0.398	0.387	0.457	0.095	0.532	0.150	-0.070	0.260	0.487	0.193	0.330	0.422	0.038	0.249
DEU	-0.208	0.710	0.683	-0.051	0.727	1.000	0.504	0.487	0.506	0.645	0.736	0.129	0.470	0.476	-0.172	0.587	0.440	0.011	0.286	0.566	0.310	-0.067
DNK	-0.042	0.717	0.546	0.060	0.521	0.534	1.000	0.503	0.603	0.636	0.405	0.576	0.683	0.374	-0.010	0.571	0.442	0.406	0.586	0.494	0.286	-0.088
ESP	-0.076	0.549	0.547	-0.027	0.502	0.510	0.482	1.000	0.443	0.600	0.350	0.550	0.455	0.705	-0.067	0.603	0.248	0.216	0.601	0.358	0.538	-0.300
FIN	-0.161	0.529	0.387	0.261	0.428	0.641	0.464	0.365	1.000	0.642	0.447	0.547	0.631	0.364	-0.218	0.501	0.590	0.382	0.518	0.708	0.480	0.127
FRA	-0.150	0.723	0.663	0.078	0.586	0.709	0.614	0.615	0.514	1.000	0.489	0.471	0.615	0.563	-0.164	0.634	0.372	0.187	0.529	0.497	0.453	-0.002
GBR	-0.088	0.677	0.672	-0.128	0.657	0.746	0.582	0.507	0.478	0.690	1.000	0.005	0.498	0.298	-0.306	0.462	0.466	0.035	0.173	0.519	0.242	0.148
GRC	-0.095	0.532	0.480	0.027	0.567	0.357	0.558	0.503	0.227	0.505	0.407	1.000	0.418	0.480	0.002	0.443	0.162	0.417	0.725	0.250	0.506	-0.202
IRL	0.080	0.409	0.508	0.098	0.351	0.424	0.485	0.392	0.393	0.546	0.488	0.273	1.000	0.370	-0.099	0.572	0.527	0.464	0.538	0.595	0.353	0.033
ITA	-0.174	0.719	0.613	-0.085	0.677	0.662	0.545	0.773	0.411	0.718	0.647	0.571	0.419	1.000	-0.022	0.396	0.147	0.017	0.588	0.194	0.573	-0.253
JPN	-0.314	0.609	0.507	-0.168	0.685	0.609	0.421	0.426	0.421	0.442	0.642	0.318	0.323	0.538	1.000	-0.009	-0.184	-0.145	-0.069	-0.184	-0.269	-0.418
NLD	-0.087	0.441	0.676	0.026	0.476	0.613	0.550	0.480	0.416	0.609	0.569	0.450	0.558	0.457	0.433	1.000	0.368	0.269	0.430	0.525	0.332	-0.093
NOR	-0.115	0.429	0.366	0.208	0.444	0.386	0.400	0.436	0.460	0.435	0.479	0.494	0.373	0.485	0.359	0.387	1.000	0.351	0.319	0.753	0.163	0.130
NZL	0.266	0.127	0.161	0.204	0.093	0.054	0.412	0.204	0.195	0.206	0.096	0.395	0.290	0.163	-0.138	0.245	0.333	1.000	0.419	0.268	0.326	0.044
PRT	-0.148	0.664	0.506	-0.117	0.548	0.469	0.678	0.644	0.296	0.509	0.484	0.587	0.330	0.665	0.469	0.356	0.334	0.214	1.000	0.345	0.474	-0.155
SWE	-0.075	0.497	0.446	0.265	0.381	0.591	0.412	0.423	0.693	0.558	0.507	0.134	0.518	0.466	0.459	0.448	0.550	0.124	0.269	1.000	0.180	0.060
TUR	-0.083	0.657	0.535	-0.040	0.528	0.468	0.555	0.631	0.273	0.556	0.452	0.659	0.212	0.654	0.435	0.391	0.248	0.235	0.668	0.184	1.000	0.003
USA	-0.146	0.122	-0.066	0.051	0.203	0.030	0.111	-0.045	0.081	0.026	-0.009	0.093	-0.084	0.000	0.133	-0.032	-0.069	-0.126	0.076	-0.020	0.103	1.000

Endnotes:

[1] See for example Balassa (1986a, 1986b), Balassa and Bauwens (1987, 1988a, 1988b), Bergstrand (1983), Culem and Lundberg (1986), Hummels and Levinsohn (1995), Loertscher and Wolter (1980), and Stone and Lee (1995).

