Factor Supplies and Specialization in the World Economy

James Harrigan and Egon Zakrajšek[1]

August 2000

Abstract

A core prediction of the Heckscher-Ohlin theory is that countries specialize in goods in which they have a comparative advantage, and that the source of comparative advantage is differences in relative factor supplies. To examine this theory, we use the most extensive dataset available and document the pattern of industrial specialization and factor endowment differences in a broad sample of rich and developing countries over a lengthy period (1970-92). Next, we develop an empirical model of specialization based on factor endowments, allowing for unmeasurable technological differences and estimate it using panel data techniques. In addition to estimating the effects of factor endowments, we also consider the alternative hypothesis that the level of aggregate productivity by itself can explain specialization. Our results clearly show the importance of factor endowments on specialization: relative endowments *do* matter.

JEL code: F1

Factor Supplies and Specialization in the World Economy

Since Ricardo, trade economists have had a persuasive explanation for international output specialization: countries specialize in goods in which they have a comparative advantage. The central problem with this elegant theory is that it links the observables to be explained, outputs, to inherently unknowable, if not metaphysical, autarky prices. The theory of comparative advantage is empirically empty unless autarky prices can be linked to observables, as they are in the Ricardian, Heckscher-Ohlin, and other versions of the theory. If trade theory is to be useful in understanding the world, it is imperative to confront these models with the data.

Our paper is a contribution to this project. We collect and analyze the most extensive data set currently available on production and factor supplies, with a focus on the question: how does the world distribution of productive resources influence the pattern of output specialization?

We look at the pattern of specialization rather than the pattern of trade, because most of the intellectual capital of trade theory is invested in explaining production. Almost all flavors of comparative advantage theory combine a sophisticated model of production with a rudimentary, if not naïve, model of consumption. The best-known example of this is the Heckscher-Ohlin theorem, the proof of which consists of the remarkable Rybczynski theorem combined with the assumption of identical and homothetic preferences[1]. Despite the fact that the bulk of the intellectual content of comparative advantage theory is about production, almost all empirical work on comparative advantage, from Leontief (1954) to Trefler (1995), has used trade data and has not directly measured production.

Reasoning that economists won't be able to understand trade until they understand specialization, Leamer (1987), Harrigan (1995, 1997), Bernstein and Weinstein (1998), and Schott (1999) have looked directly at production data. Each of these papers focused on endowment differences as a source of specialization. With the exception of Harrigan (1997), who showed the importance of industry-specific technological differences, these papers used quite restrictive models: Harrigan (1995) and Bernstein and Weinstein (1998) used the even general equilibrium model with factor price equalization, while Leamer (1987) and Schott relied on the 2- or 3-factor identical technology model. Related work on the factor content of trade by

---

[1] Plus, of course, the assumption that countries share the same technology.

Trefler (1993, 1995) also used restrictive models that relied heavily on modified forms of factor price equalization.

In contrast, Harrigan (1997) used a more flexible model that did not rely on factor price equalization assumptions and allowed for non-neutral technology differences. Using a within-country estimator, Harrigan (1997) found that technological differences were an important determinant of specialization in a panel of OECD countries. Harrigan's statistical model with country fixed effects offered consistent estimates but has the conceptual disadvantage that the model did not use cross-country variation to help identify the effects of factor supplies on specialization.

Our paper builds on Harrigan (1997) and the related literature in several ways. We ask the question: how well can relative factor endowments *alone* explain specialization? To answer this question, we begin with an extended data description that documents the pattern of industrial specialization and factor endowment differences in a broad sample of countries over a long time period (1970-92). This data description reveals the importance of country-specific influences on specialization but also suggests an important role for factor supply differences.

Next, we develop an econometric model that allows us to estimate the effects of factor endowments alone in a world where technology differences may also be important influences on specialization. Unlike Harrigan (1995), Trefler (1993, 1995), and Bernstein and Weinstein (1998), we make no use of any form of factor price equalization result. Unlike Harrigan (1997), we bypass the difficult problems of measuring technology levels, and we also dispense with the strong assumption that cross-country technology differences are exclusively Hicks-neutral at the industry level. We use panel data techniques to estimate this flexible model of specialization as a function of factor endowments.

A further contribution of our paper is that we consider an explicit alternative hypothesis. This competing explanation is the ladder-of-development or product-cycle hypothesis: a country's output mix depends on its stage of development, with countries moving from agriculture to labor-intensive manufactures to high-tech manufacturing and services as their aggregate labor productivity increases. This development story is consistent with both technology and factor supplies being important, but it is simpler and more parsimonious than models that

stress the interactions between factor supplies, factor intensities, and non-neutral technological differences.

Our results show that factor endowments *are* a major influence on specialization: for most large industrial sectors, relative factor supplies are a statistically and economically significant determinant of the location of production. However, the simple ladder-of-development model also has good explanatory power and dominates the factor proportions model on purely statistical grounds. We interpret these results as confirming the empirical relevance of factor proportions theory, and as suggesting that a full account of the workings of the global economy must assign an important role to relative factor supply differences.

1 Theory

The theory used to frame the data analysis is resolutely neoclassical: technology is assumed to be constant returns to scale, and markets are assumed to be perfectly competitive. We dispense with most of the other assumptions that are usually used in trade theory models, such as the assumptions that production is non-joint with some specific relationship between the number of goods and factors. We also make no use of any form of factor price equalization result, either relative or absolute. The use of a neoclassical model is not intended to rule out the importance of increasing returns or economic geography for specialization, but these considerations are very difficult to nest in a model that has relative factor endowments as a central driving force for specialization[2].

1.1 Technology differences and the revenue function

A convenient way to summarize the production side of the neoclassical model is with the revenue function, which gives the maximal level of national income Y for given endowments $\mathbf{v}$ and final goods prices $\mathbf{p}$[3]:

$$Y = r(\mathbf{p}, \mathbf{v}) \qquad\qquad \mathbf{p} \in \mathbb{R}^N, \mathbf{v} \in \mathbb{R}^M$$

The revenue function $r(\mathbf{p},\mathbf{v})$ is homogeneous of degree one in $\mathbf{p}$ and in $\mathbf{v}$. The net output vector $\mathbf{y}$ of the economy is given by the gradient of $r(\mathbf{p},\mathbf{v})$ with respect to prices[4]:

---

[2] For some progress on marrying economic geography and factor proportions models, see Davis and Weinstein (1999).

[3] For a careful development of the revenue function and its properties, see Dixit and Norman (1980) and Woodland (1982).

$$\mathbf{y} = r_{\mathbf{p}}(\mathbf{p}, \mathbf{v}) \qquad\qquad \mathbf{y} \in \mathbb{R}^{N} \qquad\qquad (1)$$

If technology is the same across countries, then (1) says that outputs differ only to the extent that **p** and **v** differ. In the more likely case that technology differs across countries, outputs will differ even for the same **p** and **v**. If we allow technology to differ arbitrarily across countries then (1) is useless as a framework for cross-country analysis. By restricting the way that technology differs across countries, however, modified versions of (1) can be used to study variation in outputs over time and across countries.

A simple specification is to suppose that technology differs in a Hicks-neutral fashion across industries and countries. Let θ_{ict} be a scalar productivity parameter for industry i, country c, in year t, relative to a numeraire value of $\theta_{i11} = 1$ in country 1 in year 1. By definition, Hicks-neutral technology differences mean that, given the same inputs, industry i in country c in year t produces θ_{ict} times as much output as the numeraire country/year. An appealing aspect of this specification is that Hicks-neutral technology differences are in principle measurable by applying the theory of total factor productivity measurement to data on industry inputs and outputs (see, for example, Caves, Christensen and Diewert (1982)). It is straightforward to show (e.g., Dixit and Norman (1980)) that Hicks-neutral technology differences have general equilibrium effects on outputs that can be summarized by

$$\mathbf{y} = r_{\mathbf{p}}(\boldsymbol{\Theta}\mathbf{p}, \mathbf{v}) \qquad\qquad (2)$$

where $\Theta = diag\{\boldsymbol{\theta}\}$ is a diagonal matrix of Hicks-neutral technology differences. This is the model implemented by Harrigan (1997).

Alternatively, technology may differ because of differences in factor quality across countries. For example, a hectare of agricultural land may differ in productivity across countries, or a primary school education may embody greater human capital in one country than in another. In this case, observed cross-country technology differences at the industry level arise because of the industry's use of factors that differ in their quality across countries. If we denote the quality of factor j in country c in year t, again relative to a base country/year, as λ_{jct} , then we have the result (Dixit and Norman (1980)) that

---

[4] These derivatives need not exist and will not if there are more produced goods than factors and there is no joint production. For our purposes in this paper, potential output indeterminacy is an empirical issue, and we assume differentiability of $r(\mathbf{p},\mathbf{v})$ for the rest of this section for expositional convenience.

$$\mathbf{y} = r_{\mathbf{p}}(\boldsymbol{\Lambda}\mathbf{p}, \mathbf{v}) \tag{3}$$

where $\boldsymbol{\Lambda} = diag\{\boldsymbol{\lambda}\}$ is a diagonal matrix of factor quality parameters.

If the Hicks-neutral technology parameters are constant across sectors, given by a scalar θ, then the homogeneity of $r(.,.)$ implies that aggregate nominal income is given by $Y = \theta \cdot r(\mathbf{p}, \mathbf{v})$ and

$$\mathbf{y} = \theta \cdot r_{\mathbf{p}}(\mathbf{p}, \mathbf{v}) \tag{4}$$

Similarly, if cross-country factor quality differences are the same across factors, given by a scalar λ, then we have $Y = \lambda \cdot r(\mathbf{p}, \mathbf{v})$ and

$$\mathbf{y} = \lambda \cdot r_{\mathbf{p}}(\mathbf{p}, \mathbf{v}) \tag{5}$$

Clearly, in either case the cross-country differences in technology are pure scale effects and have no impact on the *composition* of output[5]. Multiplying both sides of equations (1), (4), or (5) by a matrix $\mathbf{P} = diag\{\mathbf{p}\}$ with prices along the diagonal and dividing by nominal GDP, we can write this result as

$$\mathbf{s} = \frac{1}{Y}\mathbf{P} \cdot \mathbf{y}(\mathbf{p}, \mathbf{v}) \qquad\qquad \mathbf{s} \in S^{N-1} \tag{6}$$

where $\mathbf{s}$ is the vector of outputs as shares of GDP, and S^{N-1} is the unit simplex. It is straightforward to show that (6) is homogeneous of degree zero in $\mathbf{p}$ and in $\mathbf{v}$, meaning that only relative prices and relative endowments matter for the determination of output shares. If all countries face the same final goods prices and technology differs across countries in a neutral way, then (6) boils down to a very simple prediction: output share differences across countries and over time depend on relative factor supply differences.

---

[5] This is the model that is empirically preferred by Trefler (1995). Because technology differences, which are neutral across industries or factors, do not affect the composition of output, technology differences do not influence comparative advantage in Trefler's model.

