# Comparative advantage and economic geography : estimating the determinants of industrial location in the EU

Karen Helene Midelfart-Knarvik, Henry G. Overman, and Anthony Venables

Jan 23 2001:

# Comparative advantage and economic geography: estimating the determinants of industrial location in the EU. *


Karen Helene Midelfart-Knarvik
NHH, Bergen and CEPR

Henry G. Overman
LSE and CEPR

Anthony J. Venables
LSE and CEPR



## Abstract:

We develop and econometrically estimate a model of the location of industries across countries. The model combines factor endowments and geographical considerations, and shows how industry and country characteristics interact to determine the location of production. We estimate the model on sectoral data for EU countries over the period 1980-97, and find that endowments of skilled and scientific labour are important determinants of industrial structure, as also are forward and backward linkages to industry.



JEL number: F10
Keywords: specialization, comparative advantage, economic geography.



* This paper is produced as part of the Globalization programme of the ESRC funded Centre for Economic Performance at the LSE. Thanks to G.Ottaviano, D.Puga, S.Redding, P.Seabright, D. Trefler and seminar participants at LSE, Toronto, NHH (Bergen), IUI (Stockholm), the European Research Workshop on International Trade (Copenhagen), and the Empirical Investigations in International Trade Conference (Boulder).



Authors' addresses:

K.H. Midelfart-Knarvik
Dept of Economics
Norwegian School of Economics and Business Administration
Helleveien 30
5045 Bergen, Norway
KarenHelene.Knarvik@snf.no

H.G. Overman
Dept of Geography
LSE
Houghton Street
London WC2A 2AE
h.g.overman@lse.ac.uk

A.J. Venables
Dept of Economics
LSE
Houghton Street
London WC2A 2AE
a.j.venables@lse.ac.uk

## 1. Introduction

This paper addresses the question: what determines the location of different industries across countries? Theory tells us that it depends on supply considerations, on the cross-country distribution of demand for each sector's output, and on the ease of trade. In the case in which trade is perfectly free, then the distribution of demand becomes unimportant and supply alone determines the location of production. This is the basis of the textbook models in which comparative advantage (as driven by technology or endowment differences) determines the structure of production in each country. More generally, the presence of transport costs or other trade frictions mean that both supply and demand matter. If transport costs vary systematically with distance then geographical factors come in to play, combining with comparative advantage to determine industrial location.

The objective of this paper is to develop and econometrically estimate a model combining comparative advantage and geographical forces.[1] Our model contains countries that have differing factor endowments and face transport costs on their trade. Industrial sectors use primary factors and intermediate goods to produce differentiated goods, differentiation ensuring that there are positive trade flows, despite transport costs. The equilibrium pattern of industrial location is determined both by factor endowments and geography. Factor endowments matter for the usual reasons, although factor prices are not generally equalized by trade. Transport costs mean that the location of demand matters; countries at different locations have different market potential, and this shapes their industrial structures. Intermediate goods prices and demand vary across locations, meaning that forward and backward linkage effects are present and that industries will tend to locate close to supplier and customer industries.

Our task is to combine these effects and show how they impact differently on different sectors. All industries would, other things being equal, tend to locate in countries with abundant factor supplies, good market access, and proximity to suppliers. In general equilibrium, what are the characteristics of industries that lead them to locate in countries of different types? We illustrate the answer to this, showing how it is possible to generalise the Rybczynski and Heckscher-Ohlin effects of standard models. We then linearise the model, and show how characteristics of countries (such as their endowments or location) interact with characteristics of industries (such as their factor intensity or transport costs) to determine production structure. This linearisation provides the equation that we estim

Estimation is undertaken using data for 33 industries and 14 European Union countries, for the period 1980-97. This data set has the advantages of having a relatively straightforward geography – with a clear set of central and peripheral countries – and of covering a period of increasing economic integration. Studies of production find evidence that the specialization of European countries has increased through this period.[2] We are able to provide some insight into the roles of comparative advantage and geography in driving these changes.

Our approach can be viewed as both a synthesis and a generalization (in some directions) of two approaches in the existing empirical literature. There is a sizeable literature (dating from Baldwin 1971) that estimates the effect of industry characteristics on trade, running cross-industry regressions for a single country.[3] A more recent literature (for example Leamer 1984, Harrigan 1995, 1997) estimates the effect of country characteristics (endowments and possibly also technology) on trade and production, running cross-country regressions and estimating industry by industry. Our approach takes the panel of industries *and* countries, and estimates the way in which production depends on *both* industry characteristics and country characteristics, with the form of the interaction between these effects dictated by theory. This approach is perhaps closest to Ellison and Glaeser (1999) who analyse how industrial location in US states is affected by a range of 'natural advantages'. Our paper differs from Ellison and Glaeser in deriving the theoretical specification from trade, rather than location, theory. As a result, our interactions more clearly relate both to countries' factor endowments and to their relative locations.

Recent work by Davis and Weinstein (1998, 1999) combines comparative advantage and geography by assuming that the broad sectoral pattern of specialization (3 digit) is determined by endowments, and the finer detail of 4 digit production determined by either geography or endowments. They investigate the effect of demand shocks on production, in order to test for home market effects. Our model does not make this two level separation, and the question we address is broader, in so far as we are looking at how a variety of different forces interact to determine location. However, our model is narrower than Davis and Weinstein's in so far as we assume throughout that all sectors are perfectly competitive. While geography can, of course, have a bearing on industrial location in a competitive environment, this restriction means that some of the forces of new economic geography are absent from our approach. We make this assumption in order to have a precise and tractable link between the theory and the econometrics, whereas adding imperfect competition would raise a number of issues which go beyond the scope of this paper. For example, in such an environment there may be a multiplicity of equilibria, and hence no unique mapping from underlying characteristics of countries and industries to industrial location[4]. Addressing these issues will be the subject of future researc

The paper proceeds as follows. Section 2 outlines the analytical framework, and section 3 derives our estimating equation. Section 4 discusses the interactions in this equation, and section 5 presents econometric results. Section 6 concludes.

## *2. The model*

The model contains $I$ countries, $K$ industrial sectors and $M$ primary factors. All industries are perfectly competitive and have constant returns to scale, using primary factors and intermediate goods as inputs. Each industry produces a number of varieties of differentiated products; we denote the number of varieties produced in country $i$ by industry $k$ by $n_i^k$, and assume that this is determined exogenously. Goods are tradeable but incur transport costs, the level of which is industry specific and depends on the source and destination country; thus $t_{ij}^k$ denotes the iceberg mark-up on shipping industry $k$ products from country $i$ to country $j$.

With this structure, the value of production of each industry in each country is determined by factor supply, by the prices of intermediate goods, and by the geographical distribution of demand. One limiting case is when product varieties in all industries are perfect substitutes and the model reduces to a pure factor endowment model of trade, with all the usual properties of such a model. More generally, the presence of product differentiation means that factor prices are not independent of endowments, that there is trade in all goods (despite trade costs), and that there is a determinate structure of production (even if there are more industries than factors).

2.1: *Technology*

The $n_i^k$ industry $k$ product varieties produced in country $i$ are symmetrical, i.e. face the same cost and demand functions. Input prices in country $i$ are denoted by the vector $v_i$, and the costs of industry $k$ in country $i$ are given by unit cost function $c(v_i : k)$. F.o.b. prices equal unit costs, so

$$p_i^k = c(\boldsymbol{v_i} : k). \tag{1}$$

Iceberg transport costs of $(t_{ij}^k - 1)$ are incurred in shipping product $k$ from $i$ to $j$, so the c.i.f. price of industry $k$ goods produced in $i$ and sold in $j$ is $c(v_i : k)t_{ij}^k$

*2.2: Demand and sales*

Total expenditure on the products of industry $k$ in country $j$ is denoted $e_j^k$. This is divided between different varieties that are aggregated according to a CES function, implying a price index for industry $k$ in country $j$ of,

$$G_j^k = \left[\sum_i n_i^k \left(c(\boldsymbol{v_i}:k)\, t_{ij}^k\right)^{1-\eta}\right]^{1/(1-\eta)} \tag{2}$$

where $\eta$ is the elasticity of substitution between product varieties, assumed to be the same in all industries.[5] The value of demand for a single variety produced in $i$ and sold in $j$ is then $\left(c(\boldsymbol{v_i}:k)t_{ij}^k\right)^{1-\eta} e_j^k \left(G_j^k\right)^{\eta-1}$, as usual from a CES demand system. Summing this over all the $n_i^k$ product varieties produced by industry $k$ in country $i$, and over all markets, $j$, gives the following expression for the value of production of industry $k$ in country $i$, $z_i^k$;

$$z_i^k = n_i^k c(\boldsymbol{v_i}:k)^{1-\eta} \sum_j \left(t_{ij}^k\right)^{1-\eta} e_j^k \left(G_j^k\right)^{\eta-1}. \tag{3}$$

In what follows it will be convenient to take the total value of production as numeraire, so $\sum_i \sum_k z_i^k = 1$; $z_i^k$ is then the industry - country production share. We also define the share of country $i$ in total production as $s_i$, ($s_i \equiv \sum_k z_i^k$) and the share of industry $k$ as $s^k$, ($s^k \equiv \sum_i z_i^k$). The number of product varieties of each industry produced by each country is exogenous, and set in proportion to the size of industry and country, up to an error term $\epsilon_i^k$, so,

$$n_i^k = s_i s^k \exp[\epsilon_i^k]. \tag{4}$$

If industries were monopolistically competitive then the scale of output of each variety (and firm) would be fixed by zero profits, and the values of $n_i^k$ would be endogenously determined by free entry and exit; cross-country output variation would therefore be due to differing numbers of varieties in each country. Here numbers of varieties are set by (4), but output levels of each variety can vary according to the forces given in equation (3).

*2.3: Input prices*

Inputs consist of primary factors and a single composite intermediate good,[6] with prices $w_i$ and $q_i$ respectively, so $v_i = [w_i, q_i]$. Prices of the primary factors, $w_i$, are determined by market clearing, which can be expressed as

$$\boldsymbol{L_i} = \sum_k x_i^k \boldsymbol{c_w}(\boldsymbol{w_i}, q_i : k) \tag{5}$$

where $L_i$ is the endowment vector of country $i$, $c_w(w_i, q_i : k)$ is the vector of partial derivatives of the $k$th industry's unit cost functions with respect to primary factor prices, and country $i$ industry output levels are $x_i^k \equiv z_i^k/p_i^k$.

