

Oct 2nd 2000:

Comparative advantage and economic geography: estimating the location of production 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 Gianmarco Ottaviano, Diego Puga, Steve Redding, Paul Seabright, and seminar participants at IUI (Stockholm), LSE, and ERWIT (Copenhagen).

Authors' addresses:

K-H Midelfart-Knarvik
SIOS
Norges Handelhoyskole
Helleveien 30
5035 Bergen-Sandviken
Norway

H.G Overman
Dept of Geography
LSE
Houghton Street
London WC2A 2AE

A.J. Venables
Dept of Economics
LSE
Houghton Street
London WC2A 2AE

KarenHelene.Knarvik@snf.no

h.g.overman@lse.ac.uk

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 both comparative advantage and geographical forces. Our model contains countries that have differing factor endowments, and that have transport costs on trade between them. Industrial sectors use primary factors and intermediate goods to produce differentiated goods, differentiation ensuring that even in the presence of transport costs there are positive trade flows. The equilibrium pattern of industrial location is the outcome of both cost factors and the geographical distribution of demand. 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 will shape their industrial structures. Both the prices and the demand for intermediate goods 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

estimate.

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.¹ 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.² 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 does mean 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 further issues which go beyond the scope of this paper. For example, in such an environment there is, in general, a multiplicity of equilibria, and hence no unique mapping from underlying characteristics of countries and industries to industrial location³. Addressing these issues will be the subject of future research.

The paper proceeds as follows. Section 2 outlines our analytical framework, and section 3 illustrates the way in which country and industry characteristics interact to determine location. Section 4 derives the estimating equation, and section 5 presents econometric results. Section 6 concludes.

2. The model

The model has the following structure. There are I countries, K industrial sectors and M primary factors. All industries are perfectly competitive and operate under constant returns to scale using primary factors and intermediate goods. 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 factor on shipping industry k products from country i to country j .

With this structure, the value of production of each industry in each country (denoted z_i^k) 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 (even though there are 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 all symmetrical, i.e. face the same cost and demand functions. Input prices in country i are denoted by the vector \mathbf{v}_i , and the costs of industry k in

country i are given by unit cost function $c(\mathbf{v}_i : k)$. F.o.b. prices equal unit costs, so

$$p_i^k = c(\mathbf{v}_i : k). \quad (1)$$

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 $c(\mathbf{v}_i : k)t_{ij}^k$

2.2: Demand

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

$$G_j^k = \left[\sum_i n_i^k \left(c(\mathbf{v}_i : k) t_{ij}^k \right)^{1-\eta} \right]^{1/(1-\eta)} \quad (2)$$

where η is the elasticity of substitution between product varieties, assumed to be the same in all industries.⁴ The value of demand for a single variety produced in i and sold in j is then $\left(c(\mathbf{v}_i : k) t_{ij}^k \right)^{1-\eta} e_j^k \left(G_j^k \right)^{\eta-1}$, as usual from a CES demand system. Summing this over all markets, j , and all n_i^k industry k varieties produced by country i , gives the following expression for the value of production of industry k in country i ;

$$z_i^k = n_i^k c(\mathbf{v}_i : k)^{1-\eta} \sum_j \left(t_{ij}^k \right)^{1-\eta} e_j^k \left(G_j^k \right)^{\eta-1}. \quad (3)$$

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 ε_i^k , i.e. we assume that

$$n_i^k = s_i s^k \exp[\varepsilon_i^k]. \quad (4)$$

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).

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 form

$$r_i^k \equiv z_i^k/s_i^k = c(\mathbf{v}_i:k)^{1-\eta} \sum_j (t_{ij}^k)^{1-\eta} e_j^k (G_j^k)^{\eta-1} \exp[\varepsilon_i^k]. \quad (5)$$

This equation says that systematic cross-country variation in sectors' output, (measured by r_i^k) 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 sum in (5), which we will refer to as the market potential of industry k in country i . 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.

2.3: Input prices

Inputs consist of both primary factors and a single composite intermediate good,⁵ with prices \mathbf{w}_i and q_i respectively, so $\mathbf{v}_i = [\mathbf{w}_i, q_i]$. Prices of the primary factors, \mathbf{w}_i , are determined by market clearing, which can be expressed as,

$$\mathbf{L}_i = \sum_k x_i^k \mathbf{c}_w(\mathbf{w}_i, q_i:k) \quad (6)$$

where \mathbf{L}_i is the endowment vector of country i , $\mathbf{c}_w(\mathbf{w}_i, q_i:k)$ 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 of which has price G_i^k . The price of the intermediate good in country i is then

$$q_i = \prod_k (G_i^k)^{\varphi^k}, \quad \sum_k \varphi^k = 1, \quad (7)$$

where φ^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 we assume are fixed shares, α^k , of income, f_i , in each country. The latter is the value of total intermediate demand in country i , $q_i y_i$, times the share attributable to sector k , φ^k . Demand for the intermediate, y_i , is derived from industry output levels times partial derivatives of unit cost functions with respect to the intermediate price, giving:

$$e_i^k = \alpha^k f_i + \varphi^k q_i y_i \quad (8)$$

where $y_i = \sum_k x_i^k c_q(\mathbf{w}_i, q_i; k)$.

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

3. Properties of the model

The model captures the effects of factor endowments, geography, and industrial linkages on the location of production. In section 4 the model is linearised, this providing both the local comparative statics and our estimating equation. In this section we use numerical simulation to draw out the main relationships embodied in the model, using parameter values given in appendix A1 and a deterministic structure with $\varepsilon_i^k = 0$.