1.2 An empirical model

  Our next step is to devise an empirical model based on the above theory that can be used to draw inferences about the effects of factor supplies on specialization. To begin, we make the assumption that r($\mathbf{p,v}$) can be approximated by a translog functional form[6]. Abusing notation slightly, let lower case non-bold face p and v denote the logs of price and factor supply vectors respectively. A translog approximation for the revenue function is then given by:

$$\ln r(\mathbf{p, v}) = \mathbf{a}'p + \mathbf{b}'v + \frac{1}{2}p'\mathbf{A}p + \frac{1}{2}v'\mathbf{B}v + p'\mathbf{R}v \tag{7}$$

The matrices $\mathbf{A}$ and $\mathbf{B}$ are symmetric, and homogeneity requires that $\mathbf{a}'\iota = \iota$, $\mathbf{b}'\iota = \iota$, $A\iota = \mathbf{0}$, $B\iota = \mathbf{0}$, $R\iota = \mathbf{0}$, and $R'\iota = \mathbf{0}$, where ι is a conformable vector of ones. In what follows, let i and k index industries, c index countries, j index factors, and t index time. Assume that (7) holds for all countries and time periods. Taking the derivative of (7) with respect to p_i gives the share of industry i in country c's GDP, denoted by s_{ict}, as :

$$s_{ict} = a_i + \sum_{k=1}^{N} a_{ik} p_{kct} + \sum_{j=1}^{M} r_{ij} v_{jct} \tag{8}$$

If there are Hicks-neutral technology differences, then (2) implies that (8) becomes

$$s_{ict} = a_i + \sum_{k=1}^{N} a_{ik} p_{kct} + \sum_{k=1}^{N} a_{ik} \theta_{kct} + \sum_{j=1}^{M} r_{ij} v_{jct} \tag{8a}$$

Similarly, if there are factor-augmenting technology differences, (3) implies that (8) becomes

$$s_{ict} = a_i + \sum_{k=1}^{N} a_{ik} p_{kct} + \sum_{j=1}^{M} r_{ij} v_{jct} + \sum_{j=1}^{M} r_{ij} \lambda_{jct} \tag{8b}$$

Notice that homogeneity of the revenue function implies that $\sum_{k=1}^{N} a_{ik} = \sum_{j=1}^{M} r_{ij} = 0$. Consequently, if the Hicks-neutral parameters do not differ across sectors, or if the factor quality terms do not differ across factors, then (8a) and (8b) collapse to (8), as would be expected from the discussion in the previous section. On the other hand, if the technology differences of either type are constant over time, but otherwise unrestricted, then the influence of technology differences collapses to a country-specific constant, and (8a)-(8b) become:

---

[6] Kohli (1991) presents a comprehensive account of this methodology for time series analysis, which Harrigan (1997) adapted for panel data models.

$$s_{ict} = a_i + b_{ic} + \sum_{k=1}^{N} a_{ik} p_{kct} + \sum_{j=1}^{M} r_{ij} v_{jct} \tag{9}$$

Unfortunately, internationally comparable data on prices are not available. Instead, we assume that trade equalizes prices across countries at a point in time up to a mean-zero error term ε_{ict} and (possibly) a country-specific mean. Then the first summation in the previous four equations becomes

$$\sum_{k=1}^{N} a_{ik} p_{kct} = d_{ic} + d_{it} + \varepsilon_{ict} \tag{10}$$

Substituting (10) into (9), we arrive at the following error-components specification for output shares:

$$s_{ict} = a_i + d_{it} + \delta_{ic} + \sum_{j=1}^{M} r_{ij} v_{jct} + \varepsilon_{ict} \tag{11}$$

This is the equation that we estimate. Given data on output shares and factor supplies over time and across countries, the parameters of (11) can be estimated by a regression of output shares on log factor supplies, treating the country effects as either fixed or random. The coefficients on log factor supplies, the r_{ij}, are related to Rybczynski derivatives: a positive estimate for r_{ij} means that accumulation of factor j raises the share of industry i in national income[7]. A zero factor supply coefficient will arise if there are no factor intensity differences across sectors, so that accumulation leads all outputs to expand proportionately.

Several features of the model in equation (11) are worth emphasizing. First, the country effects $\delta_{ic} = b_{ic} + d_{ic}$ reflect the combined influence of non-neutral technology differences, plus any differences in internal relative prices, such as differences in internal or external taxes and subsidies. Second, the model potentially applies to all sectors in the economy, not just the manufacturing sectors that we focus on in our empirical work below. Third, the assumption that the translog functional form is an adequate approximation over the entire sample may be restrictive, particularly if the effects of endowments on outputs differ depending on where a

---

[7] When every factor is used in at least two sectors, and every sector uses at least two factors, and there is non-joint production, accumulation of a factor necessarily reduces the output of at least one sector (see, for instance, Jones and Scheinkman (1977)). Because we are not imposing any such assumptions, no such result holds, and it is possible that factor accumulation raises the output of all sectors (think of labor growth in the simplest specific factors model). Nevertheless, some sectors will generally expand faster than others when factor supplies change, leading to corresponding increases and decreases in the shares of output in GDP.

country is in the space of relative endowments – this is the possibility emphasized by Schott (1999). We return to this issue at the end of Section 4. Fourth, the natural definition of specialization that comes out of this model is differences in output shares of GDP. This contrasts with ad-hoc definitions of specialization that have been used by other authors, such as output shares of tradeable goods or indexes of "revealed comparative advantage" as pioneered by Balassa (1965).

1.3 An alternative hypothesis

A limitation of most of the empirical work on comparative advantage models is that there is usually no explicit alternative hypothesis. Here we consider a simple alternative hypothesis: the level of development, rather than factor abundance per se, explains a country's output mix. In this view of the world, countries develop through capital accumulation and technological progress. With development comes a change in the output mix and pattern of trade, as countries progress from specialization in agriculture through labor-intensive manufactures, capital-intensive manufactures, and finally high-tech goods and tradeable services. This ladder-of-development model is roughly consistent with a two-factor (capital and labor), many-good model and is also consistent with product-cycle models. A simple expression of the ladder-of-development model is that output shares depend only on the aggregate productivity level, or

$$s_{ict} = a_i + d_{it} + \delta_{ic} + \beta_i \theta_{ct} + \varepsilon_{ict} \tag{12}$$

where θ_{ct} is an index of aggregate output per worker. We estimate this model below along with the factor endowments model of equation (11).

2 Data

Estimation of the model of equation (11) and (12) requires data on outputs, factor supplies, and aggregate productivity for a panel of countries. The sample used covers 28 OECD and non-OECD countries over the period 1970-1992. This is the broadest sample of countries for which data on output and factor supplies are both available and includes both developed and developing countries. See Table 1 for a description and summary of the dataset.

Our output data covers twelve tradeable sectors that cover all of manufacturing output (aside from a small "miscellaneous" category). Data on other traded goods, such as mining and agricultural output, are not available on a consistent basis.

8

We consider three types of factor supplies: land, labor, and capital. Data on aggregate capital stocks come from version 5.6 of the Penn-World Tables, available from the NBER website[8]. The Penn-World Table classifies capital stocks into producer durables, non-residential and other construction, and residential construction. We aggregate only the first two capital stock measures, because for the purposes of this paper, residential construction is most appropriately regarded as a component of consumption. Aggregate output per worker also comes from the Penn-World Tables.

Our data on land comes from the World Bank's World Tables and is defined as arable land[9]. This is a very crude measure of natural resources available for productive use, but a more nuanced treatment of natural resource abundance is beyond the scope of this paper.

We classify labor endowments according to the educational levels of workers. The data on educational attainment comes from Barro and Lee (1993), whose data are also available from the NBER website. Barro and Lee construct estimates of the level of educational attainment in the population, and we use their data to classify workers into two categories:

1. Higher Educated: workers who have at least some secondary education.

2. Lower Educated: workers who have no secondary education (most of this category consists of workers with at least some primary education).

The Barro-Lee data ends in 1985, and we extend the data through the mid-1990s, following Barro and Lee's methods and using updated versions of their data sources.

Before turning to statistical analysis, we look at the data through an extended series of tables and charts. Table 2 shows each country's output share relative to the cross-country average in 1980 (a middle year in our sample). There are two primary messages to take away from Table 2. First, most of the numbers are far from 100, which means that there is a lot of cross-country variation in output shares. Second, there is also a lot of variation across a row for a given country: countries have above average output in some sectors and below average output in other sectors. These two elements of variability, across countries for a given sector and across sectors for a given country, are what the models of this paper are trying to explain.

---

[8] www.nber.org

[9] Arable land refers to land under temporary crops, temporary meadows for mowing or pasture, and land under market and kitchen gardens.

Table 3 shows factor endowments in 1980 relative to the cross-country average. As with the output data, there is tremendous variation in relative endowments, even among the developed countries. The US is only a little above average in capital per worker, below average in land per worker, but has almost twice the percentage of highly educated workers as the sample average. Japan is similar to the United States in capital abundance, but has a tiny land endowment. Australia, New Zealand, and Canada are similar to each other: each is very land abundant, has a well-educated labor force, and capital per worker similar to the US level. Korea and Taiwan have very little land, primarily low-educated workforces, and moderate levels of capital per worker.

Turning to the time-series variation in the data, the panels of Figure 1 illustrate output share data for five of the twelve sectors[10]. A glance at these figures shows that most of the variation in the data is cross-country, but that there is substantial within-country variation as well. In the food sector, there is less of a correlation with land abundance than might have been expected. However, the food category includes processed agricultural products and excludes raw grains and produce as well as subsistence agriculture. The dominance of the Asian Tigers (Taiwan and Korea) in apparel-textiles has faded somewhat over time, but not nearly fast enough to match the virtual disappearance of apparel-textile manufacturing in places like Norway, Sweden, and the Netherlands. The poorer European countries of Portugal, Spain, Greece, and Turkey have held on to or expanded their apparel-textile sectors.

In the chemicals sector, there is great heterogeneity across countries that doesn't seem to have any simple pattern. Turning to the machinery categories, we see several striking patterns: Korea and Taiwan have seen the biggest expansions, while Germany and Japan have large sectors that stayed roughly stable over time. Many other rich countries saw their big machinery sectors gradually shrink, including Canada, the US, the UK, Sweden, Norway, and the Netherlands. In electrical and non-electrical machinery, Ireland looks more like an Asian Tiger than its European neighbors.

The informal look at the data in Figure 1 necessarily focuses on one industry at a time, while the theory of comparative advantage involves double bilateral comparisons: a comparison of relative output levels within and across countries, and a comparison of relative factor abundance within and across countries. In Figures 2 and 3, we use "star charts" to make these

double bilateral comparison for a selection of industries and countries in 1980 and 1985[11]. In Figure 2, each country's output mix is represented by five vectors. The direction of each vector identifies the industry, while the length of the vector is proportional to the size of the industry. Loosely speaking, if countries have similar shaped stars, then they have similar output mixes. If on the other hand, they have very asymmetrical stars, then they are quite different from the average country. The volume of each star also gives a rough indicator of the size of the overall manufacturing sector.

Figure 2 shows that Turkey, Greece, and Norway have relatively little GDP produced by these five sectors, while Taiwan, Korea, Germany, and Japan have large manufacturing sectors. Figure 2 also illustrates that the United States has large machinery sectors and comparatively small food and apparel-textile sectors, a pattern also seen in Japan, the UK, Germany, France, Sweden, and other rich countries. Some countries have very extreme patterns of specialization: Hong Kong's output is concentrated in apparel-textile and electrical machinery, while Chile specializes in food and apparel-textiles. In contrast to richer industrial countries, Taiwan and Korea are relatively more specialized in food and apparel-textiles and less in non-electrical equipment. Figure 2 illustrates graphically what was pointed out in Tables 2: countries differ dramatically in their output mixes.

Figure 3 uses the same technique to illustrate relative factor abundance in 1980[12]. The US is most abundant in highly-educated workers and also has a lot of land. Taiwan, Japan, and Korea are all land-poor and differ in their educational mixes and capital per worker. Turkey is scarce in everything except low-educated labor, while Norway is the opposite, abundant only in capital and highly-educated workers. Many European countries have fairly similar relative factor supplies, as can be seen by the similar shapes of their endowment stars. Figure 3 vividly illustrates the range of endowment differences in our sample, and these differences are linked to the specialization differences seen in Figure 2 by the econometric analysis that follows.

---

[10] The other seven sectors are omitted from Figure 1 to save space. The five sectors shown are illustrative of patterns seen in the other sectors.

[11] Our comparison year in Figure 2 is 1985, rather than 1980 as in the rest of the data displays, because we wanted to include Japan, which has missing output data for 1980, in the figure.

[12] Australia is excluded from this Figure, because its huge land endowment obscures variation in land among the other countries.

3 Econometric Specification and Estimation

The data review above, along with the theory behind equations (11) and (12), guide our choice of estimation technique. For each of the twelve industries, the dependent variable is the share of a sector's value added in national income (expressed as a percentage), and the data is an unbalanced panel of countries and years. The regressors for equation (11) are the logs of relative factor supplies, which include capital, two types of labor, and land[13]. For equation (12), the only regressor is the log of aggregate productivity, measured as total real GDP per worker. The data review suggests at least three issues that need to be addressed: trends, measurement error, and country effects.

The models of equations (11) and (12) include time effects, and strong trends are clearly evident in the data. Although inclusion of time fixed effects in our estimators is feasible, we include linear time trends instead, for parsimony and for ease of reporting. This choice has no noticeable impact on any of our inferences on other parameters. We also allow the error term ε_{ict} to follow a stationary AR(1) process, with a common AR(1) parameter across countries for each industry:

$$\varepsilon_{ict} = \rho_i \varepsilon_{ic,t-1} + v_{ict}$$

where v_{ict} is white noise.