[2] See Amiti and Venables (2002) for analysis of models that contain both effects.

[3] See Redding and Venables (2001) for a model deriving these effects.

[4] The separability assumption says that elasticity of a trade flow with respect to a change in a country characteristic is the same at all distances. Notice that the restriction does not hold if industry *k* is composed of many sub-industries in which trade volumes fall off with distance at different rates (see below for further discussion of aggregation issues).

[5] For example, the IIT measure records no intra-industry trade if either exports from country *i* to country *j* are zero, or exports from country *j* to country *i* are zero. The extra information derived from estimating the trade equations distinguishes between these cases.

[6] Results are not sensitive to the magnitude of the sum added. Alternative strategies for handling the problem of random variables which are bounded below at zero and with a mass point at zero are to use a sample selection approach that supplements the trade flow equation (8) with a selection mechanism specifying condition for observing $x_{ij}^{k} > 0$. The difficulty here is in determining valid identification restrictions. Alternatively one could adopt a two-part model in which the Pr ($x_{ij}^{k} > 0$) is modeled parametrically by a logit or probit model and $E(\ln (x_{ij}^{k}) \mid x_{ij}^{k} > 0)$ is assumed to be a linear function of a set of explanatory variables.

[7] This industry is 'representative' in so far as it is close to the median of the industries ranked by the t-statistics on the distance variable in column 4.

[8] Using a one-sided t-test of individual significance at the 5% level.

References:

Amiti, Mary and Anthony J. Venables, (2002) ‘The Geography of Intra-industry Trade’ in P. J. Lloyd and Hyun-Hoon Lee (eds.), *Frontiers of Research in Intra-Industry Trade*. Basingstoke: Palgrave, forthcoming.

Balassa, Bela (1986a) "Intra-Industry Trade Among Exporters of Manufactured Goods" in David Greenaway and P. K. M. Tharakan (eds.) *Imperfect Competition and International Trade.* Brighton: Wheatsheaf: 108 - 128.

Balassa, Bela (1986b) "Intra-Industry Specialization: A Cross-Country Analysis" *European Economic Review* 30(1): 27 - 42.

Balassa, Bela and Luc Bauwens (1987) "Intra-Industry Specialisation in a Multi-Country and Multi-Industry Framework" *Economic Journal* 97(388): 923 - 939.

Balassa, Bela and Luc Bauwens (1988a) "Inter-Industry and Intra-Industry Specialization in Manufactured Goods" *Weltwirtschaftliches Archiv* 124(1): 1 - 13.

Balassa, Bela and Luc Bauwens (1988b) "The Determinants of Intra-European Trade in Manufactured Goods" *European Economic Review* 32(7): 1421 - 1437.

Bergstrand, Jeffrey H. (1983) "Measurement and Determinants of Intra-Industry International Trade" in P. K. M. Tharakan(ed.) *Intra-Industry Trade: Empirical and Methodological Aspects.* Amsterdam: North-Holland: 201 - 262.

Culem, Claudy and Lars Lundberg (1986) "The Product Pattern of Intra-Industry Trade: Stability among Countries and over Time" *Weltwirtschaftliches Archiv* 122(1): 113 - 130.

Feenstra, Robert C. (2000) "World Trade Flows, 1980-1997" Centre for International Data, University of California, Davis, CA. http://data.econ.ucdavis.edu/international/pdf/wtf.pdf

Grubel, Herbert G. and Peter J. Lloyd (1975) *Intra-Industry Trade*. London: Macmillan.

Harrigan, J. (1995) ‘The volume of trade in differentiated intermediate goods; theory and evidence’, *Review of Economics and Statistics*, 77, 283-293

Helpman, Elhanan and Paul R. Krugman (1985) *Market structure and foreign trade; imperfect competition and the international economy*. Cambridge, MA: MIT Press.

Hummels, David and James Levinshohn (1995) "Monopolistic Competition and International Trade: Reconsidering the Evidence" *Quarterly Journal of Economics* 110(3): 799-836.