3.1 Factor endowments.

The first experiment is to suppose that all countries are identical, except in their relative endowment

of a single factor, ℓ_i , and that all industries are identical, except in the share of this factor in costs, which we denote γ^k . (Endowments of other factors are scaled back equi-proportionately to maintain country size, as are the input shares of these factors). Figure 1 plots output levels, $\log(r_i^k)$, as a function of cross-country variation in this endowment and cross-industry variation in the factor intensity. The horizontal axes rank countries according to factor endowments $\log(\ell_i)$, (over the set of countries $i \in I$) and industries according to γ^k , (over sectors $k \in K$). The reason for working with logs and elasticities will become apparent from the linearisation undertaken in section 4.

As expected, ℓ -abundant countries have high production in industries in which the share of this factor is large (high γ^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 ℓ -abundant countries increases output for products with high γ , and decreases it for products with low γ . Some intermediate industry, with factor intensity $\bar{\gamma}$, has output level independent of the endowment of this factor.⁶ The arrow marked H shows how, for a particular country, the structure of production depends on its factor endowment; an ℓ -scarce economy has relatively high production in low γ 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. Notice that the assignment of distinct varieties of product to each country means that output levels r_i^k are determinate, regardless of the numbers of goods and factors.

An analogous pattern emerges with intermediate inputs. If countries differ in the price of the intermediate input and industries differ in the share of the intermediate in their costs, then a similar saddle-shape surface of outputs is generated. This effect corresponds to ‘forward linkages’, and we develop it fully in section 4.2.

3.2 Geography and demand.

The presence of transport costs means that industry output levels are influenced by the location of demand. These effects are contained in equation (5), referred to as the market potential of industry k in country i , and denoted $m(\mathbf{u}^k : i)$,

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

This function is indexed across countries, i , and the vector \mathbf{u}^k refers to industry characteristics that interact with the spatial distribution of demand. The first of these characteristics that we look at is the cross-industry variation in transport intensity. We suppose that transport costs are an isoelastic function of the distance between locations i and j , denoted d_{ij} , and write

$$t_{ij}^k = d_{ij}^{\theta^k}, \quad (10)$$

so θ^k is the transport intensity of industry k . We now pose the question, where do industries with high or low transport intensity locate?

To answer this it is convenient to define a measure of the market potential of an average or reference industry. If $\bar{\theta}$ is the transport intensity, and \bar{e}_j and \bar{G}_j the cross-country patterns of expenditure and price indices of this reference industry, then the reference market potential is

$$m(\bar{\mathbf{u}} : i) \equiv \sum_j \left(d_{ij}^{\bar{\theta}} \right)^{1-\eta} \bar{e}_j \left(\bar{G}_j \right)^{\eta-1}. \quad (11)$$

We will refer to this as *the* market potential of country i (although market potential is properly defined as an industry specific variable). This measure will be high in countries that have or are close to large markets.

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 have lower transport costs to other markets, and as a consequence, higher market potential. The figure plots out the surface of $\log(r_i^k)$ against industries' transport intensities, θ^k , 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.⁷

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 (8) rules this out. Alternatively, it may be due to the spatial distribution of *derived* demand varying across industries -- backwards linkages. Thus, the distribution of demand for the composite intermediate varies across countries, and this country characteristic interacts with the share of each industry's output that goes for intermediate usage. The interaction of this pair of country and industry characteristics gives rise to a saddle shaped output surface, just like figures 1 and 2. It provides the basis for our modelling of 'backwards linkages', developed fully in section 4.4.

4. Linearization

To estimate the model we log-linearise around a reference point. Equation (5) above gives the relative value of output of each industry in each country, and we now rewrite this, using (9), as

$$r_i^k = c(\mathbf{v}_i : k)^{1-\eta} m(\mathbf{u}^k : i) \exp[\varepsilon_i^k] \quad (12)$$

The functions $c(\mathbf{v}_i : k)$ and $m(\mathbf{u}^k : i)$ have the properties that: (i) there exists an input price vector, $\bar{\mathbf{v}}$, at which $c(\bar{\mathbf{v}} : k) = 1$ for all industries k , and: (ii) there exists a vector of industry characteristics, $\bar{\mathbf{u}}$, such that $m(\bar{\mathbf{u}} : i) = 1$ for all countries i . These define our reference country and industry. Linearising (12) around the reference point gives

$$\Delta r_i^k = (1 - \eta) \boldsymbol{\gamma}(\bar{\mathbf{v}} : \mathbf{k}) \cdot \Delta \mathbf{v}_i + \boldsymbol{\mu}(\bar{\mathbf{u}} : \mathbf{i}) \cdot \Delta \mathbf{u}^k + \varepsilon_i^k \quad (13)$$

where Δ denotes a proportionate change, e.g. $\Delta v_i = d \log(v_i) \approx \log(v_i) - \log(\bar{v})$. $\boldsymbol{\gamma}(\bar{\mathbf{v}} : \mathbf{k}) \equiv (\partial c / \partial v)(c / \bar{v})$ is the (row) vector of elasticities of industry k costs with respect to input prices,

equal to the shares of each input in costs, and $\boldsymbol{\mu}(\bar{\mathbf{u}}:i) \equiv (\partial m/\partial \mathbf{u})(m/\bar{\mathbf{u}})$ is the vector of elasticities of country i reference market potential with respect to industry characteristics. Notice that, evaluating the differential at the reference point, there is no cross-industry variation in costs (since $c(\bar{\mathbf{v}}:k) = 1$ for all k) or cross-country variation in market potential (since $m(\bar{\mathbf{u}}:i) = 1$ for all i).