There is no question that all of our explanatory variables are measured with substantial error. The largest amount of measurement error is surely in the cross-country dimension: land quality differs across countries but is relatively stable over time, differences in educational systems are large across countries but change only slowly within countries, and so on. If these country-specific, time-invariant measurement errors are multiplicative in levels (and therefore additive in logs), they will be absorbed into the country effects in our estimating equations. The remaining measurement error is surely important. A solution to this statistical problem requires the use of an instrumental variables estimator, but unfortunately no plausible instruments are available. This limitation should be kept in mind in reviewing our econometric results below.

The theory, the exploratory data analysis, and a consideration of measurement error all

---

[13] With four factor supplies and the imposition of homogeneity, there are three relative factor supplies. Since the normalization is irrelevant, we report four separate factor supply effects in Table 5, but the coefficients sum to zero by construction, and the covariance matrix is singular.

argue for the inclusion of country effects in our regression models. A consistent estimator under general conditions is the within or fixed effects estimator, which uses only time-series variation within countries to identify the parameters of interest. However, a key feature of our sample is that most of the variation is across countries, and in addition, a major objective of the theoretical models is to explain specialization across countries at a point in time. We therefore want an estimator that takes account of country effects but also uses at least some of the cross-sectional variation in the data, and to this end, we develop a random effects estimator. The identifying assumption that justifies the random effects estimator is that the random country effects are orthogonal to the explanatory variables. This orthogonality assumption is a strong one, but we believe that the benefits of using the cross-country variation in the data justify the discomfort caused by making this assumption. Below, we report both the fixed and random effects estimators, so that the impact of the random effects assumption on inference can be assessed by the reader. We also report estimates from a between estimator, which uses only the cross-sectional variation in the data. Details of our estimators are given in the appendix.

One of our purposes in this paper is to compare the statistical performance of the two competing models of specialization, the factor proportions and ladder-of-development models. These models are not nested, so we compute two test statistics that are designed to discriminate between non-nested models. The Akaike criterion is given by

$$AIC_k = \ln L_k - p_k$$

where $\ln L_k$ is the maximized value of the log likelihood of model k, and p_k is the number of parameters. An alternative is the Schwartz criterion,

$$Schwartz_k = \ln L_k - \frac{p_k \ln T}{2}$$

where T is the sample size. The model choice rule for both statistics is to choose the model with the larger criteria value. Both models reward goodness of fit and penalize "complexity" as measured by the number of parameters. The complexity penalty given by the Schwartz criterion, which is derived from Bayesian principles, is more severe for the sample sizes we consider.

4 Results

Each of our 12 industry equations includes 23 to 28 countries, with 6 to 23 annual observations per country. The average number of years per country is roughly 21, with overall sample sizes per industry of 472 to 590 country-year observations.

Table 4 shows results for the fixed effects (FE), random effects (RE), and between estimators of the factor proportions model given by equation (11), with statistically significant coefficients shown in **bold** and t-statistics in *italics*[14]. The parameters on relative factor supplies have the interpretation of semi-elasticities: for example, a one percent increase in the supply of low-educated labor increases the share of food in GDP by 0.194 percentage points, according to the fixed effects estimates. It may seem odd that the effect of land, which has essentially no time-series variation, is identified in the fixed effects estimates. The identification is an artifact of homogeneity, because the coefficient on land is constrained to equal minus the sum of the other three factor supply effects, which do have substantial time series variation. For the FE estimates, then, the land effect may be better thought of as the combined effect of all time-invariant factor supplies, or alternatively, as an indicator of non-homogeneity in labor and capital.

In the food sector, which includes processed food and beverages but not raw agricultural output, capital abundance reduces output, while labor abundance (especially skilled) raises it. Similar patterns are visible in the wood-paper and oil-coal sectors, at least for the RE estimates. Land has a statistically significant effect on food output in the between regressions, but a negative effect in the FE and RE results.

The apparel-textile sector exhibits some informative disagreement between the FE and between estimates. The between estimates generally confirm our intuition about this sector: countries that are land and capital scarce and abundant in labor, specialize in apparel-textiles. In contrast, the within or FE estimates show that increases in skilled labor over time reduce output. The RE estimator offers an informative compromise: unskilled labor abundance and scarcity in physical and human capital is what accounts for specialization in apparel-textiles. The large positive effect of land in the RE estimates is due to the positive FE estimate of land, which is itself a consequence of the large negative time series effects of both physical and human capital.

---

[14]In discussing the results, we will adopt the convention that a parameter is statistically different from zero if the absolute value of the t-statistic is at least 1.62, that is, the approximate 10% critical value.

Five of our small sectors can loosely be thought of as natural resource based (wood-paper, printing, oil-coal, glass-stone-clay, and primary metals). Perhaps not surprisingly, given that our empirical model does not include measures of resource supplies, the model does poorly in explaining output in these sectors - most coefficients are not statistically significant. The chemicals sector is also very poorly explained by the factor endowments model; according to the RE estimates, abundance in labor and scarcity in land is reliably associated with higher output, but no other inferences are possible.

The remaining four sectors, which include fabricated metals and the three machinery sectors, have much stronger results: capital abundance and land scarcity lead to greater output in all of these sectors. The FE estimates tend to be more precisely estimated than the between estimates, perhaps because of the importance of country-effects in these sectors. In the electrical and non-electrical machinery sectors there, is a strong positive effect of higher educated workers, which is mirrored by a negative or insignificant effect of unskilled labor. In transportation equipment, the strongest inference is that capital abundance is key; the labor effects, in contrast, are imprecisely estimated.

Overall, the factor proportions model gives a noisy but fairly consistent story about industrial specialization: human and physical capital abundance raise output in the heavy industrial sectors, while physical capital lowers output in food and apparel-textiles. The model has little success in explaining variation in output in the smaller, more resource-based sectors, probably because we have no measurements of resource abundance. It is worth comparing our results here to those of Harrigan (1997), which had a smaller sample and did not use the cross-country variation in the data, but did have direct measures of technology. In that paper, Harrigan found that technology and endowments were both important in explaining specialization among OECD countries. To the extent that technological sophistication and endowments are not orthogonal (it would be a surprise if they were), it is not surprising that we find it harder to estimate the effects of endowments than did Harrigan (1997), because we do not measure technology and include countries at widely different levels of development.

Turning to the simple ladder-of-development model of equation (12), Table 5 also presents FE, RE, and between estimates. The results are roughly in line with what we would have expected from the factor proportions results: higher aggregate productivity is associated with

lower output of food and higher output in the heavy industrial sectors (fabricated metals and the three machinery categories). In addition, the RE estimate for chemicals is positive and significant as is the between estimate for printing. As with the factor proportions model, the within and between effects are opposite in sign in the apparel-textiles sector. Aggregate productivity also has mixed success in explaining output in the more resource-based sectors, with the notable exception of primary metals.

Testing for statistical significance is important but begs the question: are the estimates economically important? To address this issue, Table 6 reports standardized coefficients. These are calculated by multiplying the regression coefficient from Tables 4 and 5 by the ratio of sample standard deviations of the right and left hand side variables, so that a standardized coefficient answers the question "by how many standard deviations does an output share change with a one standard deviation increase in an explanatory variable?". Standardized coefficients corresponding to statistically significant estimates from Tables 4 and 5 are in **bold**.

In the food and apparel-textiles sectors, the effects of high educated labor and capital are quite large: for example, a one standard deviation increase in capital lowers each sector by around half of a standard deviation according to the RE estimates. For the fabricated metals and machinery sectors, the statistically significant standardized effects are in the range of 0.1 to 0.4. This implies that the effects of endowments are economically important, but relatively small compared to the sample variation in output shares. For the other sectors (wood-paper, printing, chemicals, oil-coal, and glass-stone-clay), the standardized coefficients are so small as to be economically of second-order importance. In other words, country effects and noise are more important than measured endowments for these sectors. By contrast, the ladder-of-development model has much larger standardized effects in the sectors where aggregate productivity is statistically significant, in most cases exceeding one in absolute value.

The results of Tables 4 through 6 indicate that both the factor proportions model and the simple development model have some success in explaining the data. Table 7 reports specification test statistics, with the "winning" model indicated in **bold**. For each industry, the test statistics within a given row are comparable when the sample size is the same; that is, the estimates that use the within variation are comparable to each other, while they cannot be directly compared to the between estimates. Generally, the FE estimates of the development model are

preferred among the estimates that use the within variation; in a few cases the Schwarz criteria chooses the RE over the FE estimates, but in only one case, apparel-textiles, does either criteria choose the factor endowment model over the more parsimonious development model. The verdict is more mixed for the between estimates: in 4 of the 12 sectors (including food, apparel-textiles, and electrical equipment), both criteria choose the factor endowment model, while in the remainder of the sectors the "winner" is the development model. Overall, then, the development model seems to be preferred on these purely statistical grounds.

A final specification issue concerns our working assumption that the translog functional form is an adequate approximation to the true revenue function. We investigated this issue informally, looking for break points in the sample where factor endowment effects changed sign, to no avail. We also estimated an augmented version of equation (12), including the square as well as the level of log real GDP per worker as an explanatory variable. There was not a single statistically significant effect in the between estimates, but we did find statistically significant quadratic terms in five cases with the FE estimator. The implied nonlinear relationship between output shares and GDP per worker for these five sectors is plotted in Figure 4. The figure indicates that with the possible exception of the food sector, with a peak at around $10,000 per worker, the nonlinearities are not economically very important, and the relationship between output shares and log GDP per worker can be well-approximated by a linear model. In brief, if there *are* diversification cone effects in the world economy, we cannot find them in our data.

5 Conclusions

This paper uses a general version of the factor proportions model to organize a study of the relationship between specialization and relative factor supplies. Using a panel of 28 countries over 23 years, we report estimates using both the cross-section and time series variation in the data. We also estimate a simple ladder-of-development model, which predicts specialization by aggregate productivity alone.

The results show that factor endowments *do* help predict specialization, particularly in large industrial sectors that are not natural-resource based. These effects are economically large as well as statistically significant. The ladder-of-development model is also useful in helping to understand the data and is preferred on statistical grounds. The statistical analysis is motivated by and supplemented with an extensive graphical display of the data, which is also suggestive of the importance of factor supplies in explaining specialization.

The data analysis in this paper suggests a simple, if not very tidy, conclusion. Relative factor endowments have a large influence on specialization, in ways that are consistent with theory and stylized facts about the international economy. However, factor endowments leave much that is unexplained: there is a great degree of country-specific idiosyncracy in specialization patterns, and there is also a great deal of noise. A fuller account of specialization will probably include roles for history, geography, technology, and economic policy, but such an account will definitely include a role for relative factor supplies.

References

Balassa, Bela, 1965, "Trade Liberalization and 'Revealed' Comparative Advantage", *Manchester School* 33 (May), 99-123.

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

Bernstein, Jeffrey, and David E. Weinstein, 1998, "Do Endowments Predict the Location of Production? Evidence from National and International Data", NBER Working paper no. 6815 (November).

Caves, Douglas W., Laurits R. Christensen, and W. Erwin Diewert, 1982, "The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity", *Econometrica* 50(6): 1393-1414.

Davis, Donald R., David E. Weinstein, Scott C. Bradford, and Kazushige Shimpo, 1997, "Using International and Japanese Regional Data to Determine When the Factor Abundance Theory of Trade Works", *American Economic Review* 87(3): 421-46.

Davis, Donald R., and David E. Weinstein, 1999, "Economic geography and regional production structure: An empirical investigation", *European Economic Review* 43(2): 379-407.

Dixit, Avinash, and Victor Norman, 1980, *The Theory of International Trade*, Cambridge, England: Cambridge University Press.

Gabaix, Xavier, 1998, "The Factor Content of Trade: A Rejection of the Heckscher-Ohlin-Leontief Hypothesis", Harvard University, mimeo.

Harrigan, James, 1995, "Factor Endowments & The International Location of Production: Econometric Evidence for the OECD, 1970-1985", *The Journal of International Economics*, v. 39 no. 1/2 (August): 123-141.

Harrigan, James, 1997, "Technology, Factor Supplies and International Specialization: Estimating the Neoclassical Model", *The American Economic Review* 87(4): 475-494 (September).

Harrigan, James, 1999, "Estimation of Cross-Country Differences in Industry Production Functions", *The Journal of International Economics* 47(2): 267-293 (April).