The composite intermediate is a Cobb-Douglas aggregate of output from different industries, each with price index $G_i^k$. The price of the intermediate good in country $i$ is therefore

$$q_i = \Pi_k \left(G_i^k\right)^{\phi^k}, \qquad \sum_k \phi^k = 1, \tag{6}$$

where $\phi^k$ is the share of industry $k$ in the intermediate good.

*2.4: Expenditures*

Expenditures on each industry, $e_i^k$, come from final expenditure and intermediate demands. The former are fixed shares, $\alpha^k$, of income, $y_i$, in each country. The latter is the value of total intermediate demand in country $i$ (derived using

Shephard's lemma on industry cost functions) times the share attributable to sector $k$, $\mathrm{N}^k$,

$$e_i^k \; = \; \alpha^k y_i \; + \; \phi^k q_i \sum_k x_i^k \boldsymbol{c}_{\boldsymbol{q}}(\boldsymbol{w}_{\boldsymbol{i}}, q_i : k). \tag{7}$$

Income, $y_i$, is derived from primary factors in the usual way.

### *3. Estimating equation*

Our estimating equation is based on the output of each industry in each country, expressed relative to the size of the industry and the country. We denote this double relative measure $r_i^k$, and using (3) and (4) it takes the fo

$$r_i^k \; \equiv \; z_i^k / s_i s^k \; = \; c(\boldsymbol{v}_{\boldsymbol{i}} : k)^{1-\eta} \sum_j \left(t_{ij}^k\right)^{1-\eta} e_j^k \left(G_j^k\right)^{\eta-1} \exp[\epsilon_i^k]. \tag{8}$$

The term in the summation sign captures demand effects, and we refer to these as the market potential of industry $k$ in country $i$. We shall denote this $m(u^k : i)$,

$$m(\boldsymbol{u}^{\boldsymbol{k}} : i) \; \equiv \; \sum_j \left(t_{ij}^k\right)^{1-\eta} e_j^k \left(G_j^k\right)^{\eta-1}. \tag{9}$$

The market potential function, $m(u^k : i)$, is indexed across countries, $i$, and the vector $u^k$ refers to industry characteristics that interact with the spatial distribution of demand (such as transport costs; we discuss these in the next section). Equation (8), for the output of each industry in each country, now becomes

$$r_i^k \; = \; c(\boldsymbol{v}_{\boldsymbol{i}} : k)^{1-\eta} m(\boldsymbol{u}^{\boldsymbol{k}} : i) \exp[\epsilon_i^k] \tag{10}$$

This says that systematic cross-country variation in sectors' output is determined by two sorts of considerations. One is input price variation, captured in the unit cost function. The other is demand variation, captured by the market potential of industry $k$ in country $i$.[7]

To estimate the model we log-linearise equation (10) around a reference point. This is simplified by noting that the functions $c(v_i : k)$ and $m(u^k : i)$ can be constructed such that: (i) there exists an input price vector, $\bar{\boldsymbol{v}}$, at which $c(\bar{\boldsymbol{v}} : k) \; = \; 1$ for all industries $k$, and: (ii) there exists a vector of industry characteristics, $\bar{\boldsymbol{u}}$, such that $m(\bar{\boldsymbol{u}} : i) \; = \; 1$ for all countries $i$[8]. These define our reference country and industry. Linearising (7) around the reference point gives

$$\Delta r_i^k = (1 - \eta)\gamma(\bar{\boldsymbol{v}}:\boldsymbol{k}).\Delta \boldsymbol{v_i} + \mu(\bar{\boldsymbol{u}}:\boldsymbol{i}).\Delta \boldsymbol{u^k} + \epsilon_i^k \tag{11}$$

where $\Delta$ denotes a proportionate change, e.g. $\Delta v_i = d\log(v_i) \approx \log(v_i) - \log(\bar{v})$; $\boldsymbol{\gamma(\bar{v}:k)} \equiv (\partial c/\partial v)(c/\bar{v})$ is the (row) vector of elasticities of industry $k$ costs with respect to input prices, $v_i$, (these elasticities also being the share of each input in costs); and $\boldsymbol{\mu(\bar{u}:i)} \equiv (\partial m/\partial u)(m/\bar{u})$ is the vector of elasticities of country $i$ market potential with respect to industry characteristics, $u^k$.[9]

Since the $r_i^k$ are shares, deviations must satisfy an adding up condition, $\sum_i \sum_k z_i^k \Delta r_i^k = 0$. Using this with equation (11) gives

$$\begin{aligned}\Delta r_i^k = (1 - \eta)[\gamma(\bar{\boldsymbol{v}}:\boldsymbol{k}) - \textstyle\sum_i \sum_k z_i^k \gamma(\bar{\boldsymbol{v}}:\boldsymbol{k})].\Delta \boldsymbol{v_i} \; + \\ [\mu(\bar{\boldsymbol{u}}:\boldsymbol{i}) - \textstyle\sum_i \sum_k z_i^k \mu(\bar{\boldsymbol{u}}:\boldsymbol{i})].\Delta \boldsymbol{u^k} + \epsilon_i^k .\end{aligned} \tag{12}$$

The double summation terms in (12) do not vary over either industries or countries, so we write

$$\Delta r_i^k = (1 - \eta)[\gamma(\bar{\boldsymbol{v}}:\boldsymbol{k}) - \bar{\gamma}].\Delta \boldsymbol{v_i} + [\mu(\bar{\boldsymbol{u}}:\boldsymbol{i}) - \bar{\mu}].\Delta \boldsymbol{u^k} + \epsilon_i^k . \tag{13}$$

Terms on the right hand side of (13) are the product of country and industry characteristics, both expressed in deviation form. The first inner product gives countries' input prices times industries' input shares (elasticities of costs with respect to input prices). The second gives industries' characteristics times elasticities of countries' market potentials with respect to these characteristics. In the econometric implementation of the model we use six interactions, and we now explore each of these in turn.

## *4. Interactions*

### *4.1: Primary factors;*

On the cost side, input prices include both primary factor prices and intermediates good prices, partitioned into $\boldsymbol{v_i} = [\boldsymbol{w_i}:q_i]$. The corresponding vector of input shares is $\boldsymbol{\gamma} = [\gamma_{\boldsymbol{w}}:\gamma_q]$.

For primary factors we go back to factor endowments rather than use factor prices, since the latter are endogenous. The vector of factor price variations, $\Delta w_i$, depends on endowments according to

$$\Delta \boldsymbol{w_i} = \boldsymbol{F}.\Delta \boldsymbol{L_i} \tag{14}$$

where $\Delta L_i$ is the vector of variations in endowments from the reference point, and $F$ is the matrix of elasticities of factor prices with respect to endowments, evaluated at the reference point. Using this in equation (13) and ignoring all other effects, gives

$$\Delta r_i^k = (1 - \eta)[\gamma_w(\bar{v}:k) - \bar{\gamma}_w].F.\Delta L_i \tag{15}$$

Several points need to be made about this equation. The matrix $F$ is derived by totally differentiating the factor market clearing condition, (5), letting techniques of production and output quantities change. Details of the derivation are given in appendix A1, which also derives explicit expressions for the two-industry two-factor case. It shows how, as $\eta \to \infty$, the model produces standard Rybczynski effects, and factor prices become invariant with respect to endowments. Although the sign pattern of the matrix $F$ and of Rybczynski effects are unambiguous in the 2 x 2 case, signs in higher dimension models are not clear-cut, as Leamer (1987) has pointed out. In implementing the model we shall simply assume that diagonal elements of $F$ are much larger than off-diagonal – i.e., only include the effects of each factor endowment on the price of that factor, ignoring effects on other factors.[10] This assumption ensures that an increase in the endowment of a factor increases output in industries that are intensive users of the factor.

The relationship given in equation (15) is a linear approximation to the way in which factor intensities and factor endowments interact to determine production. It is worth comparing it to the exact relationship, which can be derived by simulation of the model (using a deterministic structure with $\varepsilon_i^k = 0$, and parameters given in appendix A2). The simulation output is given in Figure 1. Countries differ only in their relative endowment of a single factor, $\nu$, and one of the horizontal axes ranks countries according to $\log(\nu_i)$, (over the set of countries $i \in I$). Industries differ only in the share of this factor in costs, which we denote $\gamma^k$, and are ranked along the other axis in the horizontal plane.[11] The surface of the figure plots output levels, $\log(r_i^k)$, as a function of $\log(\nu_i)$ and $\gamma^k$.

As expected, $\nu$-abundant countries have high production in industries in which the share of this factor is large (high $\gamma^k$) and low production in industries where it is low, giving a saddle shaped surface. The arrow marked R on the surface indicates how, in a particular industry, production varies with factor endowments; moving to more $\nu$-abundant countries increases output for products with high $\gamma$, and decreases it for products with low $\gamma$. Some intermediate industry, with factor intensity $\bar{\gamma}$, has output level independent of the endowment of this factor.[12] The arrow marked H shows how, for a particular country, the structure of production depends on its factor endowment; an $\nu$-scarce economy has relatively high production in low $\gamma$ industries, and so on. The effects illustrated by the R and H arrows can be thought of as generalizations of Rybczynski and Heckscher-Ohlin-Samuelson effects, showing how output of each industry depends on factor endowments, and how the structure of production of each country depends on factor intensities.

The relationship of figure 1 to equation (15) should be clear. The quadratic form of (15), with deviations of endowments from a reference point multiplied by deviations of factor shares from reference values is a good approximation to the saddle-shaped surface of the figure.

In estimation we work with three primary factors. Data is available for five factor endowments which broadly correspond to researchers and scientists, skilled labour, unskilled labour, capital and agriculture. We exclude capital from estimation on the grounds that it is internationally mobile and has the same price throughout the EU, and also drop unskilled labour, since the shares of all three types of labour in the labour endowment are not independent. This leaves researchers and scientists, skilled labour and agriculture. For agriculture, rather than using land endowments we use output of agriculture, forestry and fishery products.[13] Details are given in appendices A3 (construction of variables) and A4 (data sources).

*4.2: Intermediate goods supply: forward linkages*

If there are transport costs then the prices of intermediate goods vary across countries. The model assumes a single composite intermediate good, and the cross-country variation in the price of this good, $\Delta q_i$, interacts with cross-industry variation in intermediate input shares, $\gamma_q(\bar{v}:k)$, to determine output. From equation (13) (ignoring other terms), this interaction takes the form,

$$\Delta r_i^k = (1 - \eta)[\gamma_q(\bar{v}:k) - \bar{\gamma}_q]\Delta q_i. \tag{16}$$

The relationship expressed here is simply 'forward linkages'. It could be illustrated in a figure similar to figure 1, showing how industries with high intermediate shares are drawn into locations with good access to supply of intermediates, and vice versa.