Since the z_i^k are shares, deviations from the reference point are both positive and negative and it must be the case that $\sum_i \sum_k z_i^k \Delta r_i^k = 0$. Using this with equation (13) gives

$$\begin{aligned} \Delta r_i^k = (1 - \eta) [\boldsymbol{\gamma}(\bar{\mathbf{v}}:k) - \sum_i \sum_k z_i^k \boldsymbol{\gamma}(\bar{\mathbf{v}}:k)] \cdot \Delta \mathbf{v}_i + \\ [\boldsymbol{\mu}(\bar{\mathbf{u}}:i) - \sum_i \sum_k z_i^k \boldsymbol{\mu}(\bar{\mathbf{u}}:i)] \cdot \Delta \mathbf{u}^k + \varepsilon_i^k. \end{aligned} \quad (14)$$

The double summation terms in (14) do not vary over either industries or countries, so we write

$$\Delta r_i^k = (1 - \eta) [\boldsymbol{\gamma}(\bar{\mathbf{v}}:k) - \bar{\boldsymbol{\gamma}}] \cdot \Delta \mathbf{v}_i + [\boldsymbol{\mu}(\bar{\mathbf{u}}:i) - \bar{\boldsymbol{\mu}}] \cdot \Delta \mathbf{u}^k + \varepsilon_i^k. \quad (15)$$

Thus, we express the cross-country and cross-industry variation in r_i^k as the sum of supply side and demand side considerations. On the supply side, they are given by the interaction between input shares and input prices, both expressed as deviations from some reference value. And on the demand side, they are given by the interaction between the elasticities of market potential with respect to industry characteristics, and a vector of industry characteristics. Notice also that the terms $\boldsymbol{\gamma}$ and $\boldsymbol{\mu}$ are elasticities, while $\Delta \mathbf{v}_i$ and $\Delta \mathbf{u}^k$ are in logs.

The inner products in equation (15) define a set of interactions between industry and country characteristics that will form the basis of our estimation. In the econometric implementation of the model we use six interactions, and we now explore each of these in turn, looking first at the cost side then at the demand side.

4.1: Primary factors; [Interactions 1, 2 & 3; Factor intensities and factor endowments]

On the cost side, input prices include both primary factor prices and intermediates good prices. Since the treatment of these is different, we partition the vector of input prices into $\mathbf{v}_i = [\mathbf{w}_i; q_i]$, and the

vector of shares into $\gamma = [\gamma_w : \gamma_q]$. We look first at primary factors.

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

$$\Delta w_i = \mathbf{H} \cdot \Delta \mathbf{L}_i \quad (16)$$

where $\Delta \mathbf{L}_i$ is the vector of variations in endowments from the reference point, and \mathbf{H} is the matrix of elasticities of factor prices with respect to endowments, evaluated at the reference point. Using this in equation (15) and ignoring all other effects, gives

$$\Delta r_i^k = (1 - \eta)[\gamma_w(\bar{v} : k) - \bar{\gamma}_w] \cdot \mathbf{H} \cdot \Delta \mathbf{L}_i \quad (17)$$

Several points need to be made about this equation. First, the matrix \mathbf{H} is derived by totally differentiating (6), letting both techniques of production and output quantities change. Details of the derivation of expressions (16) and (17) are given in appendix A2, which also derives explicit expressions for the two-industry two-factor case. It shows how, as $\eta \rightarrow \infty$, the model produces standard Rybczynski effects, and factor prices become invariant with respect to endowments. Second, although the sign pattern of the matrix \mathbf{H} 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 our theory we shall simply assume that diagonal elements of \mathbf{H} 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.⁸ 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 of equation (17) to figure 1 should be clear. The quadratic form of (17), 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.⁹ 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. For agriculture, rather than using land endowments we use output of agriculture, forestry and fishery products. Details are given in appendices A3 (data sources) and A4 (construction of variables).

4.2 Intermediate goods; [Interaction 4: Intermediate supply access and forward linkages]

Costs depend not only on primary factors but also the prices of intermediate good, and we now turn to the interaction between these prices and the share of the intermediate good in production, an interaction that captures ‘forward linkages’. The spatial variation of the intermediate input price depends on proximity to supplier industries, and the effect of this on each industry depends on the share of intermediates in its costs.

The model assumes a single composite intermediate good, and the cross-country variation in the price of this good, Δq_i , interacts with cross-industry variation in intermediate input shares $\gamma_q(\bar{v}:k)$ according to,

$$\Delta r_i^k = (1 - \eta)[\gamma_q(\bar{v}:k) - \bar{\gamma}_q]\Delta q_i. \quad (18)$$

Data on intermediate input shares are readily available (appendix A3). The price of the intermediate good is constructed from the price indices of each industry according to the Cobb-Douglas aggregator given in equation (7). Price indices G_i^k are defined in equation (2), and using these with equation (3) gives:

$$\left(G_i^k\right)^{1-\eta} = \sum_j \left[\frac{z_j^k \left(t_{ji}^k\right)^{1-\eta}}{\sum_\ell e_\ell^k \left(t_{j\ell}^k / G_\ell^k\right)^{1-\eta}} \right]. \quad (19)$$

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 (19) in (7) and taking logs gives,

$$\log(q_i) = A \sum_k \frac{\phi^k}{1-\eta} \log \left[\sum_j z_j^k (t_{ji}^k)^{1-\eta} \right] = A \sum_k \frac{\phi^k}{1-\eta} \log \left[\sum_j z_j^k (d_{ji}^{\theta^k})^{1-\eta} \right] \quad (20)$$

The term in square brackets gives, for each country and industry, a distance weighted measure of proximity to production in the industry. The ϕ^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 . Details of the data used in constructing this variable are given in appendix A4.