Jones, Ronald and José Scheinkman, 1977, "The Relevance of the Two-Sector Production Model in Trade Theory", *Journal of Political Economy*, 85(5): 909-935.

Kohli, Ulrich, 1991, *Technology, Duality, and Foreign Trade*, Ann Arbor: University of
Michigan Press.

Leamer, Edward E., 1987, "Paths of Development in the Three-Factor, *n*-Good General
Equilibrium Model", *Journal of Political Economy* 95 (October): 961-999.

Leontief, Wassily, 1954, "Domestic Production and Foreign Trade: The American Capital
Position Re-examined", Chapter 30 in Richard E. Caves and Harry G. Johnson, eds.,
Readings in International Economics, 1968, London: Allen and Unwin.

Schott, Peter, 1999, "One Size Fits All? Theory, Evidence, and Implications of Cones of
Diversification", mimeo, UCLA.

Summers, Robert, and Alan Heston, 1991, "The Penn World Table (Mark V): An Expanded Set
of International Comparisons, 1950-1988", *Quarterly Journal of Economics*, 106(2):
327-368.

Trefler, Daniel, 1993, "International Factor Price Differences: Leontief was Right!", *Journal of
Political Economy*, 101(6): 961-987.

Trefler, Daniel, 1995, "The Case of the Missing Trade and Other Mysteries", *American
Economic Review*, 85(5): 1029-1046.

Woodland, Alan D., 1982, *International Trade and Resource Allocation*, Amsterdam: North
Holland.

Table 1 - Data Set Description

<u>Years</u> 1970-1992

<u>Countries</u>

21 OECD   Australia (AUS), Austria (AUT), Belgium (BEL), Canada (CAN), Denmark (DEN), Finland (FIN), France (FRA), Germany (GER), Greece (GRC), Iceland (ICE), Ireland (IRL), Italy (ITA), Japan (JPN), Netherlands (NLD), New Zealand (NZL), Norway (NOR), Portugal (PRT), Spain (SPN), Sweden (SWE), the United Kingdom (UK), and the United States (US).

7 non-OECD Argentina (ARG), Chile (CHL), Hong Kong (HKG), Korea (KOR), Mexico (MEX), Turkey (TUR), Taiwan (TWN). None of these countries were in the OECD in 1970.

<u>Product Classification System:</u> Twelve categories based on the International Standard Industrial Classification (ISIC).

Food	311/2	Food manufacturing
	313	Beverage industries
	314	Tobacco manufactures
Textiles & Apparel	321	Manufacture of textiles
	322	Manufacture of wearing apparel except footwear
Wood & Paper	331	Wood products, except furniture
	332	Furniture, except metal
	341	Manufacture of paper and paper products
Printing	342	Printing, publishing, and allied industries
Chemicals	351	Manufacture of industrial chemicals
	352	Manufacture of other chemical products
Oil & Coal	353	Products of petroleum refineries
	354	Miscellaneous products of petroleum and coal
Glass, Stone & Clay	361	Pottery, china, and earthenware
	362	Glass and glass products
	369	Other non-metallic mineral products
Primary Metals	371	Iron and steel basic industries
	372	Non-ferrous metal basic industries
Fabricated Metals	381	Fabricated metal products, except machinery and equipment
Machinery ex-Electric	382	Manufacture of machinery, except electrical
Electrical Machinery	383	Electrical machinery, apparatus, appliances and supplies
	385	Professional, scientific, measuring and control equipment
Transport Equipment	384	Transportation equipment, including motor vehicles, ships, and aircraft

Shares of each industry in GDP

        Source: <u>The Industrial Statistics Yearbook 1998, 3 Digit Level of ISIC Code</u> from the United Nations Industrial Development Organization (UNIDO).

.

<u>Factor Endowments</u>

Capital   From version 5.6 of the Penn-World Table (PWT 5.6) Units: millions of 1985 international dollars. See Robert Heston and Alan Summers (1991) for details.

Labor   The economically active population (from PWT 5.6) is classified according to education level: 1) low, workers with at most primary education, 2) high, workers with at least some secondary education. Units: Thousands of workers. The educational classification for 1970, 1975, 1980, and 1985 comes from Barro and Lee (1993); intervening years are interpolated. The original data set has been updated to include the years 1990 and 1995, using the same equations and data sources used by Barro and Lee (1993).

Land   Agricultural land. Units: thousands of hectares. The source is the World Bank's World tables on CD-ROM.

<u>Aggregate GDP per Worker</u>

        From version 5.6 of the Penn-World Table (PWT 5.6) Units: thousands of 1985 international dollars per worker. See Robert Heston and Alan Summers (1991) for details.

Table 2 - 1980 Industry Output Shares Relative to the Mean Across Countries

	Food and Beverage	Textile and Apparel	Wood and Paper	Printing	Chemicals	Oil and Coal	Glass, Stone & Clay	Primary Metals	Fabricated Metals	Non-Electrical Machinery	Electrical Machinery	Transport Equipment
ARG	183	127	73	80	171	491	99	96	123	72	56	139
AUS	86	45	76	113	64	21	79	125	104	70	47	96
AUT	89	65	93	80	73	14	124	115	112	116	102	50
BEL	110	69	87	78	124	47	92	139	118	114	96	87
CAN	86	56	194	115	83	59	62	108	113	81	78	122
CHL	169	62	115	84	84	92	79	315	57	24	20	32
DEN	116	33	65	126	82	17	96	21	84	132	67	53
FIN	89	73	363	209	84	90	97	79	99	155	72	87
FRA	86	63	74	99	99	193	107	82	123	132	114	145
GER	108	57	99	76	130	183	124	156	121	229	208	211
GRC	75	122	41	38	59	43	112	65	61	17	34	61
HKG	33	400	40	101	—	—	21	14	151	36	217	34
ICE	—	43	76	102	—	—	69	78	—	—	—	—
IRL	227	80	67	131	—	—	196	14	112	121	115	51
ITA	50	85	64	66	109	28	126	126	85	111	106	124
JPN	77	75	110	160	131	68	138	188	143	200	192	165
KOR	139	216	63	68	150	143	151	141	67	57	129	99
MEX	40	—	20	7	—	—	62	64	30	16	61	55
NLD	87	25	56	143	88	35	63	43	96	74	104	61
NZL	150	70	143	130	63	15	65	46	111	56	52	77
NOR	59	19	103	105	51	23	49	105	63	80	42	86
PRT	71	148	130	63	—	—	138	50	76	32	53	82
SPN	106	87	94	71	101	73	142	117	118	92	84	122
SWE	69	25	216	145	85	37	72	96	139	169	107	158
TWN	121	205	115	52	124	231	141	137	95	56	219	116
TUR	79	90	26	14	76	216	92	92	39	40	31	43
UK	112	63	96	181	139	92	129	92	127	214	148	177
US	82	61	102	163	131	90	76	98	132	205	174	164

Notes to Table 2: This table shows each country's output shares in 1980, divided by the cross-country mean, and multiplied by 100.

Table 3 - Factor Endowment Levels Relative to the Mean, 1980

	Workers 1000's	Capital $/worker	Education level, % of total		Cropland & Pasture sq.km /100 workers
			Low	High	
Argentina (83)	69	55	130	63	278
Australia	44	139	55	155	1240
Austria	22	114	107	91	19
Belgium	26	141	93	108	7
Canada	77	130	47	164	108
Chile	25	31	121	75	79
Denmark	18	127	90	112	19
Finland	16	151	94	108	20
France	155	131	127	68	23
Germany	176	196	138	54	12
Greece	24	89	133	60	43
Hong Kong	21	55	114	83	0
Iceland	1	59	111	86	337
Ireland	8	79	96	105	79
Italy	142	114	122	73	14
Japan	473	99	84	120	1
Korea	96	44	99	101	3
Mexico	141	63	159	29	80
Netherlands	36	130	79	126	6
New Zealand	9	127	24	192	233
Norway	13	185	3	216	8
Portugal	28	36	157	31	16
Spain	84	87	146	45	42
Sweden	27	129	76	129	15
Taiwan	48	61	123	72	3
Turkey	125	29	160	27	35
UK	177	74	97	103	12
USA	720	124	13	205	67

Notes to Table 3: This table shows each country's factor endowments in 1980, divided by the cross-country mean, and multiplied by 100. Data for Argentina is from 1983.
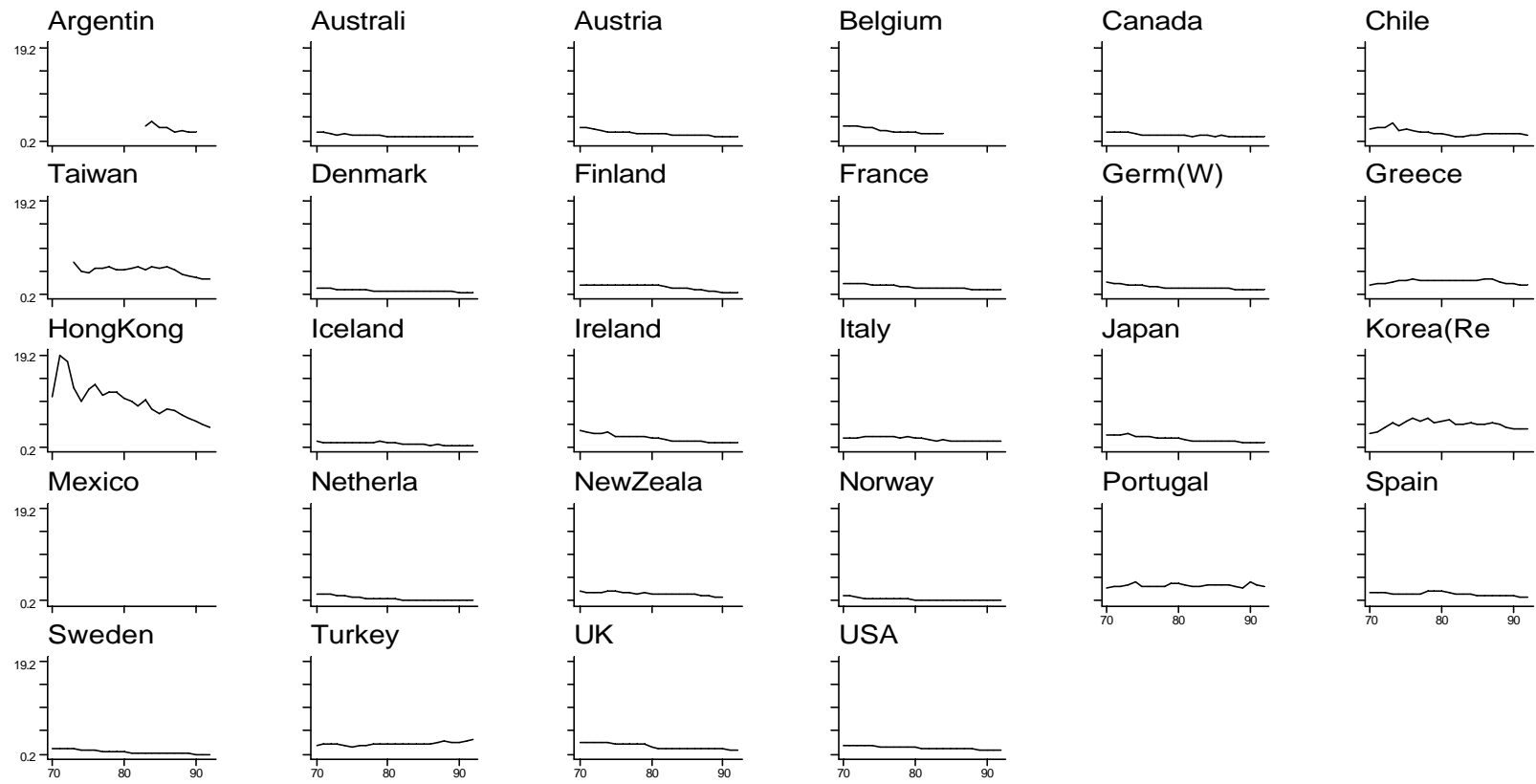
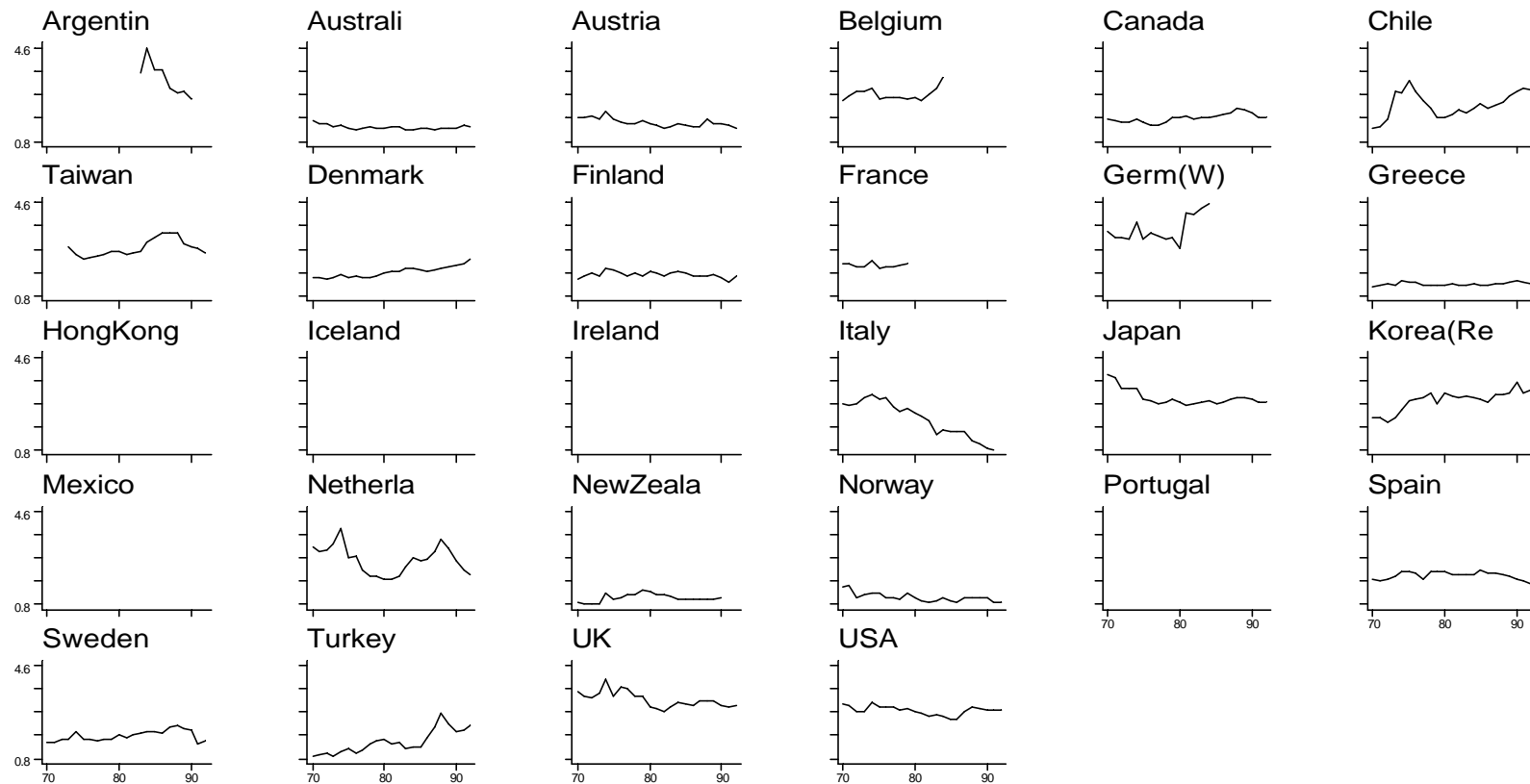Figures 1: Output Shares of GDP - By Industry