To implement this econometrically we need data on intermediate input shares, which are readily available (appendix A4). The price of the intermediate good is given by equation (6) as a Cobb-Douglas function of the price indices of each industry, and we show in appendix A3 how to construct the variable $\Delta q_i$.

*4.3: Market potential and transport costs*

Demand considerations enter our estimating equation through the interaction between industry characteristics and the elasticities of market potential with respect to these characteristics. The first industry characteristic we look at is transport intensity, which we model as follows. Suppose that transport costs are an isoelastic function of the distance between locations *i* and *j*, and that that the elasticity is industry specific and measures 'transport intensity'. Then, if distance is denoted $d_{ij}$, and the transport intensity of industry $k$, $\theta^k$, we have

$$t_{ij}^{k} = d_{ij}^{\theta^k}. \tag{17}$$

Estimates of transport intensity are obtained from the GTAP modelling project (see appendix A3). Using this expression, the market potential of industry $k$ in country $i$ (equation (9)), becomes

$$m(\boldsymbol{u}:i) \equiv \sum_j \left(d_{ij}^{\theta^k}\right)^{1-\eta} e_j^{\,k}\left(G_j^{\,k}\right)^{\eta-1}. \tag{18}$$

Linearization requires that we find the elasticity of this with respect to transport intensity, evaluated for some reference industry. Details of the procedure for doing this are given in appendix A3.

Figure 2 shows how transport intensity interacts with market potential to determine industrial location. The figure is computed for an example in which the only difference between industries is in transport intensity, and the only difference between countries is that some are closer to other markets so have higher market potential. The figure plots out the surface of $\log(r_i^k)$ against industries' transport intensities and countries' computed market potentials. For the range of transport intensities shown the surface is saddle-shaped and, as expected, production in high transport intensity industries tends to concentrate in countries with high market potential. However, we should note that this saddle shape is not a global property of the surface – a non-tradable industry would evidently have production determined solely by local demand, not by countries' reference market potential.[14]

*4.4: Intermediate goods demand; Backward linkages*

Transport intensity is not the only industry characteristic that interacts with countries' location. A further interaction arises from the fact that the spatial pattern of demand, $e_i^k$, may differ across industries. This could in principle be due to final expenditure differences, although the identical homothetic structure of preferences embodied in equation (7) rules this out. Alternatively, it may be due to the spatial distribution of *derived* demand varying across industries -- backwards linkages. We expect upstream industries to locate in countries in which derived demand is hig

As usual, equation (13) says that this should be modelled as an interaction between an industry characteristic and a country characteristic. The industry characteristic is the share of each industry's output that goes for intermediate usage. The country characteristic is the elasticity of market potential with respect to this share, which turns out to depend on the difference between a market potential based on final expenditure and a market potential based on intermediate expenditures (derived in appendix A3). Interacting these gives rise to the usual saddle shaped relationship, with upstream industries wanting to locate in countries in which the market potential from intermediate sales is high relative to the market potential from final sales.

## 5. *Estimation*

We now turn to econometric implementation and estimation of the structure outlined above. We estimate equation (13) with the interactions described in section 4. For compactness we now simply number these interactions 1 - 6, and list them in table 1. We denote the country and industry characteristics $x_i[j]$ and $y^k[j]$ respectively, with j an index running over the six interactions. The specification is then,

$$\ln(r_i^k) = \alpha + \sum_j \beta[j]\left(x_i[j] - \bar{x}[j]\right)\left(y^k[j] - \bar{y}[j]\right) + \epsilon_i^k \tag{19}$$

where a bar over a variable denotes the reference value, as before. Expanding the relationship gives the estimating equation:

$$\ln(r_i^k) = \xi + \sum_j \left(\beta[j]x_i[j]y^k[j] - \beta[j]\bar{y}[j]x_i[j] - \beta[j]\bar{x}[j]y^k[j]\right) + \epsilon_i^k. \tag{20}$$

The coefficients to be estimated are $\beta[j]$, measuring the importance of the interaction; $\beta[j]\bar{y}[j]$ and $\beta[j]\bar{x}[j]$ giving level effects in the interaction; and a constant, $\xi$, containing the sum (over $j$) of the products of all the level effects. The interactions are summarised in Table 1.

Table 1: Interactions

| | Country Characteristic: $x_i[j]$ | | Industry Characteristic: $y^k[j]$ | |
|---|---|---|---|---|
| *j = 1* | Agricultural endowment | log | Agricultural intensity | Elasticity |
| *j = 2* | Skilled labour endowment | log | Skill intensity | Elasticity |
| *j = 3* | Researchers and Scientists | log | R&D intensity | Elasticity |
| *j = 4* | Supplier access (eqn. 32) | log | Intermediate intensity | Elasticity |
| *j = 5* | Elasticity of market potential w.r.t transport costs (eqn. 34) | Elasticity | Transport costs | log |
| *j = 6* | Relative market potential (eqn. 38) | Elasticity | Share of output to industry | log |

### *5.1 Data and estimation:*

Our data is for 14 EU countries and 36 manufacturing industries, although we omit three sectors – petroleum refineries, petroleum and coal products (whose location is predominantly natural resource driven) and manufacturing not elsewhere classified - essentially a residual component. We have data on output from 1980-1997, but because we cannot get full time series for all the independent variables, we pick four time periods.[15] Within each of these periods we time average to remove business cycle fluctuations, leaving us with the cross–sections: 1980-83, 1985-88, 1990-93, 1994-97. Independent variables are measured as near to the beginning of each time period as possible. See Appendix A4 for more details.

The equation is estimated by OLS. There are potentially two important sources of heteroscedasticity - both across countries and across industries. Because we cannot be sure whether these are important, or which would dominate, we report White's heteroscedastic consistent standard errors and use these consistent standard errors for all hypothesis testing.[16]

We report standardized coefficients by conditioning on the standard deviation of the underlying variables. This normalisation means that coefficients on the interaction terms have the following two interpretations. They measure the elasticity of output with respect to a country characteristic, for an industry with corresponding industry characteristic one standard deviation above the its reference value. And symmetrically, they measure the elasticity of output with respect to an industry characteristic, for a country with corresponding characteristic one standard deviation above the reference level. The use of standardised variables allows us to compare across different time periods and different interactions without having to worry about differences in the variances of the underlying endowments and intensities.

### *5.2 Results*

We start by pooling across the four time periods.[17] The results are shown in the second column of Table 2, where we give the coefficients on each of the six interaction terms. The ordering of the variables is as per Table 1, so the first three coefficients correspond to the comparative advantage interactions (section 4.1) and the last three coefficients correspond to economic geography interactions (sections 4.2-4.4). Table A2 in Appendix A5 reports the additional coefficients on country and industry characteristics. Thus, in terms of Figures 1 and 2, Table 2 reports the coefficients that give the slope of the surface, while relegating the coefficients which determine the position of the saddle (the reference points) to the appendix.

We can see from the results (second column) that the comparative advantage variables have the same signs as predicted by theory and are significant at the 5% level or better. The coefficients are smaller for agricultural than for skill and R&D intensity, indicating lower elasticities, i.e., that the related endowments have a weaker impact on production shares. Turning to the economic geography variables, we see that the coefficients for forward ($\beta[4]$) and backward linkages ($\beta[6]$) have the right sign and are significant at the 5% and 1% level respectively. The coefficient for backward linkages is much smaller than that for forward linkages, suggesting that supplier access is more important in determining location of production than access to industrial customers. The final economic geography variable $\beta[5]$ capturing the interaction between industry transport costs and the distribution of demand is significant, but has the opposite sign to that expected from theory. It is unclear how to interpret this result, particularly as the coefficient becomes insignificant when we allow for parameter heterogeneity across time – the issue to which we now

Table 2: Regression results: Dependent variable $\ln(r_{ij}^{k})$

| | Variable | Pooled | 1980-83 | 1985-88 | 1990-93 | 1994-97 |
|---|---|---|---|---|---|---|
| | Interactions: β[j] | | | | | |
| β[1] | Agric. endowment | 0.111** | 0.078 | 0.140 | 0.166** | 0.158** |
| | * agricultural inputs | (0.046) | (0.114) | (0.097) | (0.085) | (0.079) |
| β[2] | Skill endowment | 1.600*** | 1.503*** | 1.484*** | 1.479*** | 1.663*** |
| | * skill intensity | (0.228) | (0.439) | (0.420) | (0.463) | (0.582) |
| β[3] | Researchers+scientists | 0.602*** | 0.584* | 0.741** | 1.108** | 1.624*** |
| | * R&D intensity | (0.196) | (0.325) | (0.389) | (0.536) | (0.581) |
| β[4] | Supplier access | 0.763** | 0.570 | 0.754 | 0.799 | 1.096* |
| | * intermed. intensity | (0.356) | (0.811) | (0.771) | (0.667) | (0.689) |
| β[5] | Market pot. elasticity | -0.356** | -0.395 | -0.270 | -0.319 | -0.382 |
| | * transport costs | (0.148) | (0.315) | (0.299) | (0.290) | (0.275) |
| β[6] | Relative market pot. | 0.138*** | 0.182*** | 0.171*** | 0.130*** | 0.083** |
| | * output to industry | (0.024) | (0.059) | (0.052) | (0.043) | (0.041) |
| | Diagnostics | | | | | |
| | $R^2$ | 0.145 | 0.140 | 0.151 | 0.177 | 0.171 |
| | Adjusted $R^2$ | 0.136 | 0.105 | 0.116 | 0.143 | 0.137 |
| | Number of obs | 1824 | 456 | 456 | 456 | 456 |

Note: Standard errors reported in brackets; *** = significant at 1% level; ** = significant at 5% level; * = significant at 10%. All regressions are overall significant according to standard F-tests.

Pooling across years implicitly assumes that the parameters of equation (20) are constant across time. However, there are three potential sources of variation in the underlying system – the characteristics that define the reference country can change, those defining the reference industry can change, or industries can become more or less responsive to country and industry characteristics, so $\beta[j]$ changes. Given the increasing economic integration of the EU in the period 1980-1997 any or all of these are possible. To test for the validity of the assumption of constant coefficients, we include a full set of time dummies and time-dummy interactions to allow the reference country/industry characteristics or responsiveness to change over time. Testing for the stability of equation (20) then reduces to a joint test for the significance of all of the time dummy variables. Under heteroscedasticity, the standard F-test is not appropriate, but calculation of the appropriate White heteroscedastic consistent covariance matrix allows us to test for significance using a Wald test. The assumption of constant parameters across time involves imposing 57 restrictions, producing a Wald statistic of 2003, which is clearly significant (the Wald test is distributed Chi-squared with 57 degrees of freedom), leading to rejection of the hypothesis that parameters are constant. Given that the parameters vary over time in all three dimensions, we split the sample and estimate separately for each of four periods, 1980-83, 1985-8, 1990-93 and 1994-97. Separating the years also reduces the degree of endogeneity of some of the explanatory variables.[18] These estimates are given in remaining columns of Table 2.