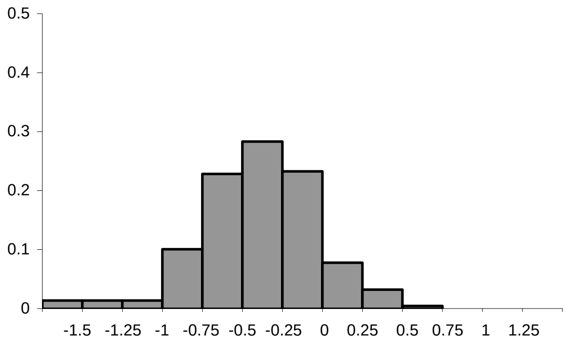
Loertscher, Rudolf and Frank Wolter (1980) "Determinants of Intra-Industry Trade: Among Countries and Across Industries" *Weltwirtschaftliches Archiv* 116(2): 280-293.

Pagoulastos, Emilio and Robert Sorensen (1975) "Two-Way International Trade: An Econometric Analysis" *Weltwirtschaftliches Archiv* 111(3): 454-465.

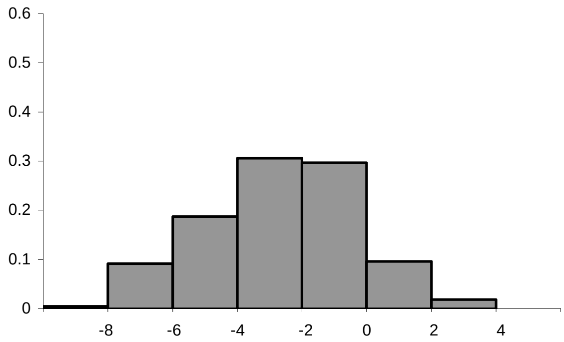
Redding, Stephen and Anthony J. Venables, (2001) "Economic Geography and International Inequality" CEPR, London.

Stone, Joe A. and Hyun-Hoon Lee (1995) "Determinants of Intra-Industry Trade: A Longitudinal, Cross-Country Analysis" *Weltwirtschaftliches Archiv* 131(1): 67 - 85.

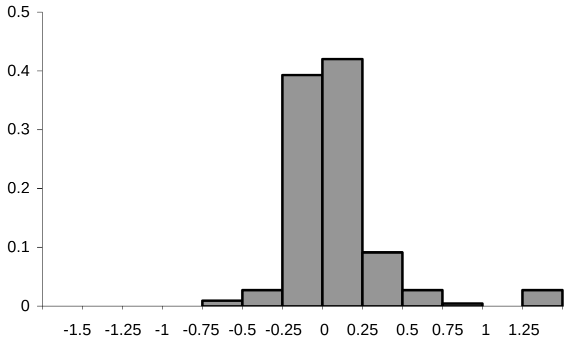
Verdoorn, P. J. (1960) "The Intra-Bloc Trade of Benelux" in E. A. G. Robinson (ed.) *Economic Consequences of the Size of Nations.* London: Macmillan: 291 - 329.



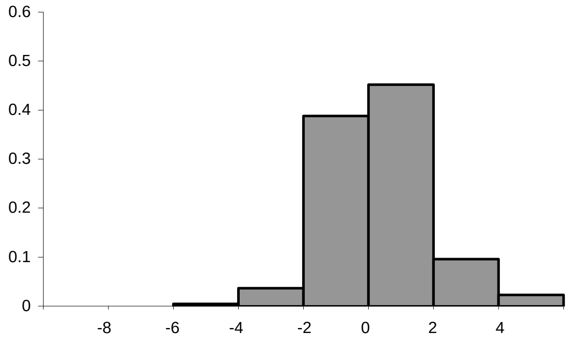
1a: distance coefficient - without $\hat{r}_{ij}^{k}$