Food and Beverages
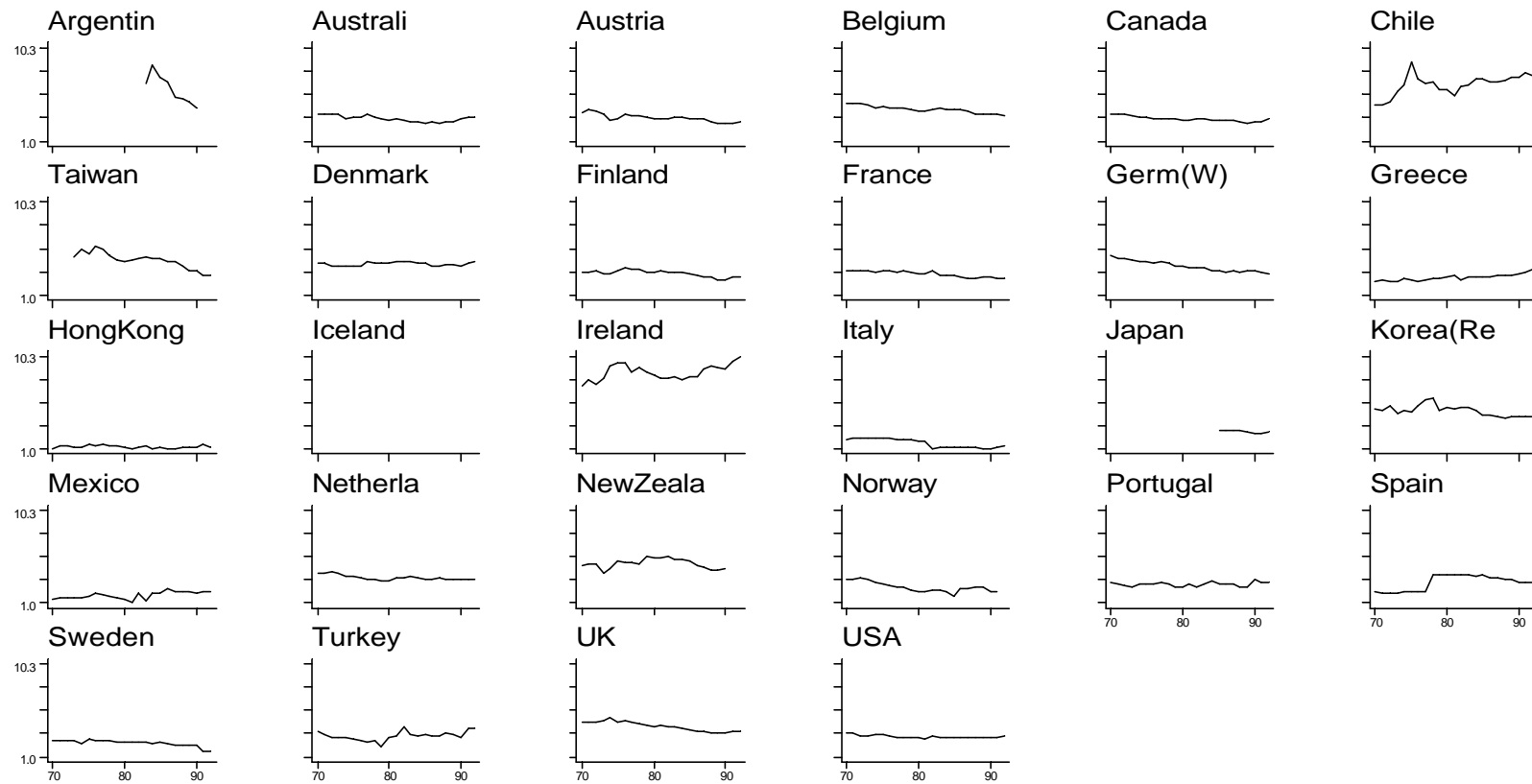
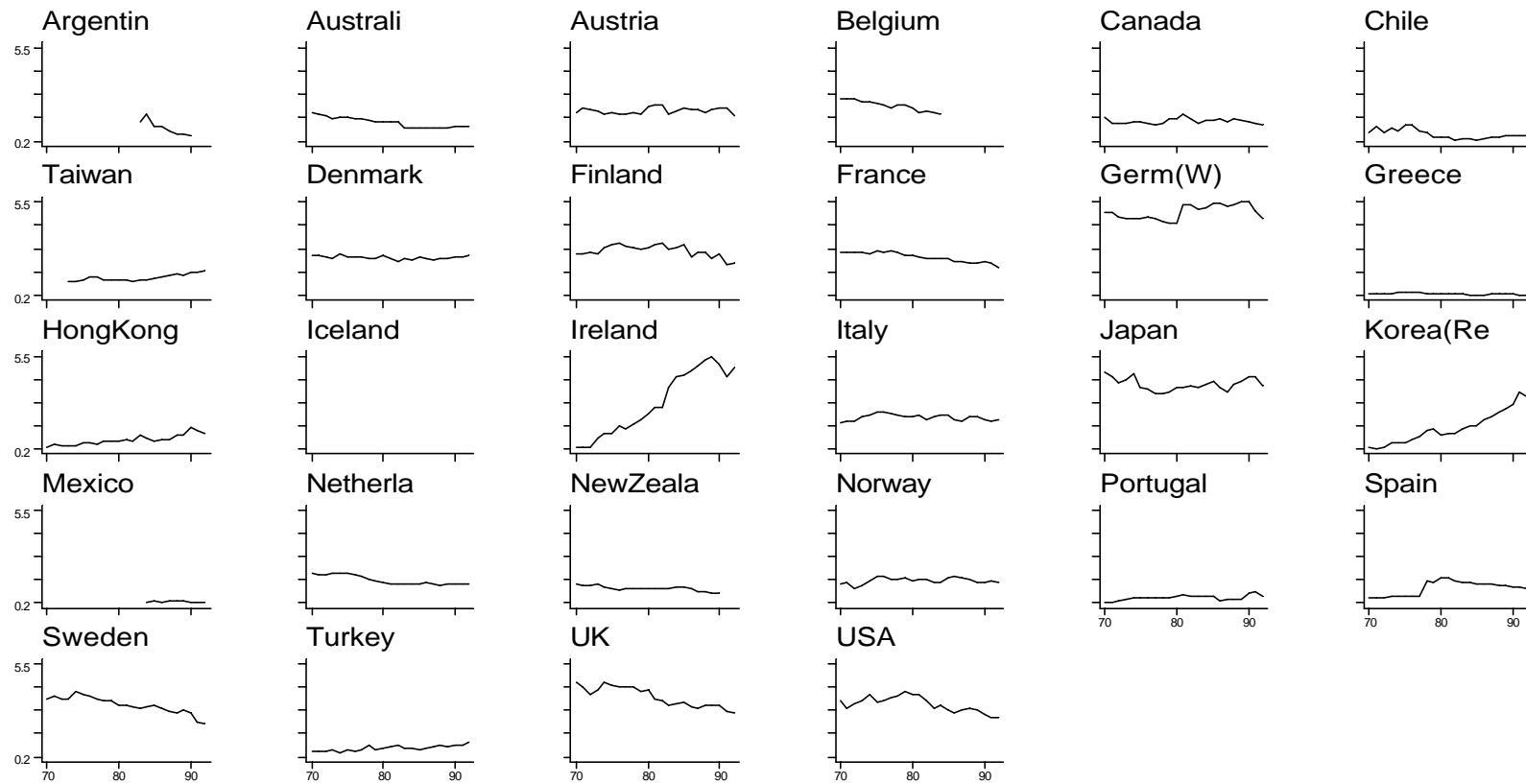Figures 1: Output Shares of GDP - By Industry