4.3: Demand and location; [Interaction 5: Market potential and transport intensity]

On the demand side, we focus on two industry characteristics that interact with country market potential. One is transport intensity θ^k , and the other is the share of the industry's output going to intermediate production, denoted ψ^k , which is our basis for assessing backwards linkages. We therefore have $\mathbf{u}^k = [\theta^k, \psi^k]$, and our estimation strategy requires that we interact observations of these industry characteristics with the elasticities of $m(\mathbf{u}^k : i)$ with respect to the characteristic, $\boldsymbol{\mu}_\theta(\bar{\mathbf{u}} : i)$ and $\boldsymbol{\mu}_\psi(\bar{\mathbf{u}} : i)$.

Transport intensity has already been defined and discussed in section 3.2, and data comes from the GTAP trade modelling project (see Appendix A3). We interact transport intensity with the elasticity of $m(\mathbf{u}^k : i)$ with respect to transport intensity evaluated at the reference point, i.e., with $\boldsymbol{\mu}_\theta(\bar{\mathbf{u}} : i)$. Using (2), (4) and (10) in (9) gives:

$$m(\mathbf{u}^k : i) \equiv \sum_j (t_{ij}^k)^{1-\eta} e_j^k (G_j^k)^{\eta-1} = \sum_j \left[\frac{(e_j^k / s^k) (d_{ij}^{\theta^k})^{1-\eta}}{\sum_\ell s_\ell (c(v_\ell : k) d_{\ell j}^{\theta^k})^{1-\eta}} \right]. \quad (21)$$

We find the elasticity of this with respect to θ by constructing values of the numerator at two different values of θ , $\tilde{\theta}$ and $\tilde{\theta} + \Delta\theta$, while assuming that the denominator is constant. Evaluating this at the reference point gives elasticity,

$$\mu_{\theta}(\bar{u}:i) = \left(\frac{\sum_j \bar{e}_j d_{ij}^{-(1-\eta)(\bar{\theta}+\Delta\theta)} - \sum_j \bar{e}_j d_{ij}^{-(1-\eta)\bar{\theta}}}{\sum_j \bar{e}_j d_{ij}^{-(1-\eta)\bar{\theta}}} \right) \frac{\tilde{\theta}}{\Delta\theta} \quad (22)$$

Once again, appendix A4 gives details of the empirical construction of this variable.

4.4: Backward linkages [Interaction 6: Relative market potential and backward linkages].

The backwards linkage effect arises as industries that have a high share of their output going to intermediate production, ψ^k , may tend to locate in countries in which demand for intermediates is relatively high. To implement this, we split expenditure, e_i^k , into its two components. Using (8) in (21) and approximating the denominator by a constant, $1/A$,

$$m(\mathbf{u}^k:i) = A \sum_j \left(\alpha^k f_i + \varphi^k q_i y_i \right) \left(t_{ij}^k \right)^{1-\eta} \quad (23)$$

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

$$\psi^k \equiv \varphi^k \sum_i q_i y_i / \sum_i e_i^k, \quad \text{and} \quad 1 - \psi^k = \alpha^k \sum_i f_i / \sum_i e_i^k \quad (24)$$

then using (24) in (23) gives,

$$m(\mathbf{u}^k:i) = A \left[(1-\psi^k) \sum_j \left(t_{ij}^k \right)^{1-\eta} f_j / \sum_j f_j + \psi^k \sum_j \left(t_{ij}^k \right)^{1-\eta} q_j y_j / \sum_j q_j y_j \right] \quad (25)$$

The elasticity of this with respect to ψ^k is

$$\mu_{\psi}(\bar{u}:i) = \left(\sum_j d_{ij}^{-(1-\eta)\bar{\theta}} \frac{q_j y_j}{\sum_j q_j y_j} - \sum_j d_{ij}^{-(1-\eta)\bar{\theta}} \frac{f_j}{\sum_j f_j} \right) \frac{m(\bar{u}:i)}{\bar{\psi}}, \quad (26)$$

which we can compute by constructing separate market potential measures for final expenditure and

for intermediate demands (appendix A4). This interacts with cross-industry variation in ψ^k , observable from input-output data. It says that the difference between each country's market potential computed for intermediate expenditures minus market potential computed on final expenditure, should be interacted with the each industry's share of output going to industry, rather than as final sales.

5. Estimation

We now turn to econometric implementation and estimation of the structure outlined above. The dependent variable is the 'double relative' measure of output, $\ln(r_{ij}^k)$, which takes into account the differing size of countries and industries, and the independent variables are the interactions between country and industry characteristics.

Denoting the country and industry characteristics $x_i[j]$ and $y^k[j]$ respectively, with j an index running over the six interactions, gives the following specification,

$$\ln(r_i^k) = \sum_j \beta[j] (x_i[j] - \bar{x}[j]) (y^k[j] - \bar{y}[j]) + \varepsilon_i^k \quad (27)$$

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

$$\ln(r_i^k) = \xi + \sum_j (\beta[j] x_i[j] y^k[j] - \beta[j] \bar{y}[j] x_i[j] - \beta[j] \bar{x}[j] y^k[j]) + \varepsilon_i^k. \quad (28)$$

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, ξ , 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. 20)	log	Intermediate intensity	Elasticity
$j = 5$	Elasticity of market potential w.r.t. transport costs (eqn. 22)	Elasticity	Transport costs	log
$j = 6$	Relative market potential (eqn. 26)	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. The equation is estimated by OLS, and we report standardized coefficients by conditioning on the standard deviation of the underlying variables. 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.