From the first three rows, we see that the comparative advantage variables have the same signs as predicted by theory and that the coefficients are mostly significant and increasing in magnitude as we move to later time periods. By the final period, agriculture, skills, and R&D are significant at the 5%, 1% and 1% levels respectively. As with the pooled results, the coefficients for agriculture are consistently smaller than for skill and R&D intensity. Skill intensity is initially more important than R&D, although the two factors are equally important by the final period.

Results for the economic geography variables are more mixed. Backward linkages ($\beta[6]$) are always significant at the 5% level (indeed at the 1% level for the first three periods). Forward linkages $\beta[4]$ are significant at the 10% level, but only for the last period. Changes in the coefficients suggest that the backward linkage has become less strong through time, while the forward linkage has become stronger. This says that sectors highly intensive in intermediate goods are moving towards central locations to get better access to these goods, while access to industrial customers has decreased in importance. The transport cost coefficient, $\beta[5]$ is insignificant in all sub-periods suggesting that the interaction between market potential and transport costs has no significant impact on the location of indu

So far, we have concentrated on the direct estimates of the coefficients on the interaction terms. However, we can also calculate industry specific Rybczynski effects and country specific Heckscher-Ohlin effects. From equation (20) the Rybczynski effects are

$$R_k \equiv \frac{\partial \ln(r_i^k)}{\partial x_i[j]} = \beta[j]\left(y^k[j] - \bar{y}[j]\right) \quad for\ k = 1,...,33, \qquad j = 1,...,6. \tag{21}$$

while the Heckscher-Ohlin effects take the form

$$H_i \equiv \frac{\partial \ln(r_i^k)}{\partial y^k[j]} = \beta[j]\left(x_i[j] - \bar{x}[j]\right) \quad for\ i = 1,...,14, \qquad j = 1,...,6. \tag{22}$$

To calculate these we need, in addition to the interaction coefficients $\beta[j]$, the level effects $\beta[j]\bar{y}_j$ and $\beta[j]\bar{x}_j$, given in appendix table A2, and industry and country characteristics $y^k[j]$ and $x_i[j]$, from appendix A5.[20]

Table 3 reports R-effects for all industries for each of the six country characteristics. The numbers given in the table are the elasticity of each industry's output share with respect to the corresponding country characteristic (for 1994-1997). The size of the R-effect depends on the value of the industry characteristic relative to its estimated reference value, and on the strength of the relevant interaction as captured by $. Looking at, for example, skilled labour, we see positive R-effects for 26 of the 33 industries, with the largest positive effects occurring in Professional Instruments, followed by Drugs and Medicines, and Printing and Publishing. For the researchers and scientists, only three industries have positive R-effects – Aircraft, Drugs and Medicines, and Radio, TV and Communications equipment.[21] As a final example, we see that supplier access has a positive effect on all but one industry (Professional instruments), and the effects are especially strong for Iron and Steel and for Non-ferrous metals.

Note that we are calculating marginal R-effects with respect to changes in one country characteristic at a time. In some cases, when the reference intensity is outside the range of observed intensities, all of the marginal effects will be of one sign. Initially this would seem inconsistent with our use of a double-relative measure of production shares. However, it purely reflects the fact that when there is correlation across the country characteristics any given marginal R-effect can be more than offset by the R-effects from other characteristics. We see an example of this in Table 3 for relative market potential where all of the R-effects are negative. In such cases, the ranking and magnitude of the effects are still informative.

Table 3: R-effects (1994-97)

| | Agricultural endowment | Skilled labour endowment | Researchers and Scientists | Supplieraccess | Elasticity w.r.t. MP | Relative MP |
|---|---|---|---|---|---|---|
| Food | 4.571 | -0.718 | -3.350 | 4.806 | -0.866 | -4.282 |
| Beverages | 4.571 | 0.766 | -3.350 | 4.806 | -1.173 | -4.282 |
| Tobacco | 4.571 | -1.041 | -3.350 | 4.806 | -1.173 | -4.282 |
| Textiles | -0.838 | -0.027 | -3.410 | 2.847 | 0.126 | -2.770 |
| Wearing Apparel | -0.838 | -0.092 | -3.410 | 2.847 | 0.095 | -2.770 |
| Leather & Products | -0.838 | -0.602 | -3.410 | 2.847 | -0.366 | -2.770 |
| Footwear | -0.838 | -0.534 | -3.410 | 2.847 | -0.366 | -2.770 |
| Wood Products | 0.049 | 0.101 | -3.411 | 2.471 | 0.359 | -3.675 |
| Furniture & Fixtures | 0.049 | 0.420 | -3.411 | 2.471 | 0.359 | -3.675 |
| Paper & Products | -1.081 | 0.292 | -3.378 | 1.988 | -1.023 | -2.549 |
| Print & Publishing | -1.081 | 4.332 | -3.378 | 1.988 | -1.023 | -2.549 |
| Industrial Chemicals | -1.161 | 1.064 | -1.604 | 3.963 | 0.970 | -1.400 |
| Drugs & Medicine | -1.175 | 4.515 | 1.944 | 1.742 | 0.970 | -5.195 |
| Chemical Products nec | -1.161 | 1.990 | -1.604 | 3.963 | 0.970 | -1.400 |
| Rubber Products | -1.026 | 1.419 | -2.964 | 1.960 | 0.970 | -1.189 |
| Plastic Products | -1.026 | 0.535 | -2.964 | 1.960 | 0.970 | -1.189 |
| Pottery & China | -1.171 | 1.505 | -3.192 | 0.553 | 3.229 | -3.171 |
| Glass & Products | -1.171 | 0.646 | -3.192 | 0.553 | 3.229 | -3.171 |
| Non-Metallic minerals nec | -1.171 | 0.638 | -3.192 | 0.553 | 3.229 | -3.171 |
| Iron & Steel | -1.177 | 0.370 | -3.186 | 5.927 | 1.441 | -0.536 |
| Non-Ferrous Metals | -1.178 | -0.324 | -3.180 | 5.959 | -2.412 | -0.580 |
| Metal Proucts | -1.177 | 1.546 | -3.114 | 0.859 | -0.925 | -1.709 |
| Office & Computing | -1.177 | 3.944 | -0.298 | 4.096 | -1.900 | -3.918 |
| Machinery & Equipment | -1.175 | 2.137 | -2.367 | 1.628 | -1.900 | -2.828 |
| Radio,TV& Communication | -1.177 | 3.444 | 1.909 | 1.677 | -1.142 | -2.850 |
| Electrical Apparutus nec | -1.175 | 2.107 | -1.614 | 0.346 | -1.142 | -2.647 |
| Shipbuilding | -1.176 | 1.614 | -3.050 | 3.720 | -2.255 | -5.585 |
| Railroad Equipment | -1.178 | 2.858 | -2.009 | 2.865 | -2.255 | -6.821 |
| Motor Vehicles | -1.178 | 0.178 | -1.633 | 4.471 | -2.567 | -3.231 |
| Motorcycles | -1.178 | 0.202 | -2.009 | 2.865 | -2.255 | -6.821 |
| Aircraft | -1.178 | 4.210 | 2.531 | 3.052 | -2.255 | -2.308 |
| Transport Equipment | -1.178 | 0.819 | -2.009 | 2.865 | -2.255 | -6.821 |
| Professional Instruments | -1.173 | 4.907 | -2.042 | -1.362 | -1.142 | -4.391 |

The H-effects are reported in Table 4 for all countries for each of the industry characteristics. The numbers give the elasticity of a country's output share with respect to the industry intensities (for 1994-1997). For example, the H-effect for skill intensity tells us that if an industry becomes more intensive in the use of skilled labour, then Germany sees the largest increase in share, followed by Sweden, while Spain and Portugal loose shares. To take another example, if an industry's intermediate intensity increases, the production share of France, the UK and Germany in that industry will increase the most, while the production share of Ireland, Portugal and Greece will fall the most. .

Table 4: H-effects (1994-97)

| | Agricultural intensity | Skill intensity | R&D intensity | Intermediate intensity | Transport intensity | Share of output to industry |
|---|---|---|---|---|---|---|
| Austria | 0.146 | 2.763 | -0.956 | -0.641 | 0.863 | 0.022 |
| Belgium | 0.086 | 1.764 | -0.015 | -0.514 | 0.823 | 0.039 |
| Denmark | 0.543 | 3.179 | -0.411 | -1.219 | 0.901 | 0.043 |
| Spain | 0.284 | -0.782 | -2.939 | 0.155 | 0.675 | 0.005 |
| Finland | 0.366 | 2.606 | 1.074 | -1.380 | 0.933 | 0.027 |
| France | 0.239 | 1.923 | -0.194 | 1.064 | 0.231 | -0.090 |
| UK | 0.132 | 1.337 | -1.045 | 0.800 | 0.365 | 0.014 |
| Germany | 0.048 | 3.179 | -0.411 | 1.430 | -0.371 | 0.288 |
| Greece | 0.522 | 0.802 | -2.205 | -1.634 | 0.945 | 0.008 |
| Ireland | 0.487 | 0.987 | 1.496 | -1.929 | 0.948 | 0.038 |
| Italy | 0.270 | -0.012 | -2.805 | 0.868 | 0.399 | -0.057 |
| Netherlands | 0.291 | 2.154 | -1.494 | -0.272 | 0.773 | 0.020 |
| Portugal | 0.342 | -2.593 | -3.812 | -1.439 | 0.923 | 0.039 |
| Sweden | 0.192 | 2.862 | 1.228 | -0.736 | 0.873 | 0.015 |

*5.3 Robustness*

Before considering the overall fit and the relative importance of comparative advantage variables to economic geography variables, we briefly consider the robustness of our results. In estimating the coefficients in Table 2, our specification of the error structure allowed for the possibility of heteroscedasticity due to differences across industries or countries, but ignored the fact that we have an industry-country panel for each of the years. That is, we ignored the possibility that shocks might be correlated across industries and/or countries. There are two possible sources for such country/industry specific shocks. First, a particular industry or country might experience a shock to its share in European wide production. Looking back to equation (10) it is clear that our use of the double relative measure means that our specification is robust to such shocks. However, it is possible that country or industry characteristics might be consistently mismeasured for one particular country or industry. Again, from equation (10) it is clear that these measurement errors would translate in to fixed effects for the country or industry concerned. To test the robustness of our results to this form of specification error, we include a full set of country dummies and industry dummies and re-estimate equation (20), dropping the 12 country and industry levels variables. The results for the interaction variables for each of the years are reported in Table 5. They indicate that our results on the interaction terms are robust to the inclusion of industry and country fixed effects. The explanatory power of the equation is increased, as would be expected, with $R^2$ rising from around 17% to 24%, while the changes in the estimates of $[j] are negligible.