Textiles and Apparel

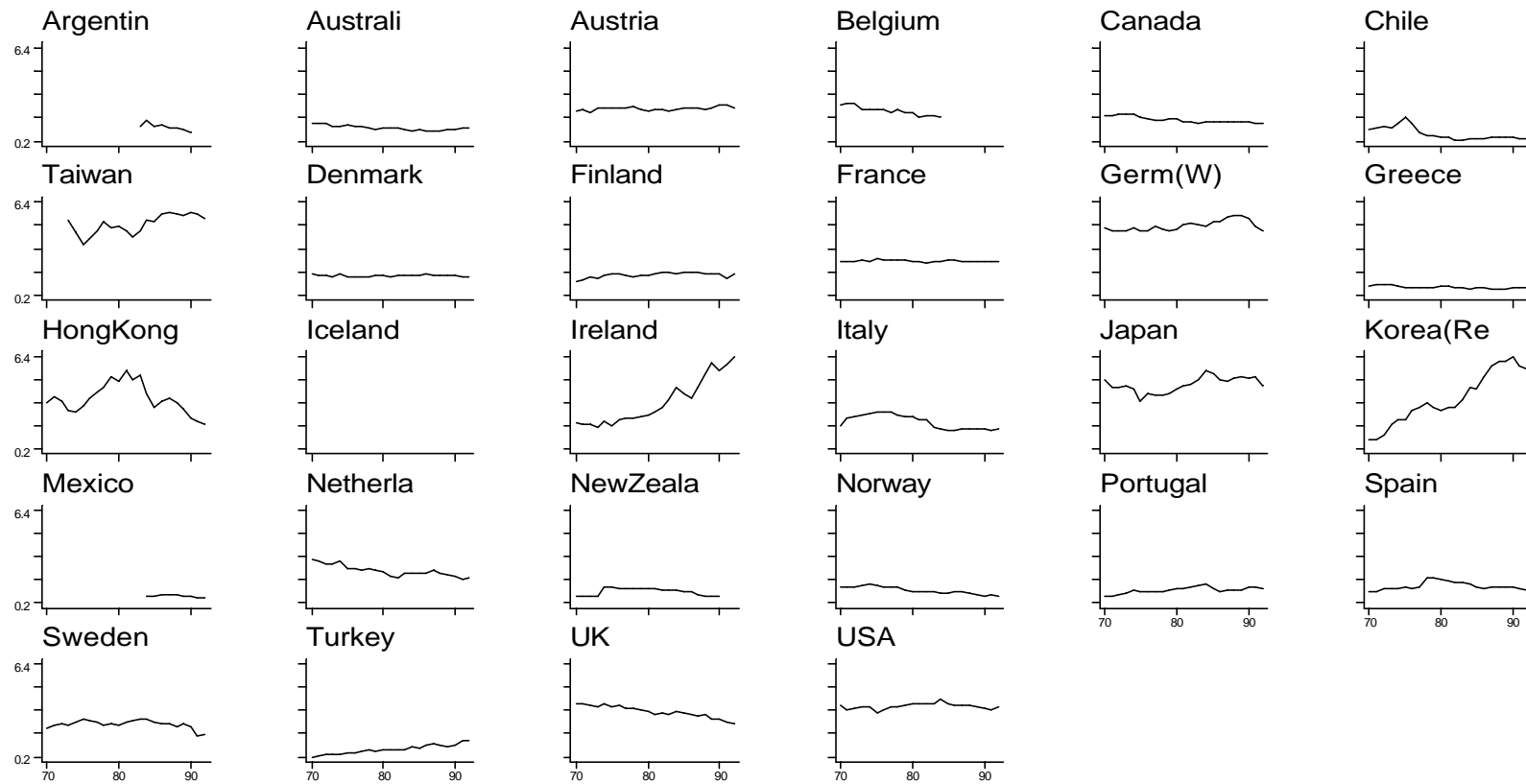Figures 1: Output Shares of GDP - By Industry

Chemicals

Figures 1: Output Shares of GDP - By Industry
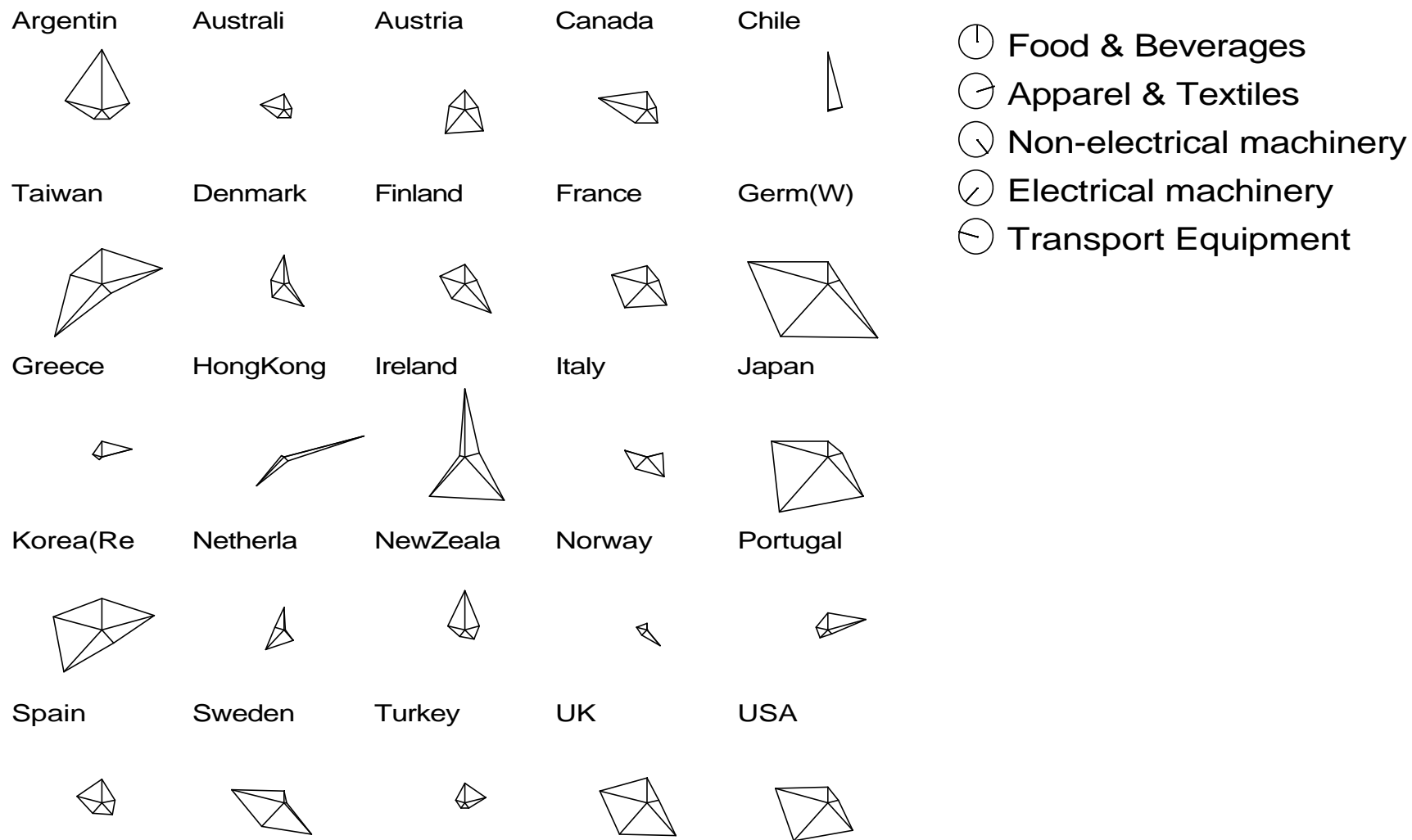
Non-Electrical Machinery

Figures 1: Output Shares of GDP - By Industry
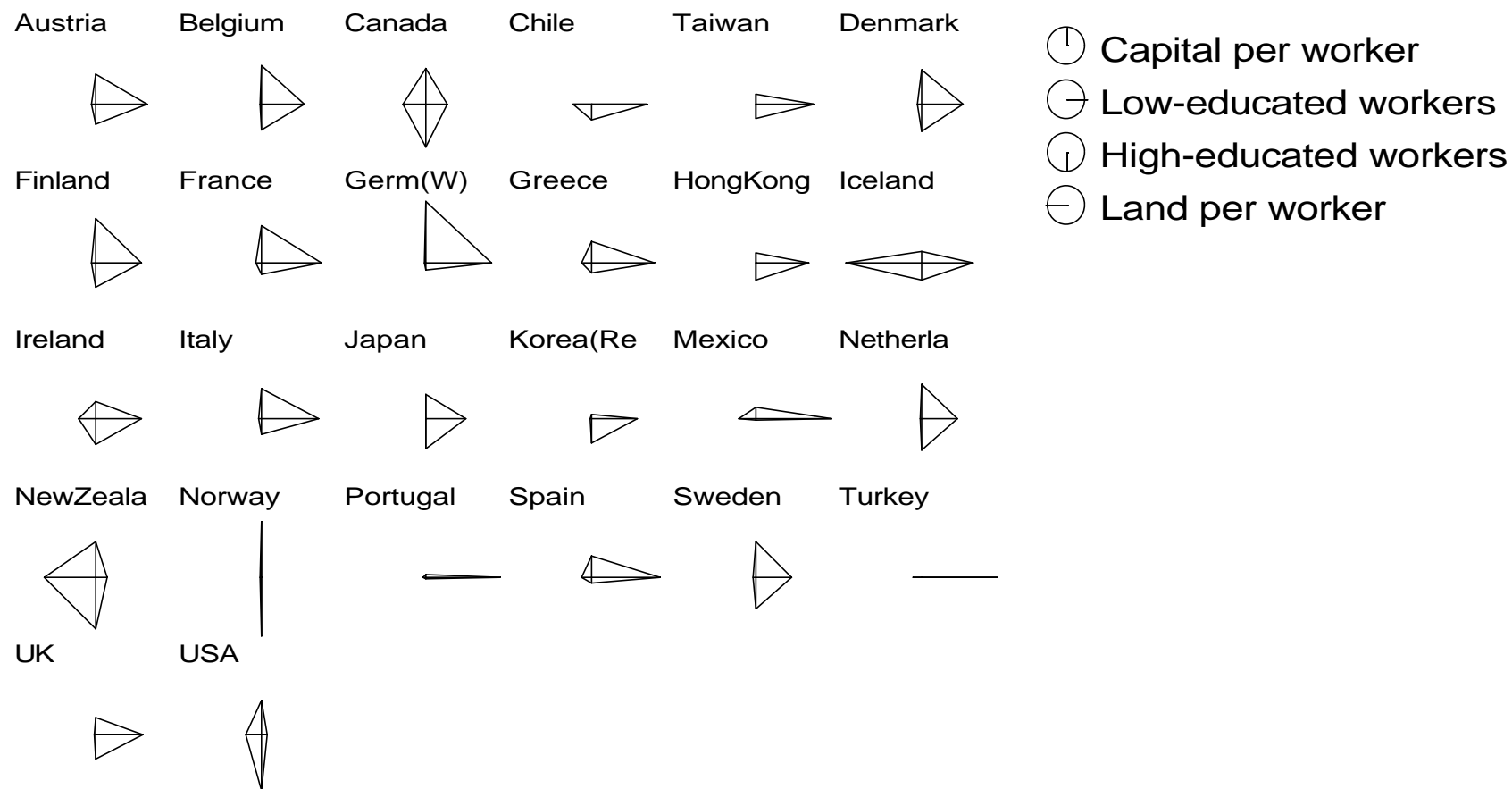
Electrical Machinery & Instruments

Figure 2 - Relative Output Shares, 1985

Notes to Figure 2: For a country, each of the five vectors is proportional to the country's output share in that sector relative to the average country's output in that sector.

Figure 3 - Relative Endowments, 1980

Austria   Belgium   Canada   Chile   Taiwan   Denmark

⊙ Capital per worker
⊖ Low-educated workers
⊙ High-educated workers
⊖ Land per worker

Finland   France   Germ(W)   Greece   HongKong   Iceland

Ireland   Italy   Japan   Korea(Re   Mexico   Netherla

NewZeala   Norway   Portugal   Spain   Sweden   Turkey

UK   USA

Notes to Figure 3: For a country, each of the four vectors is proportional to the country's endowment of that factor relative to the average country's endowment of that factor.