1b: distance t-statistic - without $\hat{r}_{ij}^{k}$

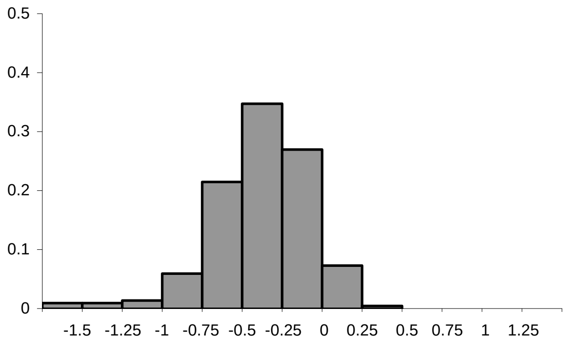


1c: distance coefficient - with $\hat{r}_{ij}^{k}$

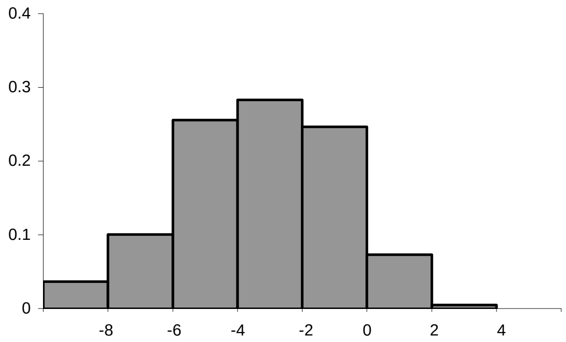


1d: distance t-statistic - with $\hat{r}_{ij}^{k}$

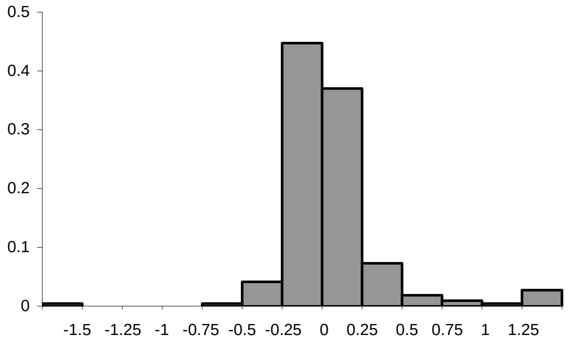
Figure 1: Industry analysis using the ratio measure



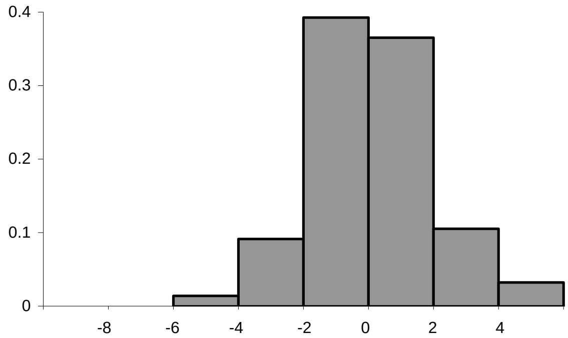
2a: distance coefficient - without $\hat{g}_{ij}^{k}$



2b: distance t-statistic - without $\hat{g}_{ij}^{k}$

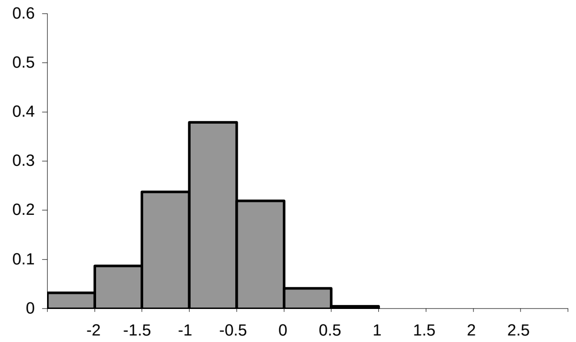


2c: distance coefficient - with $\hat{g}_{ij}^{k}$

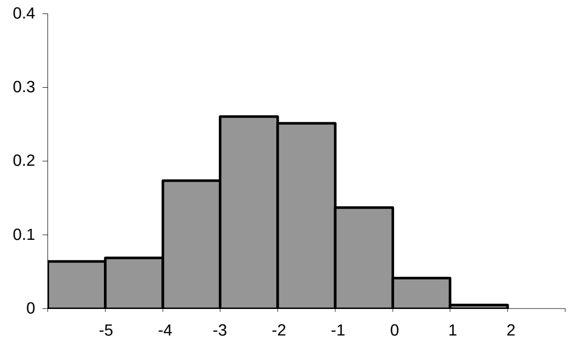


2d: distance t-statistic - with $\hat{g}_{ij}^{k}$

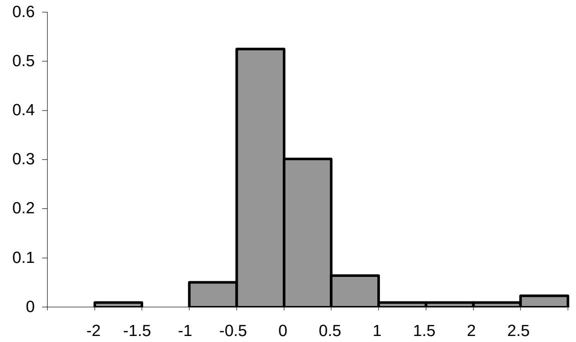
Figure 2: Industry analysis using the Grubel-Lloyd index