We also test the robustness of our specification by dropping each of the interactions in turn from the estimating equation. We undertake this just for the 1994-97 data set, and report only the interaction coefficients, $[j], in Table A3 in the Appendix. Once again, we see that the coefficients are stable across the specifications.

Table 5: $[j], Robustness Check I: Fixed effects: Dependent variable $\ln(r_{ij}^{k})$

| | Variable | 1980-83 | 1985-88 | 1990-93 | 1994-97 |
|---|---|---|---|---|---|
| $[1] | Agriculture endowment * agricultural intensity | 0.077 (0.126) | 0.135 (0.106) | 0.163* (0.087) | 0.153** (0.080) |
| $[2] | Skill endowment * skill intensity | 1.492*** (0.380) | 1.479*** (0.389) | 1.463*** (0.437) | 1.658*** (0.559) |
| $[3] | Researchers+scientists *R&D intensity | 0.588** (0.301) | 0.744** (0.376) | 1.112** (0.506) | 1.630*** (0.546) |
| $[4] | Supplier access * intermed. intensity | 0.564 (0.787) | 0.757 (0.753) | 0.801 (0.659) | 1.101* (0.684) |
| $[5] | Market pot. elasticity * transport costs | -0.405 (0.307) | -0.275 (0.291) | -0.323 (0.282) | -0.380 (0.267) |
| $[6] | Relative market pot. * output to industry | 0.187*** (0.059) | 0.176*** (0.056) | 0.130*** (0.047) | 0.084* (0.048) |
| | Country dummies | yes | yes | yes | yes |
| | Industry dummies | yes | yes | yes | yes |
| | Diagnostics | | | | |
| | $R^2$ | 0.233 | 0.235 | 0.249 | 0.237 |
| | Adjusted $R^2$ | 0.136 | 0.138 | 0.155 | 0.141 |
| | Number of obs | 456 | 456 | 456 | 456 |

*5.4 Comparative Advantage and Economic Geography: Explaining the Location of Production in the EU*

We have seen that both comparative advantage and economic geography variables are significant in determining the location of production in the EU. In this section, we consider how well the regression performs in explaining variation in production shares, as well as attempting to assess the relative importance of the two types of explanator

In terms of the overall regression, we are able to explain between 14 and 18 percentage of country specialization using just the six interaction variables. The proportion of variation in production shares that is explained through the model rises over time as Europe becomes increasingly specialized.[22] For comparison, note that Ellison and Glaeser (1999) are able to explain around 20 percentage of the location of US production using 16 interactions between characteristics of industries and of US states.

Similarly to Harrigan (1997) we construct predicted shares using the estimated coefficients and compare this to the actual shares by calculating the correlation between predicted and actual. In terms of the overall regression, the correlation rises from 0.375 in 1980-83 to 0.421 in 1994-97. We can also calculate the correlation for each country and each industry - the results are given in Tables 6 and 7 respectively for 1994-97.

Table 6: Correlation - Countries

| | | | | | | |
|---|---|---|---|---|---|---|
| Austria | Belgium | Denmark | Spain | Finland | France | UK |
| 0.125 | 0.267 | 0.358 | 0.480 | -0.022 | 0.246 | 0.414 |
| Germany | Greece | Ireland | Italy | Netherlands | Portugal | Sweden |
| 0.150 | 0.420 | 0.708 | 0.186 | 0.155 | 0.638 | 0.251 |

Table 7: Correlation - Industries

| | | | | | |
|---|---|---|---|---|---|
| Food | 0.089 | Industrial Chemicals | 0.134 | Office & Computing Machinery | 0.499 |
| Beverages | 0.397 | Drugs & Medicine | 0.164 | Machinery & Equipment | 0.005 |
| Tobacco | 0.340 | Chemical Products nec | -0.104 | Radio,TV & Communication | 0.543 |
| Textiles | 0.640 | Rubber Products | 0.621 | Electrical Apparatus nec | -0.110 |
| Wearing Apparel | 0.377 | Plastic Products | 0.519 | Shipbuilding & Repairing | 0.160 |
| Leather & Products | 0.783 | Pottery & China | 0.691 | Railroad Equipment | 0.752 |
| Footwear | 0.723 | Glass & Products | 0.411 | Motor Vehicles | 0.777 |
| Wood Products | -0.126 | Non-Metallic Minerals nec | 0.531 | Motorcycles | 0.498 |
| Furniture & Fixtures | 0.337 | Iron & Steel | 0.421 | Aircraft | 0.280 |
| Paper & Products | -0.114 | Non-Ferrous Metals | 0.444 | Transport Equipment nec | 0.576 |
| Printing & Publishing Publishing | 0.168 | Metal Products | 0.316 | Professional Instruments | 0.213 |

The results are somewhat mixed, both when it comes to countries and industries. With respect to countries, thirteen out of the fourteen countries have positive correlations, with only Finland showing a (tiny) negative correlation. Some countries appear to specialise more in line with their comparative advantage and economic geography than others. The most in line are the four peripheral European economies – Ireland, Portugal, Spain, Greece. Least in line are Finland, Austria and Germany.

In terms of industries 29 out of 33 show positive correlations - many of them much larger than the average correlation we saw for countries, suggesting that our specification does better at predicting distribution of industry shares across countries than on predicting the specialisation patterns of countries. Of the industries that do badly, two are resource based - wood products and paper products – and two are residual non-elsewhere classified.

It is clear from Table 2 that the comparative advantage variables are generally more significant than the economic geography variables. However, economic geography variables play a part – we can reject the null hypothesis that there are no economic geography effects at the 5% level. Taking the coefficients as given we conduct a further experiment. First, we set the geography variables to their reference value and calculate the correlation between predicted and actual shares. Second, we set the endowment variables to their reference value and calculate the correlation between predicted and actual shares. Setting a variable to its reference value assumes away the impact of that variable, but without altering the remaining estimated coefficients (something we want to avoid when both sets of variables are significant). The results show that disregarding the impact of economic geography variables reduces the correlation between predicted and actual shares by approximately 13%, while disregarding the impact of comparative advantage variables reduces the correlation between predicted and actual shares by approximately 44%, again confirming that economic geography does matte

## 7. *Concluding comments*

The theoretical model developed in this paper provides a rigorous framework in which comparative advantage can be combined with transport costs and geography, to provide a more general theory of trade and location. Results of the theory are intuitive, and enable Heckscher-Ohlin insights to be generalised to environments with more trade frictions than is common in such models. Linearization of the model provides an estimating equation in which country characteristics, industry characteristics, and most importantly the interaction of the two, combine to determine the shares of each industry in each country.

Implementing this equation on EU data, we find that a substantial part of the EU's cross-country variation in industrial structure can be explained by the forces captured in the model. Factor endowments are important. In particular, countries' endowments of highly skilled labour are important in attracting high skill intensive industries. Geography also matters, as industries dependent on forward and backward linkages locate close to centres of manufacturing supply and demand. Economic integration and falling levels of national government intervention in EU industry suggests that economic forces should have become increasingly important in determining industrial location, and we find some evidence that this is so.

Our approach is based on industries that are perfectly competitive, and the omission of imperfect competition is important. However, to include imperfect competition would create significant complexities that we have sought to avoid at this stage. For example, theory suggests that in such an environment, it is generally industries with intermediate levels of transport costs that are drawn into central locations, creating a non-monotonic relationship between transport intensity and location (this perhaps accounting for the poor performance of our transport intensity variable). General cases in which there are many industries, some perfectly and others imperfectly competitive, and all subject to transport costs have yet to be worked out. And we know that in such environments intermediate goods create a multiplicity of equilibria, as agglomerations may form. All of these issues are the subject of our ongoing research.

Appendix A1: Factor endowments, factor prices, and outputs.

We focus on a single country, so drop subscripts and write the output of industry $k$ as

$$x^k = A^k c(\boldsymbol{v}:k)^{-\eta}. \tag{23}$$

Comparing this with equation (3), we see that this is expressed in physical units not value (hence the different exponent on unit costs), and that a number of terms are combined in $A^k$, assumed constant. This means that differentiation is undertaken along a compensated demand curve, holding price indices constant. Suppose that there are just two factor inputs and no intermediates. Call the factor inputs $L$ and $K$ with factor prices $w$, $r$ and factor shares in sector $k$, $\gamma_w^k$ and $\gamma_r^k$. Considering the effects of factor price changes on outputs gives

$$\Delta x^k = -\eta\,(\gamma_w^k \Delta w + \gamma_r^k \Delta r) \tag{24}$$

Factor demand equations are

$$L^k = \frac{\partial c(w,r:k)}{\partial w} x^k, \qquad K^k = \frac{\partial c(w,r:k)}{\partial r} x^k \tag{25}$$

so the effect of a change in factor prices on factor demands in each industry are,

$$\begin{aligned} \Delta L^k &= \gamma_r^k \sigma^k (\Delta r - \Delta w) - \eta\,(\gamma_w^k \Delta w + \gamma_r^k \Delta r), \\ \Delta K^k &= \gamma_w^k \sigma^k (\Delta w - \Delta r) - \eta\,(\gamma_w^k \Delta w + \gamma_r^k \Delta r) \end{aligned} \tag{26}$$

where $\sigma^k$ is the elasticity of substitution between factors. These equations are for each sector, and their production-share-weighted sum must equal any change in factor endowments, so

$$\frac{wL}{Y}\Delta L = \sum\nolimits_k \gamma_w^k s^k \Delta L^k, \qquad \frac{rK}{Y}\Delta K = \sum\nolimits_k \gamma_r^k s^k \Delta K^k \tag{27}$$

Using (26) in (27) and applying Cramer's rule we can express changes in factor prices as a function of changes in endowments, $\Delta L$ and $\Delta K$. This relationship is the matrix $F$ of the text. In general, we can solve for factor prices as a function of endowments, and then use the result back in (24) for the associated changes in production le