31

Table 4 - Regression Results, Factor Endowment Model

	Fixed Eff.	Random Eff.	Between	Fixed Eff.	Random Eff.	Between	Fixed Eff.	Random Eff.	Between
	Food and Beverages			Textiles and Apparel			Wood and Paper		
Low Ed	**0.194**	**0.165**	0.136	0.013	**0.019**	**0.860**	**0.181**	**0.112**	**-0.346**
	3.86	*4.62*	*0.69*	*0.93*	*1.64*	*4.24*	*2.54*	*1.94*	*-2.25*
High Ed	**0.395**	**0.468**	**1.325**	**-0.229**	**-0.215**	**1.380**	**0.207**	**0.176**	-0.141
	2.22	*5.04*	*2.04*	*-2.80*	*-3.79*	*2.63*	*2.18*	*3.29*	*-0.36*
Capital	-0.272	**-0.456**	**-1.751**	**-0.478**	**-0.304**	**-1.658**	-0.048	**-0.142**	0.495
	-1.28	*-4.21*	*-2.45*	*-3.50*	*-4.51*	*-3.41*	*-0.43*	*-2.1*	*1.06*
Land	**-0.318**	**-0.178**	0.289	**0.694**	**0.500**	**-0.582**	**-0.340**	**-0.146**	-0.008
	-2.05	*-6.13*	*1.69*	*5.14*	*16.81*	*-2.31*	*-1.88*	*-10.46*	*-0.12*
Year	**-0.027**	**-0.022**		**-0.024**	**-0.031**		**-0.027**	**-0.021**	
	-2.87	*-4.01*		*-4.94*	*-15.69*		*-3.92*	*-6.81*	
	Printing			Chemicals			Oil and Coal		
Low Ed	**0.043**	0.015	**-0.247**	0.037	**0.090**	0.237	-0.002	**0.027**	**0.493**
	2.38	*0.94*	*-3.07*	*0.76*	*2.67*	*1.55*	*-0.11*	*2.03*	*2.98*
High Ed	**-0.073**	-0.025	**0.239**	0.044	**0.181**	0.265	0.038	**0.119**	0.120
	-1.69	*-0.86*	*1.78*	*0.24*	*2.28*	*0.73*	*0.39*	*2.74*	*0.34*
Capital	-0.025	0.007	0.033	0.047	0.018	-0.310	-0.083	**-0.190**	**-0.710**
	-0.76	*0.28*	*0.21*	*0.18*	*0.15*	*-0.81*	*-0.92*	*-3.59*	*-1.71*
Land	0.055	0.003	-0.025	0.127	**-0.289**	**-0.192**	0.048	**0.045**	0.097
	0.84	*0.33*	*-0.98*	*0.64*	*-14.16*	*-2.75*	*0.37*	*4.67*	*0.74*
Year	0.004	0.001		-0.002	-0.006		-0.003	-0.001	
	1.30	*0.33*		*-0.16*	*-1.36*		*-0.54*	*-0.30*	

Notes to Table 4 This table presents fixed effects, random effects, and between estimates of the factor proportions model. T-statistics are in italics. The dependent variable is output as a percentage of GDP, and the explanatory variables are log relative factor supplies. The coefficients on the four factor supplies sum to zero by construction. See text for details.

	Fixed Eff.	Random Eff.	Between	Fixed Eff.	Random Eff.	Between	Fixed Eff.	Random Eff.	Between
	Glass, Stone, & Clay			Primary Metals			Fabricated Metals		
Low Ed	**0.039**	**0.046**	**0.119**	0.067	-0.001	0.152	**0.056**	-0.010	-0.119
	3.13	3.83	2.39	0.67	-0.04	0.66	3.29	-0.26	-0.99
High Ed	**0.258**	**0.170**	-0.224	-0.070	0.003	0.583	**0.288**	0.097	-0.015
	2.70	3.65	-1.44	-0.64	0.04	0.74	2.29	1.25	-0.10
Capital	0.055	-0.073	0.098	-0.056	-0.026	-0.793	**0.371**	0.054	0.212
	0.48	-1.27	0.61	-0.45	-0.31	-0.76	2.44	0.92	1.10
Land	**-0.351**	**-0.142**	0.007	0.059	0.024	0.058	**-0.715**	**-0.140**	**-0.078**
	-5.23	-23.40	0.11	1.22	0.64	0.39	-5.11	-6.30	-2.20
Year	**-0.030**	**-0.021**		0.008	**-0.013**		**-0.715**	**-0.023**	
	-8.42	-8.81		-1.37	-3.27		-3.80	-6.56	
	Non-Electrical Machinery			Electrical Machinery			Transport Equipment		
Low Ed	**-0.103**	**-0.128**	-0.333	**0.056**	0.014	0.012	-0.074	**-0.057**	-0.220
	-2.91	-6.76	-1.43	3.62	0.44	0.04	-0.80	-2.98	-1.32
High Ed	**0.272**	**0.234**	-0.561	**0.316**	**0.254**	0.310	-0.032	-0.102	-0.330
	2.78	2.70	-1.08	4.05	4.56	0.54	-0.23	-1.17	-0.74
Capital	**0.279**	**0.310**	**1.053**	**0.386**	**0.224**	0.100	0.334	**0.281**	0.600
	2.95	3.89	1.95	2.17	3.73	0.16	1.55	3.49	1.30
Land	**-0.448**	**-0.416**	-0.160	**-0.758**	**-0.493**	**-0.421**	-0.227	**-0.123**	-0.050
	-3.72	-13.60	-1.30	-4.26	-18.47	-4.17	-0.90	-4.85	-0.51
Year	**-0.035**	**-0.035**		**-0.037**	**-0.028**		**-0.034**	**-0.028**	
	-5.13	-9.45		-4.82	-10.35		-2.98	-6.29	

Table 5 - Regression Results, GDP per Worker Model

	Fixed Eff.	Random Eff.	Between	Fixed Eff.	Random Eff.	Between	Fixed Eff.	Random Eff.	Between
	Food and Beverages			Textiles and Apparel			Wood and Paper		
Y/W	**-1.074**	**-1.098**	-0.727	**0.461**	**0.249**	**-2.511**	**0.738**	**0.634**	**0.725**
	-3.21	*-6.78*	*-1.17*	*2.54*	*1.81*	*-2.69*	*4.24*	*6.55*	*2.12*
Year	**-0.010**	**-0.009**		**-0.058**	**-0.055**		**-0.033**	**-0.031**	
	-1.91	*-2.43*		*-26.67*	*-25.60*		*-8.21*	*-11.04*	
	Printing			Chemicals			Oil and Coal		
Y/W	-0.014	-0.006	**0.635**	0.304	**0.201**	-0.166	**-0.336**	**-0.335**	**-1.409**
	-0.25	*0.11*	*5.59*	*1.48*	*2.66*	*-0.46*	*-3.66*	*-8.42*	*-5.17*
Year	-0.000	-0.001		-0.003	-0.001		0.001	0.001	
	-0.41	*-0.63*		*-0.6*	*-0.36*		*0.29*	*0.37*	
	Glass, Stone & Clay			Primary Metals			Fabricated Metals		
Y/W	**0.688**	**0.560**	-0.205	-0.021	-0.027	-0.632	**1.014**	**0.690**	**0.541**
	5.67	*6.87*	*-1.37*	*-0.20*	*-0.39*	*-0.87*	*4.39*	*12.10*	*4.01*
Year	**-0.029**	**-0.027**		**-0.013**	**-0.013**		**-0.033**	**-0.027**	
	-11.22	*-12.90*		*-2.84*	*-3.22*		*-5.88*	*-7.09*	
	Non-Electrical Machinery			Electrical Machinery			Transport Equipment		
Y/W	**0.648**	**0.783**	**1.292**	**0.584**	**0.648**	0.125	0.289	**0.403**	**0.764**
	2.99	*7.41*	*3.80*	*2.91*	*10.56*	*0.19*	*1.29*	*3.72*	*2.22*
Year	**-0.020**	**-0.023**		**-0.016**	**-0.017**		**-0.025**	**-0.026**	
	-3.72	*-5.84*		*-3.79*	*-4.82*		*-4.13*	*-5.30*	

<u>Notes to Table 5</u> This table presents fixed effects, random effects, and between estimates of the ladder-of-development model. T-statistics are in italics. The dependent variable is output as a percentage of GDP, and the explanatory variable Y/W is the log of aggregate GDP per worker. See text for details.

Table 6 - Standardized Parameter Estimates, Both Models

|	Fixed Eff.	Random Eff.	Between	Fixed Eff.	Random Eff.	Between	Fixed Eff.	Random Eff.	Between
	Food and Beverages			Textiles and Apparel			Wood and Paper		
Low Labor	**0.229**	**0.195**	0.161	0.016	**0.025**	**1.126**	**0.127**	**0.078**	**-0.242**
High Labor	**0.540**	**0.639**	**1.810**	**-0.317**	**-0.298**	**1.913**	**0.154**	**0.132**	-0.105
Capital	-0.356	**-0.597**	**-2.294**	**-0.639**	**-0.407**	**-2.216**	-0.034	**-0.102**	0.356
Land	**-0.240**	**-0.134**	**0.218**	**0.688**	**0.495**	**-0.576**	**-0.179**	**-0.077**	-0.004
Y/W	**-4.010**	**-4.102**	-2.715	**2.206**	**1.190**	**-12.018**	**1.897**	**1.629**	**1.864**
	Printing			Chemicals			Oil and Coal		
Low Labor	**0.013**	0.004	**-0.071**	0.020	**0.049**	0.128	-0.002	**0.018**	**0.325**
High Labor	**-0.023**	-0.008	**0.074**	0.028	**0.119**	0.174	0.030	**0.095**	0.096
Capital	-0.007	0.002	0.010	0.031	0.012	-0.203	-0.067	**-0.152**	**-0.568**
Land	0.012	0.001	-0.006	-0.058	**-0.131**	-0.087	0.027	**0.025**	0.054
Y/W	-0.015	0.006	**0.680**	0.537	**0.355**	-0.293	**-0.726**	**-0.723**	**-3.041**
	Glass, Stone & Clay			Primary Metals			Fabricated Metals		
Low Labor	**0.010**	**0.012**	**0.031**	0.058	-0.001	0.134	**0.020**	-0.004	-0.043
High Labor	**0.072**	**0.047**	-0.062	-0.068	0.003	0.572	**0.117**	0.040	-0.006
Capital	0.015	-0.020	0.026	-0.053	-0.024	-0.751	**0.147**	0.021	0.084
Land	**-0.068**	**-0.028**	0.001	0.037	0.015	0.037	**-0.172**	**-0.034**	**-0.019**
Y/W	**0.657**	**0.535**	-0.196	-0.066	-0.086	-1.992	**1.188**	**0.809**	**0.635**
	Non-Electrical Machinery			Electrical Machinery			Transport Equipment		
Low Labor	-0.087	**-0.109**	-0.283	**0.053**	0.013	0.011	-0.051	**-0.039**	-0.151
High Labor	**0.261**	**0.225**	-0.538	**0.341**	**0.274**	0.334	-0.025	-0.079	-0.256
Capital	**0.261**	**0.290**	**0.985**	**0.405**	**0.236**	0.105	0.253	**0.213**	0.454
Land	**-0.254**	**-0.236**	-0.091	**-0.484**	**-0.314**	**-0.269**	-0.105	**-0.056**	-0.023
Y/W	**1.790**	**2.162**	**3.569**	**1.812**	**2.011**	0.388	0.647	**0.901**	**1.707**

Notes to Table 6 This table presents coefficients from Tables 4 and 5, re-scaled to units of sample standard deviations. See text for details.

Table 7 - Specification Test Statistics

	FE		RE		Between		FE		RE		Between		FE		RE		Between	
	factors	Y/W	factors	Y/W	factors	Y/W	factors	Y/W	factors	Y/W	factors	Y/W	factors	Y/W	factors	Y/W	factors	Y/W
	Food						**Textiles and Apparel**						**Lumber & Wood**					
Akaike	8.3	**19.8**	-19.3	-9.5	**-2.33**	-2.50	**215.4**	214.2	186.6	191.4	**-2.20**	-2.65	228.9	**243.2**	207.4	223.2	-1.73	**-1.64**
Schwarz	-117.5	**-101.6**	-147.2	-133.1	**-2.53**	-2.60	89.1	**92.3**	58.1	67.3	**-2.39**	-2.75	97.5	**116.2**	73.8	94.0	-1.92	**-1.73**
p	58	56	59	57	4	2	58	56	59	57	4	2	60	58	61	59	4	2
N	565				27		575				27		590				28	
	Printing & Publishing						**Chemicals**						**Oil & Coal**					
Akaike	684.9	**691.0**	676.1	686.0	0.28	**0.33**	101.4	**109.0**	87.1	92.0	**-0.87**	-1.00	150.2	**157.1**	145.6	156.1	-1.38	**-1.17**
Schwarz	553.4	**563.9**	542.4	556.7	0.09	**0.24**	-2.5	**9.2**	-8.5	0.5	**-1.07**	-1.10	46.3	57.3	50.0	**64.6**	-1.58	**-1.27**
p	60	58	61	59	4	2	50	48	46	44	4	2	50	48	46	44	4	2
N	592				28		472				23		472				23	
	Stone, Clay & Glass						**Primary Metals**						**Fabricated Metals**					
Akaike	497.1	**523.8**	495.5	515.2	**0.44**	0.43	6.4	**9.9**	-14.2	-8.3	-2.06	**-1.96**	330.5	**346.3**	312.2	343.2	0.09	**0.31**
Schwarz	366.8	397.8	373.9	**397.9**	0.25	**0.33**	-124.6	**-116.8**	-136.5	-126.2	-2.25	**-2.05**	204.9	225.1	195.3	**230.6**	-0.10	**0.22**
p	60	58	56	54	4	2	60	58	56	54	4	2	58	56	54	52	4	2
N	569				28		583				28		561				27	
	Non-electrical Machinery						**Electrical Machinery**						**Transportation Equipment**					
Akaike	**227.1**	**227.1**	200.4	200.7	-1.79	**-1.66**	202.6	**209.6**	177.6	182.5	**-1.78**	-2.11	119.9	**129.9**	97.8	108.4	-1.46	**-1.30**
Schwarz	101.5	**105.9**	83.5	88.1	-1.98	**-1.75**	77.0	**88.4**	60.7	69.9	**-1.97**	-2.21	-5.7	**8.7**	-19.1	-4.2	-1.65	**-1.40**
p	58	56	54	52	4	2	58	56	54	52	4	2	58	56	54	52	4	2
N	561				27		561				27		561				27	

Notes to Table 7 This table presents specification test statistics for three estimates of each model for each industry. The "winning" model is indicated in **bold**. p is the number of parameters in each model, and N is the sample size. The formula for the Akaike and Schwarz criteria are given in the text.