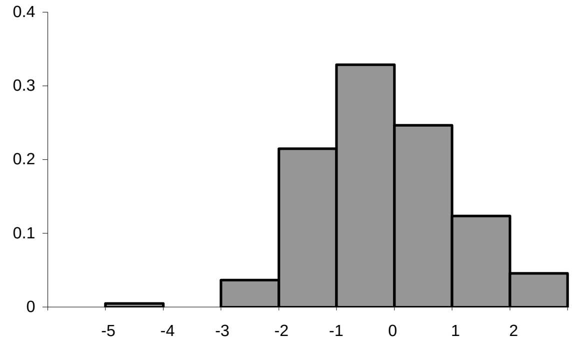
3a: distance coefficient - without $\hat{r}_{ij}^{k}$



3b: distance t-statistic - without $\hat{r}_{ij}^{k}$



3c: distance coefficient - with $\hat{r}_{ij}^{k}$



3d: distance t-statistic - with $\hat{r}_{ij}^{k}$

Figure 3: Industry analysis of Europe using the ratio measure

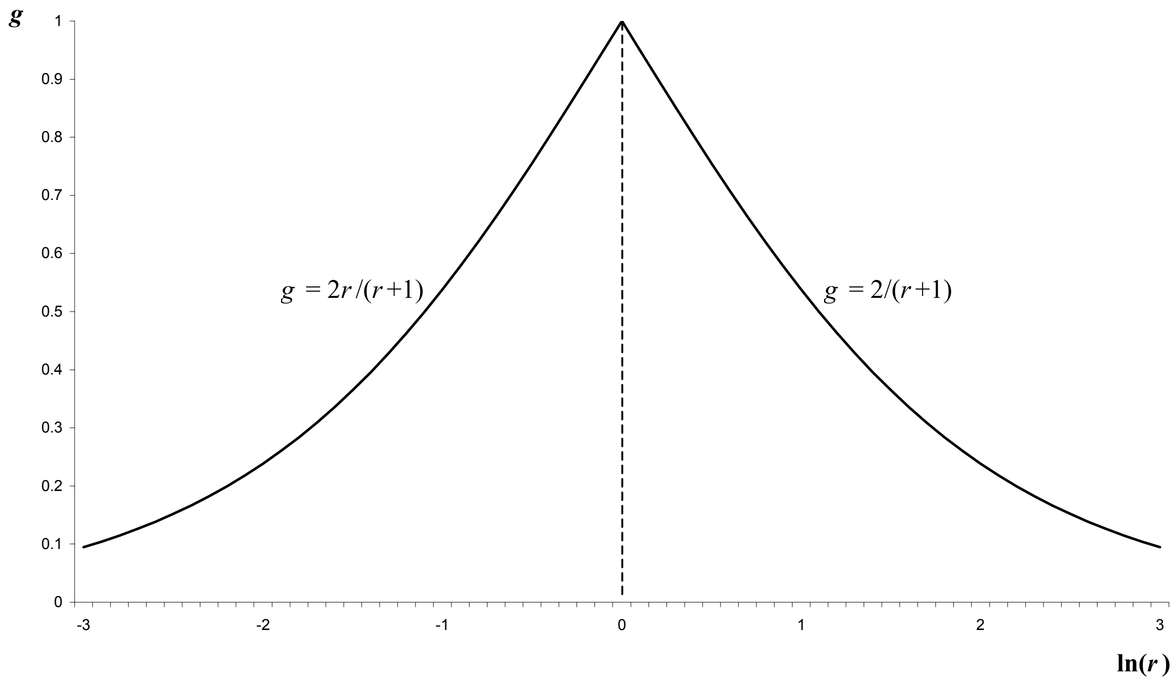


Figure 4: Grubel-Lloyd and ratio measures of IIT