General expressions are not very insightful, but if we assume that there are just two industries and that $\eta$ is very large relative to $\sigma^k$ (so $\sigma^k = 0$ in equations (26)) then

$$\begin{bmatrix} (\gamma_w^1)^2 s^1 + (\gamma_w^2)^2 s^2 & \gamma_w^1 \gamma_r^1 \mu^1 + \gamma_w^2 \gamma_r^2 s^2 \\ \gamma_w^1 \gamma_r^1 s^1 + \gamma_w^2 \gamma_r^2 s^2 & (\gamma_r^1)^2 s^1 + (\gamma_r^2)^2 s^2 \end{bmatrix} \cdot \begin{bmatrix} \Delta w \\ \Delta r \end{bmatrix} = \begin{bmatrix} -\dfrac{wL}{Y}\dfrac{\Delta L}{\eta} \\ -\dfrac{rK}{Y}\dfrac{\Delta K}{\eta} \end{bmatrix} \tag{28}$$

(Exponents are always written outside brackets, to distinguish them from superscripts). The determinant of this matrix is, $det = s^1 s^2 (\gamma_w^1 - \gamma_w^2)^2$. Now consider the effect of a change in capital endowment on factor prices. From (28) we derive

$$\frac{\Delta r}{\Delta K} = -\left(\frac{rK}{\eta Y}\right) \frac{(\gamma_w^1)^2 s^1 + (\gamma_w^2)^2 s^2}{s^1 s^2 (\gamma_w^2 - \gamma_w^1)^2}, \qquad \frac{\Delta w}{\Delta K} = \left(\frac{rK}{\eta Y}\right) \frac{(\gamma_w^1 \gamma_r^1 s^1 + \gamma_w^2 \gamma_r^2 s^2)}{s^1 s^2 (\gamma_w^2 - \gamma_w^1)^2} \tag{29}$$

These are two terms in the matrix $F$. Notice that they are inversely proportional to 0. Thus, as 0 6 4, so factor prices become invariant with respect to endowments, as expected. The terms are unambiguously signed, again as expected in a two-sector two-factor framework, although in higher dimension models this is not necessarily so.

Using (29) in (24) we can derive the effects of factor endowments on outputs. This is simply

$$\Delta x^1 = -\Delta K \frac{\gamma_r^1 \gamma_w^2}{\gamma_w^1 - \gamma_w^2} \tag{30}$$

which is exactly the Rybczynski effect of standard 2-by-2 Heckscher-Ohlin trade theory, expressed for proportional changes and value shares. This is then, a special case of the more general model of the paper.

## Appendix A2: The simulation model:

The model is constructed with 9 countries, 5 industries, 2 factors ($L$ and $K$) and Cobb-Douglas unit cost functions. The elasticity of substitution between varieties is set at 0 = 5, and in both figures 1 and 2 there is no production or use of the intermediate good. Consumers' expenditure is divided equally between the goods.

For figure 1, $t_{ij}$ =1.1 and $t_{ii}$ =1.0. All countries have the same endowment of $K$ ( = 1) and $L$ endowments in the range 0.75 - 1.25. Across industries, the share of $L$ in costs varies from 0.33 to 0.66, and the share of $K$ correspondingly from 0.66 to 0.33.

For figure 2, all countries have the same endowments $L = K = 1$ and all industries the same factor shares (0.5 for both factors). Transport costs vary across goods and countries, and the extreme values of transport costs are given in the table below.

| | Least transport intensive good | Most transport intensive good |
|---|---|---|
| 2 closest economies | 1.003 | 1.03 |
| 2 furthest economies | 1.045 | 1.49 |

The horizontal axes measure the transport costs between the two closest economies for different industries, $2^k$, and the market potential of different countries, computed from equation (9) for the middle ranked industry.

Appendix A3: Construction of variables

1) Dependent variable: $\Delta r_i^k$: log of industry output levels, expressed relative to both the EU output of industry $k$ as a whole, and to the total manufacturing output of country $i$. This value is calculated from the production data for each of the 36 sectors (see Appendix A4).

2) Primary factors: We use three factor intensity/ factor endowment interactions, for skilled labour, researchers and scientists, and agriculture

A) Share of factors in costs of each industry, $(_w$ :

i) Skilled labour intensity: proxied by the proportion of non-manual workers in the sector's employment times labour compensation as % gross output.

ii) R&D intensity: R&D expenditure as % gross output. This includes some non-labour components, although the major share of R&D expenditure is personnel costs.

iii) Agricultural intensity: Inputs from agriculture, fishery and forestry as % gross output.

B) Endowments, $)L_i$:

i) Skilled labour: proportion of the population with secondary education or higher (logs).

ii) Researchers and scientists per ten thousand labour force (logs).

iii) Agricultural abundance: proxied by gross value added of agriculture, forestry and fishery products as % of all branches (logs).

3) Intermediate supply: Forward linkages

A) Share of intermediate in costs of each industry, $(_q$: from input-output tables.

B) Price of the intermediate good, $)q_i$: The intermediate price in each country is given by equation (6), and is a function of price indices $G_i^k$, as defined in equation (2). Using (3) with equation (2) gives:

$$\left(G_i^k\right)^{1-\eta} = \sum_j \left[ \frac{z_j^k \left(t_{ji}^k\right)^{1-\eta}}{\sum_\ell e_\ell^k \left(t_{j\ell}^k / G_\ell^k\right)^{1-\eta}} \right]. \tag{31}$$

We assume that variation in the term in square brackets comes mainly from the numerator. Holding the denominator constant (and equal to $1/A$), using (31) in (6) and taking logs gives,

$$\log(q_i) = A\sum_k \frac{\phi^k}{1-\eta}\log\left[\sum_j z_j^k \left(t_{ji}^k\right)^{1-\eta}\right] = A\sum_k \frac{\phi^k}{1-\eta}\log\left[\sum_j z_j^k \left(d_{ji}^{\theta^k}\right)^{1-\eta}\right] \qquad (32)$$

The term in square brackets gives, for each country and industry, a distance weighted measure of proximity to production in the industry. The $\phi^k$ weighted average of these gives each country's proximity to suppliers of the product mix that goes into the composite intermediate, and is an overall measure of the 'supplier access' of country $i$. Implementation of equation (32) requires:

i) Production levels : $z_j^k$; value of output data, see A4.

ii) Shares of each industry in intermediate, $\phi^k$: Sales to aggregate manufacturing industry as share of gross output for each sector, from input-output data.

iii) Distance, $d_{ij}$: Distance between the economic centre of gravity of countries. Centres of gravity computed from subnational GDP (NUTS2). 'Internal distance', $d_{ii} = 1$.

iv) Elasticity with respect to distance: $\theta^k(1-\eta) = -1$. This value chosen in line with estimates from gravity models of trade and from the geographical tradition of market potential, and assumed the same in a

4) Market potential and transport costs

A) Transport intensities, $\theta^k$: Transport costs as percentage of fob priced sales, see A4.

B) Elasticity of market potential, $\mu_\theta(\bar{u}:i)$: Equation (18) of the text is,

$$m(\boldsymbol{u}^k : i) \equiv \sum_j \left(d_{ij}^{\theta^k}\right)^{1-\eta} e_j^k \left(G_j^k\right)^{\eta-1}. \qquad (33)$$

We find the elasticity of this with respect to $\theta$ by evaluating it at two distinct values of $\theta$, $\tilde{\theta}$ and $\tilde{\theta} + \Delta\theta$, holding other terms at their reference industry value, $e_j^k\left(G_j^k\right)^{\eta-1} = \bar{e}_j\left(\bar{G}_j\right)^{\eta-1}$:

$$\mu_\theta(\bar{u}:i) = \left(\frac{\sum_j \bar{e}_j\left(\bar{G}_j\right)^{\eta-1}\left(d_{ij}^{-(1-\eta)(\tilde{\theta}+\Delta\theta)} - d_{ij}^{-(1-\eta)\tilde{\theta}}\right)}{\sum_j \bar{e}_j\left(\bar{G}_j\right)^{\eta-1} d_{ij}^{-(1-\eta)\tilde{\theta}}}\right)\frac{\tilde{\theta}}{\Delta\theta} \qquad (34)$$

To evaluate this we need:

i) Reference industry expenditures, $\bar{e}_j\left(\bar{G}_j\right)^{\eta-1}$; proportional to $GDP_j$.

ii) Distance, reference transport intensity, as above, i.e., $\tilde{\theta}(1-\eta) = -1$.

iii) Evaluated at: $\theta = 0.7$ and $\Delta\theta = 0.6$.

5) Intermediate goods demand: Backwards linkages

A) Share of industry's sales going to manufacturing, denoted $\psi^k$: from input-output tables.

B) Elasticity of market potential, $\mu_{\psi}(\bar{u}:i)$

From equation (7), expenditure consists of two components, final and derived demand. Using (7) in (9), and denoting the quantity of the intermediate used in location $j$ by $h_j$ gives,

$$m(\boldsymbol{u^k}:i) = \sum_j \left( \alpha^k y_j + \phi^k q_j h_j \right) \left( d_{ij}^{\theta^k} / G_j^k \right)^{1-\eta} \tag{35}$$

If the share of industry $k$'s output going to intermediate sales is $\psi^k$, defined by

$$\psi^k \equiv \phi^k \sum_j q_j h_j / \sum_j e_j^k, \quad and \quad 1 - \psi^k = \alpha^k \sum_j y_j / \sum_j e_j^k \tag{36}$$

then using (36) in (35) gives,

$$m(\boldsymbol{u^k}:i) = \sum_j \left( (1-\psi^k) y_j / \sum_j y_j + \psi^k q_j h_j / \sum_j q_j h_j \right) \left( d_{ij}^{\theta^k} / G_j^k \right)^{1-\eta} \sum_j e_j^k \tag{37}$$

The elasticity of this with respect to $\psi^k$ is

$$\mu_{\psi}(\bar{u}:i) = \sum_j \left( d_{ij}^{\bar{\theta}} / \bar{G}_j \right)^{1-\eta} \left( \frac{q_j h_j}{\sum_j q_j h_j} - \frac{y_j}{\sum_j y_j} \right) \frac{m(\bar{u}:i)}{\bar{\psi}}, \tag{38}$$

which we can compute by constructing separate market potential measures for final expenditure and for intermediate demands. Implementation of equation (38) requires:

i) Distance, reference transport intensity, as above, i.e., $\tilde{\theta}(1-\eta) = -1$.

ii) Spatial distribution of final expenditure, $y_i / \sum_i y_i$: use GDP$_i$.

iii) Spatial distribution of intermediate expenditures, $q_j h_j / \sum_j q_j h_j$. From input-output tables.