Our first estimates (column 2 of Table 2) are derived by pooling across the four time periods giving us 1824 observations.¹⁰ Pooling across years implicitly assumes that the parameters of equation (28) 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 ($\bar{x}[j]$), those defining the reference industry can change ($\bar{y}[j]$), 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 (28) 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 1994-97.¹¹ These estimates are given in remaining columns of Table 2.

5.2 Results

The first row of Table 2 gives the constant term. The next six rows give the estimated coefficients on $x_i[j]$, the country characteristics. From the estimating regression, we see that this is an estimate of $-\beta[j]\bar{y}[j]$. The following six rows give the estimated coefficients on industry characteristics, $y^k[j]$, estimates of $-\beta[j]\bar{x}[j]$. Finally, the next six rows give the coefficients on the interaction variables, the estimates of $\beta[j]$. We first discuss these interaction coefficients, and then turn to discussion of the estimates of $-\beta[j]\bar{y}[j]$ and $-\beta[j]\bar{x}[j]$.

The first three $\beta[j]$ coefficients cover factor endowments and factor intensities. They all have the same signs as predicted by theory and in the pooled sample are significant at the 5% level or better. Looking at the estimates for separate years we see that coefficients are increasing in magnitude, and in the last period agriculture, skills, and R&D are significant at the 5%, 1% and 1% levels respectively. The coefficients are smaller for agricultural intensity than for skill and R&D intensity, indicating lower elasticities, and that the related endowments have a weaker impact on production shares. We discuss the economic interpretation of the magnitude of these coefficients later.

$\beta[4]$ and $\beta[6]$ are the forward and backward linkages respectively. They have the right sign and are significant at the 5% and 1% level in the pooled sample and the 10% and 5% level in 1994-97. The coefficients measure the elasticity of production share with respect to location (measured by supplier access or by relative market potential) for an industry with intermediate intensity (forward linkage) or share of output going to industry (backward linkage) one standard deviation above the

corresponding $\bar{y}[j]$. There is evidence 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.

$\beta[5]$ is the interaction coefficient on market potential interacted with transport intensity. This coefficient has the wrong sign, significantly so in the pooled sample, although not in any of the separate sub-periods. Thus, it suggests that high transport intensity industries tend to locate in countries with lower market potential, the opposite of the case illustrated in figure 2. There are two likely reasons for this. One is the quality of the data on cross-industry variation in 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). However, we also experimented with measures based on tradability (defined as the ratio of the sum of exports and imports to gross value of output), which had little effect on results. The second possible reason is that, as we noted in section 3.2, the saddle relationship between transport intensity and market potential is a local relationship, and very high transport intensity industries locate just according to local demand. We experimented with splitting the sample between high and lower transport intensity industries, but again without great success.

While the $\beta[j]$ coefficients are the main focus of interest, it is worth making a few points about the estimates $\beta[j]\bar{y}[j]$ and $\beta[j]\bar{x}[j]$ given in the upper part of the table. Dividing by estimates of $\beta[j]$ gives estimates $\bar{y}[j]$ and $\bar{x}[j]$, which are the points along which the surfaces in figures 1 and 2 are flat (see also equation (27)). For around 80% of our estimates, these lie 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 responds to a change in country characteristics included both positive and negative responses.

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.¹² 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.

As we have already noted, the interaction coefficients, $\beta[j]$, measure the response of production to either country or industry characteristics. This is most easily seen for factor endowments/ intensities,

where they can be related to the H-effects and R-effects of figure 1. Consider an industry with skilled labour intensity (characteristic $j=2$) one standard deviation above $\bar{y}[2]$, so $y^k[2] - \bar{y}[2] = 1$ (recall that variables are conditioned on their standard deviation). $\beta[2]$ then measures the R-effect (our generalization of the Rybczynski effect) for this industry, and says that the elasticity of this industry's output share in each country with respect to the share of that country's labour force that is skilled, is 1.66 (using the 1994-97 estimate). Clearly, for an industry with skilled labour intensity one standard deviation below $\bar{y}[2]$ the R-effect changes sign, and the elasticity becomes -1.66. The H-effects are analogous. For a country with skilled labour share one standard deviation above $\bar{x}[2]$ the elasticity of output share with respect to the skilled labour intensity of the industry is 1.66.

Looking across the distribution of industry characteristics, $y^k[j]$, we can calculate R-effects for each industry, and these are reported in Table 3 for changes in endowments of skilled labour and of researchers and scientists. The numbers given in the table are the elasticity of each industry's output share with respect to share of labour force that is skilled (columns 2 and 3) or the number of researchers and scientists per thousand employees (columns 4 and 5). Looking first at skilled labour, we see positive R-effects for 26 of the 33 industries (1994-97 data), with the largest effects occurring in Professional Instruments, followed by Drugs and Medicines and Printing and Publishing. The ranking directly reflects the skill-intensity of the industries. For the R&D endowment, only three industries have positive R-effects – Aircraft, Drugs and Medicines, and Radio, TV and Communications equipment.¹³ (The common values for some of the R&D un-intensive industries reflect the fact that we do not have R&D data for each of each of these industries).

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

Variable	Pooled	1980-83	1985-88	1990-93	1994-97
CONSTANT, ξ	7.753*** (2.471)	6.798 (5.257)	7.647 (5.264)	5.985 (4.979)	12.820** (5.637)
Country Characteristic: $-\beta[j]y[j]$					
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: $-\beta[j]x[j]$					
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: $\beta[j]$					
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%, one sided tests. All regressions are overall significant according to standard F-tests.