## Appendix A4: Data sources

**Manufacturing production:** The data set is based on production data from two sources: the OECD STAN database and the UNIDO database. The OECD STAN database provides production data for 13 EU countries and 36 industries, from 1980 to 1997. We combine this with production data for Ireland from the UN UNIDO database, giving us data on 14 EU countries (the EU 15, excluding Luxembourg). Due to missing observations, a small number of data points had to be estimated (see Midelfart-Knarvik, Overman, Redding and Venables, 2000, for details on missing data and estimation procedures).

OECD STAN (Structural Analysis) database

Data: National industrial data on value of output.

Period: 1970-1997, annual data.

Countries: 13 European countries: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Italy, Netherlands, Portugal, Spain, Sweden, United Kingdom.

Sectors: 36 industrial sectors, as per Table A1.

UNIDO database

Data: National industrial data on value of output.

Period: 1970-1997, annual data.

Countries: Ireland.

Sectors: 27 industrial sectors; specification adjusted to be consistent with STAN database.

Country and industry characteristics

(A) Industry characteristics

- *R&D as percentage of total costs*: R&D expenditures as share of gross value of output*: Source: ANBERD and STAN, OECD
- *Skill intensity*: Source: STAN, OECD, and COMPET, Eurostat
- *Transport costs (intensity)*; Transport costs as share of fob priced sales within the EU (i.e. basis for calculation is intra-EU trade). Source: The GTAP 4 Data Base (McDougal et al, 1998).
- *Agricultural input share:* Use of agricultural inputs (incl. fishery and forestry) as share of gross value of output**: Source: Input-output tables, OECD
- *Forward linkage*: Total use of intermediates as a share of gross value of output**. Source: Input-output tables, OECD
- *Backward linkage*: Sales to manufacturing as share of total sales: Percentage of domestic sales to domestic manufacturing as intermediates**. Source: Input-output tables, OECD

(B) Country characteristics: 1980, 1985, 1990, 1997

- *Market potentials*: Indicators of economic potential (see Appendix A4).
  Source: Regio database, Eurostat
- *Researcher and Scientists*: Researchers per 10,000 labour force
  source: OECD Science, Technology and Industry Scoreboard 1999
- *Education of population*: Share of population aged 25-59 with at least secondary education. Source: Eurostat Yearbook (levels for 1996-7), and Barro and Lee (1993) (for growth rates used to calculate other year values).
- *Agricultural production*: Gross value added of agriculture, forestry and fishery products as % of all branches.
  Source: Eurostat

Notes:

*) As industry intensities are assumed to be equal across countries R&D shares of gross value of output are calculated as weighted averages. We use data for Denmark, Finland, France, Germany (former FRG), Italy, Netherlands, Spain, Sweden and the UK for the year 1990.

**) We use a weighted average of 1990 IO tables for Denmark, France, Germany and the UK to calculate intermediate input shares and the destination of final output (intermediate usage vs final usage). Intermediates include both domestically purchased and imported inputs. The data needed to calculate the industry intensities were in general not available for the 36 sectors disaggregation, so intensities calculated at a cruder level of disaggregation, were mapped into the 36 sectors classification.

Appendix A5: Tables

Table A1: Industry characteristics

| Industries | ISIC | Share of non-man workers in work-force | Labour compensation, share of costs | R&D expend, share of costs | Use of intermed share of costs | Input of agric, fish & forestry share of costs | Transport costs, share of fob value shipped | Sales to manuf, share of ouput |
|---|---|---|---|---|---|---|---|---|
| Food | 3110 | 0.336 | 0.116 | 0.0011 | 0.708 | 0.2664 | 0.044 | 0.175 |
| Beverages | 3130 | 0.48 | 0.167 | 0.0011 | 0.708 | 0.2664 | 0.041 | 0.175 |
| Tobacco | 3140 | 0.351 | 0.085 | 0.0011 | 0.708 | 0.2664 | 0.041 | 0.175 |
| Textiles | 3210 | 0.248 | 0.234 | 0.0006 | 0.643 | 0.0158 | 0.056 | 0.341 |
| Wearing Apparel | 3220 | 0.207 | 0.272 | 0.0006 | 0.643 | 0.0158 | 0.055 | 0.341 |
| Leather & Products | 3230 | 0.21 | 0.201 | 0.0006 | 0.643 | 0.0158 | 0.050 | 0.341 |
| Footwear | 3240 | 0.155 | 0.285 | 0.0006 | 0.643 | 0.0158 | 0.050 | 0.341 |
| Wood Products | 3310 | 0.266 | 0.231 | 0.0005 | 0.630 | 0.0569 | 0.059 | 0.229 |
| Furniture & Fixtures | 3320 | 0.258 | 0.272 | 0.0005 | 0.630 | 0.0569 | 0.059 | 0.229 |
| Paper & Products | 3410 | 0.349 | 0.192 | 0.0009 | 0.614 | 0.0045 | 0.043 | 0.376 |
| Printing & Publishing Publishing | 3420 | 0.539 | 0.331 | 0.0009 | 0.614 | 0.0045 | 0.043 | 0.376 |
| Industrial Chemicals | 3510 | 0.542 | 0.163 | 0.0165 | 0.680 | 0.0008 | 0.068 | 0.625 |
| Drugs & Medicine | 3522 | 0.714 | 0.257 | 0.0476 | 0.606 | 0.0002 | 0.068 | 0.117 |
| Chemical Products nec | 3528 | 0.542 | 0.210 | 0.0165 | 0.680 | 0.0008 | 0.068 | 0.625 |
| Rubber Products | 3550 | 0.294 | 0.333 | 0.0045 | 0.614 | 0.0071 | 0.068 | 0.686 |
| Plastic Products | 3560 | 0.297 | 0.248 | 0.0045 | 0.614 | 0.0071 | 0.068 | 0.686 |
| Pottery & China | 3610 | 0.318 | 0.316 | 0.0025 | 0.567 | 0.0003 | 0.114 | 0.286 |
| Glass & Products | 3620 | 0.255 | 0.300 | 0.0025 | 0.567 | 0.0003 | 0.114 | 0.286 |
| Non-Metallic Minerals nec | 3690 | 0.318 | 0.240 | 0.0025 | 0.567 | 0.0003 | 0.114 | 0.286 |
| Iron & Steel | 3710 | 0.32 | 0.215 | 0.0026 | 0.745 | 0.0001 | 0.076 | 0.915 |
| Non-Ferrous Metals | 3720 | 0.32 | 0.155 | 0.0026 | 0.746 | 0.0000 | 0.031 | 0.898 |
| Metal Products | 3810 | 0.282 | 0.360 | 0.0032 | 0.577 | 0.0001 | 0.044 | 0.545 |
| Office & Computing Machinery | 3825 | 0.665 | 0.252 | 0.0279 | 0.684 | 0.0001 | 0.035 | 0.206 |
| Machinery & Equipment | 3829 | 0.421 | 0.280 | 0.0098 | 0.603 | 0.0002 | 0.035 | 0.333 |
| Radio,TV & Communication | 3832 | 0.512 | 0.301 | 0.0474 | 0.604 | 0.0001 | 0.042 | 0.330 |
| Electrical Apparatus nec | 3839 | 0.373 | 0.314 | 0.0164 | 0.560 | 0.0001 | 0.042 | 0.361 |
| Shipbuilding & Repairing | 3841 | 0.280 | 0.369 | 0.0037 | 0.672 | 0.0001 | 0.032 | 0.099 |
| Railroad Equipment | 3842 | 0.294 | 0.468 | 0.0129 | 0.643 | 0.0000 | 0.032 | 0.057 |
| Motor Vehicles | 3843 | 0.265 | 0.24 | 0.0162 | 0.697 | 0.0000 | 0.030 | 0.279 |
| Motorcycles | 3844 | 0.253 | 0.255 | 0.0129 | 0.643 | 0.0000 | 0.032 | 0.057 |
| Aircraft | 3845 | 0.547 | 0.32 | 0.0529 | 0.650 | 0.0000 | 0.032 | 0.419 |
| Transport Equipment nec | 3849 | 0.318 | 0.256 | 0.0129 | 0.643 | 0.0000 | 0.032 | 0.057 |
| Professional Instruments | 3850 | 0.439 | 0.443 | 0.0126 | 0.504 | 0.0003 | 0.042 | 0.167 |

Table A2: Regression results: Dependent variable $\ln(r_{ij}^{k})$

| Variable | Pooled | 1980-83 | 1985-88 | 1990-93 | 1994-97 |
|---|---|---|---|---|---|
| CONSTANT, > | 7.753*** (2.471) | 6.798 (5.257) | 7.647 (5.264) | 5.985 (4.979) | 12.820** (5.637) |
| Country Characteristic: - $\$[j]\bar{y}[j]$ | | | | | |
| Agricultural endowment | -0.027 (0.044) | -0.017 (0.097) | -0.066 (0.098) | 0.048 (0.098) | -0.060 (0.092) |
| Skilled labour endowment | -0.306*** (0.069) | -0.404*** (0.151) | -0.315** (0.144) | -0.216 (0.141) | -0.180 (0.133) |
| Researchers and Scientists | -0.265*** (0.060) | -0.161 (0.166) | -0.273* (0.147) | -0.285** (0.116) | -0.318*** (0.111) |
| Supplier access | -0.379 (0.303) | -0.255 (0.690) | -0.379 (0.655) | -0.192 (0.576) | -0.772 (0.622) |
| Market potential transport cost elas. | -0.374** (0.152) | -0.402 (0.325) | -0.277 (0.324) | -0.493 (0.349) | -0.365 (0.347) |
| Relative market potential | 0.065** (0.031) | 0.138** (0.070) | 0.079 (0.071) | 0.107 (0.080) | -0.008 (0.089) |
| Industry Characteristic: - $\$[j]\bar{x}[j]$ | | | | | |
| Agricultural intensity | 0.007 (0.042) | -0.023 (0.107) | -0.039 (0.091) | -0.031 (0.081) | 0.026 (0.069) |
| Skill intensity | -1.471*** (0.225) | -1.428*** (0.427) | -1.363*** (0.405) | -1.351*** (0.460) | -1.507*** (0.579) |
| R&D intensity | -0.709*** (0.197) | -0.708** (0.324) | -0.870** (0.397) | -1.212** (0.558) | -1.697*** (0.571) |
| Intermediate intensity | -0.421** (0.208) | -0.303 (0.461) | -0.404 (0.449) | -0.461 (0.405) | -0.652 (0.429) |
| Transport costs | 0.116*** (0.034) | 0.124* (0.073) | 0.098 (0.070) | 0.108* (0.067) | 0.127** (0.066) |
| Share of output to industry | -0.035 (0.029) | -0.074 (0.052) | -0.062 (0.054) | -0.033 (0.063) | 0.015 (0.058) |
| Interactions: $\$[j]$ | | | | | |
| Agric. endowment * agricultural inputs | 0.111** (0.046) | 0.078 (0.114) | 0.140 (0.097) | 0.166** (0.085) | 0.158** (0.079) |
| Skill endowment * skill intensity | 1.600*** (0.228) | 1.503*** (0.439) | 1.484*** (0.420) | 1.479*** (0.463) | 1.663*** (0.582) |
| Researchers+scientists * R&D intensity | 0.602*** (0.196) | 0.584* (0.325) | 0.741** (0.389) | 1.108** (0.536) | 1.624*** (0.581) |
| Supplier access * intermed. intensity | 0.763** (0.356) | 0.570 (0.811) | 0.754 (0.771) | 0.799 (0.667) | 1.096* (0.689) |
| Market pot. elasticity * transport costs | -0.356** (0.148) | -0.395 (0.315) | -0.270 (0.299) | -0.319 (0.290) | -0.382 (0.275) |
| Relative market pot. * output to industry | 0.138*** (0.024) | 0.182*** (0.059) | 0.171*** (0.052) | 0.130*** (0.043) | 0.083** (0.041) |
| Diagnostics | | | | | |
| $R^2$ | 0.145 | 0.140 | 0.151 | 0.177 | 0.171 |
| Adjusted $R^2$ | 0.136 | 0.105 | 0.116 | 0.143 | 0.137 |
| Number of obs | 1824 | 456 | 456 | 456 | 456 |

Note: Standard errors reported in brackets; *** = significant at 1% level; ** = significant at 5% level; * = significant at 10%. All regressions are overall significant according to standard F-tests.

Table A3: $\beta$[j], Robustness Check II, 1994-97: Dependent variable $\ln(r_{ij}^{k})$

| Variable | | -1 | -2 | -3 | -4 | -5 | -6 |
|---|---|---|---|---|---|---|---|
| Agriculture endowment | 0.158** | | 0.174** | 0.159** | 0.121* | 0.160** | 0.163** |
| * agricultural intensity | (0.079) | | (0.081) | (0.083) | (0.071) | (0.079) | (0.079) |
| Skill endowment | 1.663*** | 1.732*** | | 2.554*** | 1.601*** | 1.704*** | 1.655*** |
| * skill intensity | (0.582) | (0.575) | | (0.485) | (0.587) | (0.595) | (0.583) |
| Researchers+scientists | 1.624*** | 1.627*** | 2.394*** | | 1.670*** | 1.597*** | 1.617*** |
| *R&D intensity | (0.581) | (0.590) | (0.488) | | (0.590) | (0.594) | (0.580) |
| Supplier access | 1.096* | 0.790 | 0.978 | 1.193* | | 1.300* | 1.147* |
| * intermed. intensity | (0.689) | (0.674) | (0.683) | (0.699) | | (0.671) | (0.686) |
| Market pot. elasticity | -0.382 | -0.401 | -0.450* | -0.329 | -0.563** | | -0.313 |
| * transport costs | (0.275) | (0.277) | (0.263) | (0.281) | (0.274) | | (0.267) |
| Relative market pot. | 0.083** | 0.097** | 0.078** | 0.081* | 0.096** | 0.066* | |
| * output to industry | (0.041) | (0.039) | (0.035) | (0.044) | (0.042) | (0.039) | |
| Diagnostics | | | | | | | |
| $R^2$ | 0.160 | 0.149 | 0.148 | 0.141 | 0.165 | 0.164 | 0.168 |
| Adjusted $R^2$ | 0.125 | 0.120 | 0.119 | 0.111 | 0.137 | 0.136 | 0.140 |
| Number of obs | 456 | 456 | 456 | 456 | 456 | 456 | 456 |

## *Endnotes:*

1. Unlike recent work by e.g. Trefler (1993, 1995) the aim of this paper is not to provide a test of the HOV theorem, but to estimate how factor endowments, trade frictions and the geographical distribution of demand interact in determining the location of production and international specialization.

2. See Midelfart-Knarvik, Overman, Redding and Venables, (2000).

3. See Leamer and Levinsohn (1995) for a discussion and critique of this and other approaches.

4. See Fujita, Krugman and Venables (1999).

5. Letting this elasticity differ across industries would be straightforward in the theoretical sections, but a common value is assumed in the empirical estimation.

6. Having many intermediate goods and a full input-output structure would be easy in theory, but is difficult to implement in the econometrics. The reason is that diagonal elements often dominate the input-output matrix, so that examination of forward and backward linkages encounters severe endogeneity problems.

7. If there are no transport costs (all $t_{ij}^k = 1$) then price indices and market potential take the same value in all locations, so production is determined by cost factors alone; otherwise, geography matters.

8. These properties both hold by appropriate choice of units.

9. Notice that, at the reference point, there is no cross-industry variation in costs (since $c(\bar{\boldsymbol{v}} : \boldsymbol{k}) = 1$ for all $k$) or cross-country variation in market potential (since $m(\bar{\boldsymbol{u}} : i) = 1$ for all $i$); the linearisation captures the variation in costs and industry characteristics away from the reference point.

10. We can in principle estimate with the full matrix $F$, not just the diagonal, but the resulting specification is beset by multi-collinearity problems.

11. Endowments of other factors are scaled back equi-proportionately to maintain country size, as are the input shares of these factors

12. Local perturbation of the endowment in this direction has no effect on $r_i^k$.

13. Since we are focussing only on the structure of manufacturing, we take agricultural production as an exogenous measure of 'agriculture abundance', rather than going back to an underlying endowment such as land.

14. The figure only illustrates the range in which the saddle shape holds. Increasing transport intensity further causes a flattening of the surface. As $2^k$ 6 4 so the market potential of industry *k* becomes equal to local demand

15. The resulting data has time varying country characteristics, while industry intensities are constant across all four time periods. Thus, we are assuming that there have not been major changes in production technologies given the level of aggregation of our industrial classification.

16. The results on the significance of coefficients are largely unchanged when we move from OLS standard errors to heteroscedastic robust standard errors. This reflects the fact that heteroscedasticity is unlikely to be a major problem as we measuring shares (the left hand side variable) relative to industry and country size.

17. This gives us a total of 1824 observations. In each of the four time periods, there are 6 missing observations: Denmark: ISIC 3842, 3845, 3849; France: ISIC 3849; Ireland: ISIC 3130; Netherlands: ISIC 3842.

18. Midelfart-Knarvik, Overman, Redding and Venables (2000) show that the industrial production structure of Europe changes over this time period. Assuming that this is in response to EU integration and in line with our model, then pooling across years is problematic, as period (t+1)'s explanatory variables are a function of period t's production structure. The lack of appropriate instruments, and the short length of the panel then rules out GMM estimation of a suitably specified panel

19. We have experimented with alternative definitions of transport intensity. Results reported use data from the GTAP 4 Database, which provide transport costs as a percentage of fob priced sales (see Appendix A3). Measures based on tradability (defined as the ratio of the sum of exports and imports to gross value of output), and measures based on Hummels(1998) had no impact on sign or significance of the results.

20. Table A2 gives estimates of $\beta[j]\bar{y}[j]$ and $\beta[j]\bar{x}[j]$ Dividing by estimates of $\beta[j]$ gives estimates of the reference points $\bar{y}[j]$ and $\bar{x}[j]$, that determine the location of the saddle. For around 80% of our estimates, these lie with within the range of observations on the corresponding variables, $x_i[j]$ and $y^k[j]$, and none are significantly outside. If our sample of industries covered the entire economy (services as well as manufactures), then lying within the range would be required by theory, as it would ensure that industry output responses to a change in country characteristics included both positive and negative responses.

21. The preponderance of negative values reflects the use of Scientists in non-manufacturing sectors of the economy. See Harrigan (1997) for a similar finding.

22. See Midelfart-Knarvik, Overman, Redding and Venables, (2000).

*References*

Baldwin, R.E. (1971), 'Determinants of the commodity structure of US trade', *American Economic Review*, 126-146.

Barro, Robert and Jong-Wha Lee (1993): "International Comparisons of Educational Attainment", NBER Paper no. 4349

Davis, D. and D. Weinstein (1998): "Market access, economic geography and comparative advantage: an e assessment", NBER Working Paper no. 6787

Davis, D. and D. Weinstein (1999): "Economic geography and regional production structure: an empirical investigation", *European Economic Review* 43: 379-408

Ellison, Glenn and Edward L. Glaeser (1999): "The geographic concentration of industry: Does natural ad explain agglomeration?", *American Economic Review* 89, Papers and Proceedings: 311-316

Fujita, M., P. Krugman and A.J. Venables (1999) *The spatial economy; cities, regions and international trade*, MIT press, Cambridge MA.

Harrigan, J. (1995) 'Factor endowments and the international location of production; econometric eviden OECD', *Journal of International Economics*,39,123-141.

Harrigan, J. (1997) 'Technology, factor supplies and international specialization; estimating the neoclassic *American Economic Review*, 87, 475-494.

Hummels, David (1998), "Towards a geography of transport costs", mimeo University of Chicago.

Leamer, E. (1984), *Sources of International Comparative Advantage*, MIT press, Cambridge MA.

Leamer, E. (1987): "Paths Of Development in the 3 x n General Equilibrium Model", *Journal of Political Economy* 95: 961-999

Leamer, E. and J. Levinsohn (1995), 'International trade theory; the evidence', in G. Grossman and K. Ro *Handbook of International Economics*, vol. 3, North Holland, Amsterdam.

McDougall, R., A. Elbehri, and T. P. Truong (1998) (eds): *Global trade, assistance, protection. The GTAP 4 Data Base*, Center for Global Trade Analysis, Purdue University.

Midelfart-Knarvik, K-H, H.G. Overman, S.J. Redding and A.J. Venables (2000): "*The location of European industry*", report prepared for the Directorate General for Economic and Financial Affairs, Europea Commission, Economic papers No. 142. April 2000, European Commission, Brussels.

Trefler, D. (1993): "International factor price differences: Leontief was right!", *Journal of Political Economy* 961-987

Trefler, D. (1995): "The case of the missing trade and other mysteries", *American Economic Review*, 85, 102 1046.