Table 3: R-effects: $\beta[j](y^{k[j]} - \bar{y}[j]), j = 2, 3.$

	Skill intensity	Skill intensity	R&D intensity	R&D intensity
	pooled	1994-97	pooled	1994-97
Food	-1.281	-0.718	-1.607	-3.350
Beverages	0.147	0.766	-1.607	-3.350
Tobacco	-1.591	-1.041	-1.607	-3.350
Textiles	-0.616	-0.027	-1.629	-3.410
Wearing Apparel	-0.678	-0.092	-1.629	-3.410
Leather & Products	-1.168	-0.602	-1.629	-3.410
Footwear	-1.103	-0.534	-1.629	-3.410
Wood Products	-0.492	0.101	-1.630	-3.411
Furniture & Fixtures	-0.186	0.420	-1.630	-3.411
Paper & Products	-0.309	0.292	-1.617	-3.378
Printing & Publishing	3.578	4.332	-1.617	-3.378
Industrial Chemicals	0.434	1.064	-0.959	-1.604
Drugs & Medicine	3.753	4.515	0.356	1.944
Chemical Products nec	1.325	1.990	-0.959	-1.604
Rubber Products	0.775	1.419	-1.464	-2.964
Plastic Products	-0.075	0.535	-1.464	-2.964
Pottery & China	0.858	1.505	-1.548	-3.192
Glass & Products	0.032	0.646	-1.548	-3.192
Non-Metallic Minerals nec	0.024	0.638	-1.548	-3.192
Iron & Steel	-0.233	0.370	-1.546	-3.186
Non-Ferrous Metals	-0.901	-0.324	-1.544	-3.180
Metal Products	0.898	1.546	-1.519	-3.114
Office & Computing Machinery	3.204	3.944	-0.475	-0.298
Machinery & Equipment	1.466	2.137	-1.242	-2.367
Radio,TV & Communication	2.723	3.444	0.343	1.909
Electrical Apparatus nec	1.437	2.107	-0.963	-1.614
Shipbuilding & Repairing	0.963	1.614	-1.496	-3.050
Railroad Equipment	2.160	2.858	-1.110	-2.009
Motor Vehicles	-0.419	0.178	-0.970	-1.633
Motorcycles	-0.395	0.202	-1.110	-2.009
Aircraft	3.460	4.210	0.574	2.531
Transport Equipment nec	0.198	0.819	-1.110	-2.009
Professional Instruments	4.131	4.907	-1.122	-2.042

5.3 Robustness

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 (5) 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 endowments or industry intensities might be consistently mismeasured for one particular industry or country. Again, from equation (5) 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 (28), dropping the 12 country and industry levels variables. The results for the interaction variables for each of the years are reported in Table 4. 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 $\beta[j]$ 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, $\beta[j]$, in Table 5. Once again, we see that the coefficients are stable across the specifications.

Table 4: $\beta[j]$, Robustness Check I: Fixed effects: Dependent variable $\ln(r_{ij}^k)$

Variable	1980-83	1985-88	1990-93	1994-97
Agriculture endowment * agricultural intensity	0.077 (0.126)	0.135 (0.106)	0.163* (0.087)	0.153** (0.080)
Skill endowment * skill intensity	1.492*** (0.380)	1.479*** (0.389)	1.463*** (0.437)	1.658*** (0.559)
Researchers+scientists *R&D intensity	0.588** (0.301)	0.744** (0.376)	1.112** (0.506)	1.630*** (0.546)
Supplier access * intermed. intensity	0.564 (0.787)	0.757 (0.753)	0.801 (0.659)	1.101* (0.684)
Market pot. elasticity * transport costs	-0.405 (0.307)	-0.275 (0.291)	-0.323 (0.282)	-0.380 (0.267)
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

Table 5: $\beta[j]$, Robustness Check II, 1994-97: Dependent variable $\ln(r_{ij}^k)$

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

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, including imperfect competition creates 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: 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 $\eta = 5$, 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, θ^k , and the market potential of different countries, computed from equation (11) for the middle ranked industry.

Appendix A2: Factor endowments, factor prices, and outputs.

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

$$x^k = A^k c(\mathbf{v};k)^{-\eta}. \quad (29)$$

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 , γ_w^k and γ_r^k . Considering the effects of factor price changes on outputs gives

$$\Delta x^k = -\eta(\gamma_w^k \Delta w + \gamma_r^k \Delta r) \quad (30)$$

Factor demand equations are

$$L^k = \frac{\partial c(w,r;k)}{\partial w} x^k, \quad K^k = \frac{\partial c(w,r;k)}{\partial r} x^k \quad (31)$$

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

$$\begin{aligned} \Delta L^k &= \gamma_r^k \sigma^k (\Delta r - \Delta w) - \eta(\gamma_w^k \Delta w + \gamma_r^k \Delta r), \\ \Delta K^k &= \gamma_w^k \sigma^k (\Delta w - \Delta r) - \eta(\gamma_w^k \Delta w + \gamma_r^k \Delta r) \end{aligned} \quad (32)$$

where σ^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

$$\frac{wL}{Y} \Delta L = \sum_k \gamma_w^k s^k \Delta L^k, \quad \frac{rK}{Y} \Delta K = \sum_k \gamma_r^k s^k \Delta K^k \quad (33)$$

Using (32) in (33) and applying Cramer's rule we can express changes in factor prices as a function of changes in endowments, ΔL and ΔK . This relationship is the matrix H . In general, we can solve for factor prices as a function of endowments, and then use the result back in (30) for the associated changes in production levels.

General expressions are not very insightful, but if we assume that there are just two industries and that η is very large relative to σ^k (so $\sigma^k = 0$ in equations (32)) then

$$\begin{bmatrix} \left(\gamma_w^1\right)^2 s^1 + \left(\gamma_w^2\right)^2 s^2 & \gamma_w^1 \gamma_r^1 \mu^1 + \gamma_w^2 \gamma_r^2 s^2 \\ \gamma_w^1 \gamma_r^1 s^1 + \gamma_w^2 \gamma_r^2 s^2 & \left(\gamma_r^1\right)^2 s^1 + \left(\gamma_r^2\right)^2 s^2 \end{bmatrix} \begin{bmatrix} \Delta w \\ \Delta r \end{bmatrix} = \begin{bmatrix} -\frac{wL}{Y} \frac{\Delta L}{\eta} \\ -\frac{rK}{Y} \frac{\Delta K}{\eta} \end{bmatrix} \quad (34)$$

(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 (34) we derive

$$\frac{\Delta r}{\Delta K} = -\left(\frac{rK}{\eta Y}\right) \frac{\left(\gamma_w^1\right)^2 s^1 + \left(\gamma_w^2\right)^2 s^2}{s^1 s^2 \left(\gamma_w^2 - \gamma_w^1\right)^2}, \quad \frac{\Delta w}{\Delta K} = \left(\frac{rK}{\eta Y}\right) \frac{\left(\gamma_w^1 \gamma_r^1 s^1 + \gamma_w^2 \gamma_r^2 s^2\right)}{s^1 s^2 \left(\gamma_w^2 - \gamma_w^1\right)^2} \quad (35)$$

These are two terms in the matrix H . Notice that they are inversely proportional to η . Thus, as $\eta \rightarrow \infty$, 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 (35) in (30) we can derive the effects of factor endowments on outputs. This simply takes the form

$$\Delta x^1 = -\Delta K \frac{\gamma_r^1 \gamma_w^2}{\gamma_w^1 - \gamma_w^2} \quad (36)$$

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 A3: 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 specification 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 percentage of fob priced sales within the EU (i.e. basis for calculation is intra-EU trade). source: The GTAP 4 Data Base (McDougall 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 percentage of total sales)*: Percentage of domestic sales to domestic manufacturing as intermediates and capital goods** 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:

*) R&D expenditure is not available for all EU countries. We use data for Denmark, Finland, France, Germany (former FRG), Italy, Netherlands, Spain, Sweden and the UK. The calculated R&D share of gross value of output is a weighted average for these countries for the year 1990.

***) IO tables are not available for all EU countries. 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, had to be mapped into the 36 sectors classification.

Table A1: Industry intensities

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 output
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	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	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 &	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

Appendix A4: Construction of variables:

1) Dependent variable: Δr_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 A3).

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

A) *Share of factors in costs of each industry, γ_w :*

- i) Skilled labour intensity: proxied by the product of 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:*

- 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).

2) Intermediate composite good:

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

B) Supplier access/ Prices of intermediate: Implementation of equation (20) requires:

- i) Production levels : z_j^k ; value of output data for the following 25 sectors: manufacturing (22 sectors), agriculture, mining and quarrying, and services.
- ii) Shares of each industry in intermediate, ϕ^k : Sales to aggregate manufacturing industry by each of the 25 sectors above, expressed as share of gross output.
- iii) Distance, d_{ij} : Distance between the economic centre of gravity of countries. Centres of gravity computed from GDP at subnational (NUTS2) data. 'Internal distance', d_{ii} is set to one.
- 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. Assumed the same in all sectors.

3) Demand and location:

A) Transport intensities, θ^k : Transport costs as percentage of fob priced sales.

B) Elasticity of market potential, $\mu_\theta(\bar{u}: i)$ Implementation of equation (22) requires:

- i) Reference expenditure \bar{e}_j ; proportional to GDP_j .

- ii) Distance, reference transport intensity, as above, i.e., $\tilde{\theta}(1 - \eta) = -1$.
- iii) Computation of elasticity: we experiment with different values, and results are reported for $\theta = 0.7$ and $\Delta\theta = 0.6$.

4) Backwards linkages:

- A) Share of industry's sales going to manufacturing industry, ψ^k : from input-output tables.
- B) Elasticity of market potential, $\mu_\psi(\bar{u}:i)$ Implementation of equation (26) requires:
 - i) Distance, reference transport intensity, as above, i.e., $\tilde{\theta}(1 - \eta) = -1$.
 - ii) Spatial distribution of final expenditure, $f_i / \sum_i f_i$: use GDP_i .
 - iii) Spatial distribution of intermediate expenditures, $q_i y_i / \sum_i q_i y_i$. From equation (8), expressed in value terms; z_i^k and γ_q as in 2 above.

Endnotes:

1. See Midelfart-Knarvik, Overman, Redding and Venables, 2000
2. See Leamer and Levinsohn (1995) for a discussion and critique of this and other approaches.
3. See Fujita, Krugman and Venables (1999).
4. Letting this elasticity differ across industries would be straightforward in the theoretical sections, but a common value is assumed in the empirical estimation.
5. 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.
6. Local perturbation of the endowment in this direction has no effect on r_i^k .
7. The figure only illustrates the range in which the saddle shape holds. Increasing transport intensity further causes a flattening of the surface. As $\theta^k \rightarrow \infty$ so the market potential of industry k becomes equal to local demand.
8. We can in principle estimate with the full matrix \mathbf{H} , not just the diagonal, but the resulting specification is beset by multi-collinearity problems.
9. 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.
10. Observations for the following countries/industries are missing: Denmark: ISIC 3842, 3845, 3849; France: ISIC 3849; Ireland: ISIC 3130; Netherlands: ISIC 3842.
11. Output and country characteristics vary through time, although industry characteristics are held constant (as in table A1). Separating the years also reduces the degree of exogeneity of some of the explanatory variables. 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
12. See Midelfart-Knarvik, Overman, Redding and Venables, 2000
13. The preponderance of negative values reflects the use of Scientists in non-manufacturing sectors of the economy. See Harrigan (1997) for a similar finding.

References

- Baldwin, R.E. (1971), 'Determinants of the commodity structure of US trade', *American Economic Review*, 61, 126-146.
- Barro, Robert and Jong-Wha Lee (1993): "International Comparisons of Educational Attainment", NBER Working Paper no. 4349
- Davis, D. and D. Weinstein (1998): "Market access, economic geography and comparative advantage: an empirical 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 advantage 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 evidence for the OECD', *Journal of International Economics*, 39, 123-141.
- Harrigan, J. (1997) 'Technology, factor supplies and international specialization; estimating the neoclassical model', *American Economic Review*, 87, 475-494.
- 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. Rogoff (eds) *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, European Commission, Economic papers No. 142. April 2000, European Commission, Brussels.

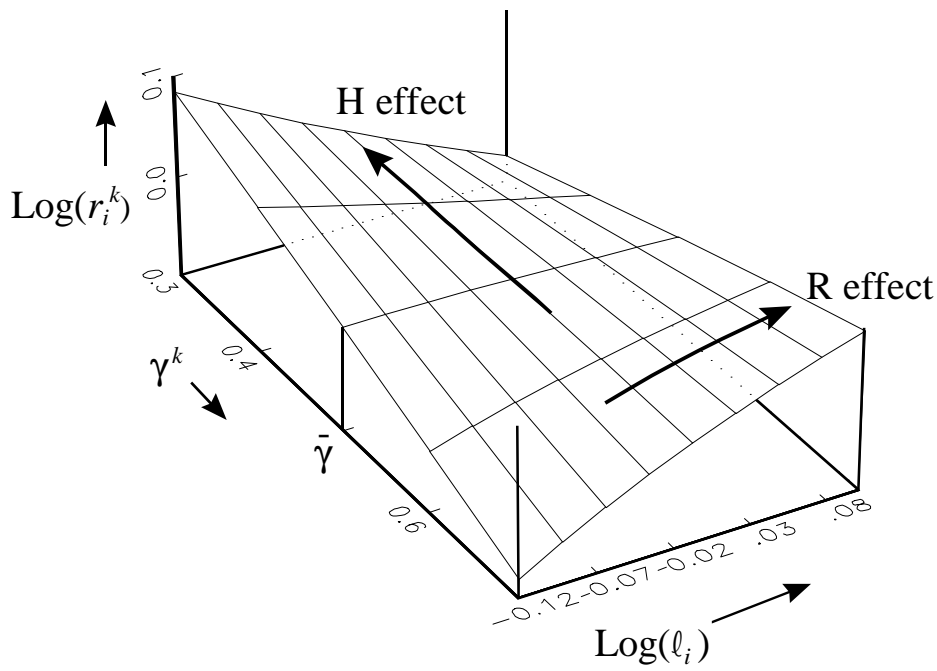


Figure 1: Cost share and endowment

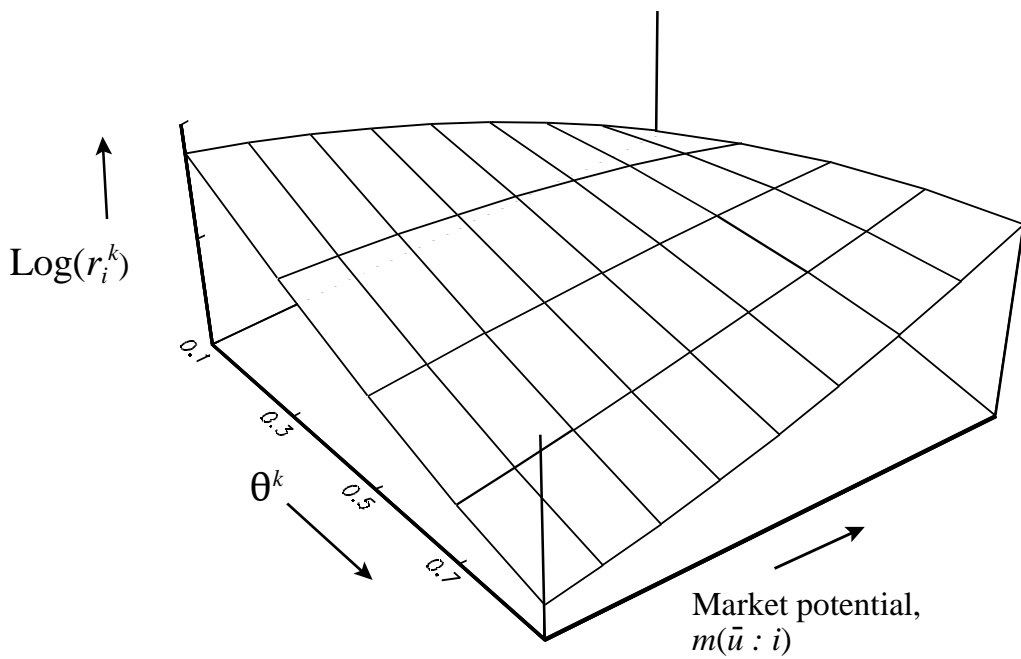


Figure 2: Transport intensity and market potential