Figure 4 - Output shares and GDP per worker when non-linearities are significant
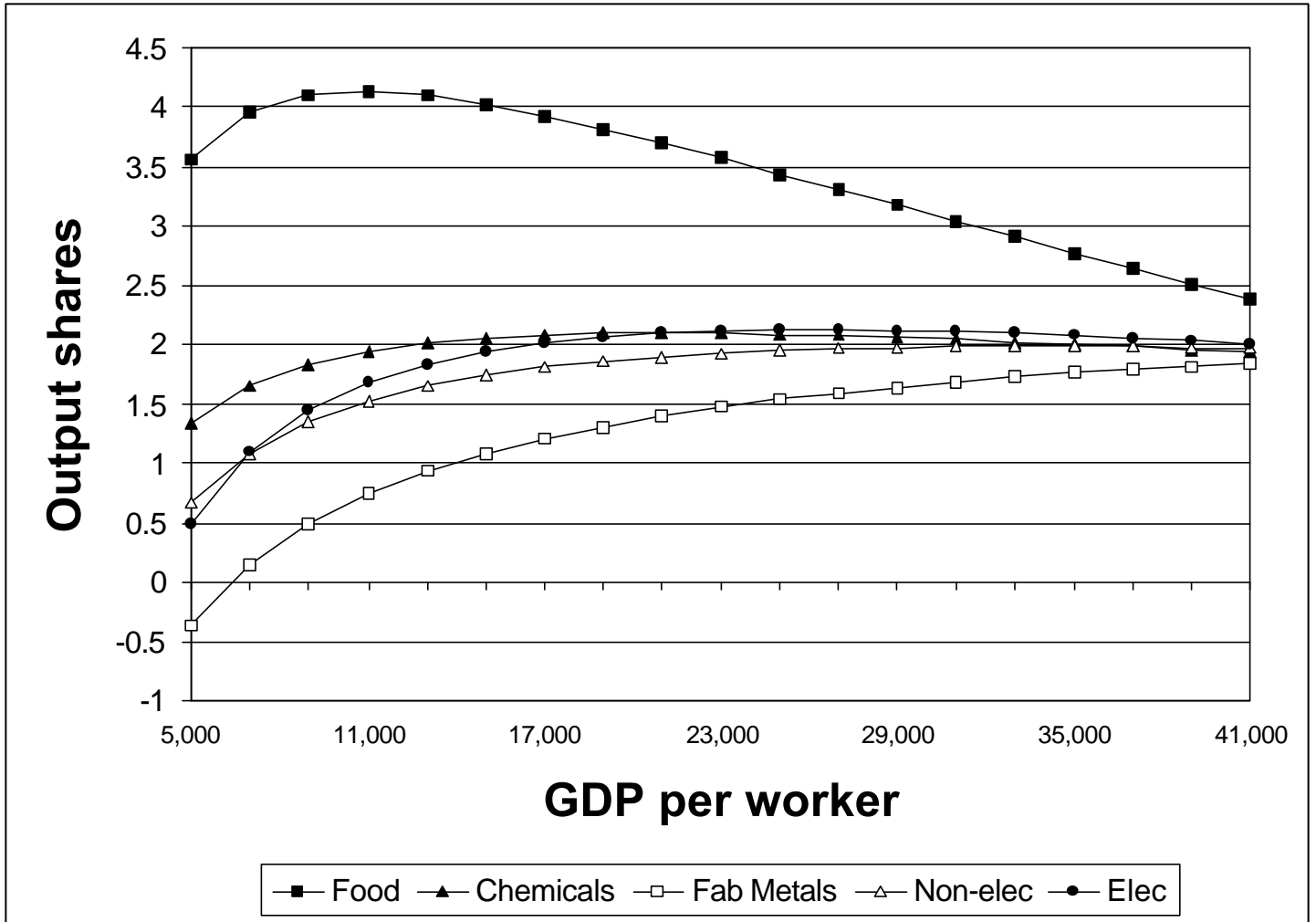
Appendix

In this appendix, we discuss our estimation methods for equations (11) and (12). For the purposes of this appendix, we will generically write an unbalanced panel model for output shares for industry i as

$$s_{ict} = \alpha_{ic} + \beta_i' x_{ct} + \varepsilon_{ict} \qquad\qquad c = 1,...,C \qquad t = 1,...,T_{ic} \qquad \text{(A1)}$$

where $\beta_i \in \mathbb{R}^K$ is a vector of parameters to be estimated, and $x_{ct} \in \mathbb{R}^K$ is the corresponding vector of explanatory variables. The time index notation emphasizes that the panels are unbalanced.

A.1 Autocorrelation

As noted in the text, we allow the error terms ε_{ict} to follow a stationary AR(1) process, with a common AR(1) parameter ρ_i across countries for each industry:

$$\varepsilon_{ict} = \rho_i \varepsilon_{ic,t-1} + v_{ict} \qquad\qquad \text{(A2)}$$

where v_{ict} is white noise. Our estimator first transforms the model so that the resulting equation errors are serially uncorrelated. We use the panel data modification of the Prais-Winsten (PW) transformation proposed by Baltagi and Li (1991). The PW-transformation, which amounts to quasi-differencing of equation (A1) with separate treatment of the initial observation for each country, is valid for both the fixed and random effects specifications. The transformation is applied to each industry-specific equation separately and consists of two steps[1].

First, we estimate each industry equation by least squares with country fixed effects and collect the residuals $\hat{\varepsilon}_{ict}$. Our estimate of ρ_i is then simply the least squares estimate of ρ_i from the following regression:

$$\hat{\varepsilon}_{ict} = \rho_i \hat{\varepsilon}_{ic,t-1} + e_{ict} \qquad\qquad \text{(A3)}$$

The second step of the Baltagi-Li procedure is to transform (A1) as follows: for $t = 1$, we multiply each observation by $\sqrt{1 - \rho_i^2}$, and for $t = 2,...,T_{ic}$, we multiply each observation by $(1-\rho_i L)$ where L is the lag operator. All of our fixed and random effects estimators use the PW-transformed data, so that by construction the statistical models have error processes free from first-order serial correlation. The error term of the PW-transformed model is denoted by u_{ict} and is defined as

$$u_{ict} = \varepsilon_{ict}\sqrt{1-\rho_i^2} \qquad\qquad t = 1$$

$$u_{ict} = \varepsilon_{ict} - \rho_i \varepsilon_{ic,t-1} \qquad\qquad t = 2,...,T_{ic}$$

(A4)

A.2 Random effects estimation

For notational convenience we will drop the industry subscripts i in what follows, with the understanding that each industry is estimated separately. We assume that the error term u_{ct} has a one-way error components structure with heterogeneous residual variance:

$$u_{ct} = \theta_c + \zeta_{ct}$$
$$\theta_c \sim N(0,\gamma_c^2)$$
$$\zeta_{ct} \sim N(0,\sigma_c^2)$$

(A5)

Note that we allow the variance of both the random country effect θ_c and the remainder disturbance ζ_{ct} to vary across countries. As it happens, allowing for heterogeneous variances leads to large efficiency gains in our application. We also assume that the random effects and the disturbances are uncorrelated.

The C country-specific equations observed over T_c time periods can be conveniently written in matrix form as

$$\mathbf{s} = \mathbf{X}\beta + \mathbf{u} = \mathbf{X}\beta + \mathbf{Z}\theta + \zeta$$

(A6)

where $\mathbf{s}$ is the (n x 1) vector of observed PW-transformed output shares, $\mathbf{X}$ is the (n x (K+1)) matrix of observed PW-transformed explanatory variables, $\mathbf{Z}$ is the known (n x C) design matrix, θ is the (C x 1) vector of unknown random effects, ζ is the (n x 1) unobserved vector of remainder disturbances, and $n = \sum_{c=1}^{C} T_c$ is the number of country/year observations in the panel. The $n \times n$ covariance matrix of $\mathbf{u}$ is

$$\Omega = \mathbf{Z}G\mathbf{Z}' + R$$

(A7)

where the $C \times C$ matrix G and the $n \times n$ matrix R are defined as

$$G = diag[\gamma_c^2]$$
$$R = diag[\sigma_c^2 I_{T_c}]$$

---

[1] As noted in the context of the time-series literature by Maeshiro (1979), the separate treatment of the first observation in the PW transformation results in significant efficiency gains, especially if the explanatory variables are trending (as they are in our dataset).

Given some estimates for G and R, the feasible generalized least squares estimates of θ and β are given by

$$\hat{\beta} = (\mathbf{X}'\hat{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\Omega}^{-1}\mathbf{s} \tag{A8}$$

$$\hat{\theta} = \hat{G}\mathbf{Z}'\hat{\Omega}^{-1}(\mathbf{s} - \mathbf{X}\hat{\beta}) \tag{A9}$$

Our normality assumptions are used in deriving an estimator for Ω. We use the restricted/residual maximum likelihood (REML) method proposed by Patterson and Thompson (1971). As discussed in detail by Harville (1977), the REML approach to variance component estimation offers several significant advantages over ANOVA-type (i.e., method-of-moments) estimators. The first favorable theoretical property of the REML approach is that it accommodates an unbalanced panel design. Second, non-negativity constraints on the variance components or other constraints on the parameter space cause no conceptual difficulties. Third, REML estimates of variance components take into account the loss in degrees of freedom resulting from the estimation of the parameter vector β, thus yielding unbiased estimates of G and R in finite samples.[2]

The REML approach reduces the maximization problem to one over only the parameters in Ω. The REML log-likelihood function is given by

$$l(\Omega) = \text{constant} - \frac{1}{2}\ln|\Omega| - \frac{1}{2}\ln|\mathbf{X}'\Omega^{-1}\mathbf{X}| - \frac{n-K}{2}\ln\left(\mathbf{b}'\Omega^{-1}\mathbf{b}\right) \tag{A10}$$

where $\mathbf{b} = \mathbf{s} - \mathbf{X}[\mathbf{X}'\Omega^{-1}\mathbf{X}]\,\mathbf{X}'\Omega^{-1}\mathbf{s}$. We use a ridge-stabilized Newton-Raphson algorithm to maximize the log-likelihood function in equation (A10). Using the REML estimate of Ω, we use equations (A8) and (A9) to calculate the FGLS estimators of θ and β. The heteroscedasticity-consistent asymptotic covariance matrix of $\hat{\beta}$ is computed according to Diggle, Liang, and Zeger (1995):

$$V\left(\hat{\beta}\right) = (\mathbf{X}'\hat{\Omega}^{-1}\mathbf{X})^{-1}\left[\sum_{c=1}^{C}\mathbf{X}'_c\hat{\Omega}_c^{-1}\hat{u}_c\hat{u}'_c\hat{\Omega}_c^{-1}\mathbf{X}_c\right](\mathbf{X}'\hat{\Omega}^{-1}\mathbf{X})^{-1} \tag{A11}$$

where $\hat{\Omega}_c^{-1}$ and $\hat{u}_c$ are the components of $\hat{\Omega}^{-1}$ and $\hat{u}$, respectively that correspond to country c.

---

[2] The fact that likelihood-based estimators are derived under the assumption of a particular parametric form, generally normality, for the distribution of the data vector is not as restrictive as it may seem at first glance. As discussed by Harville (1977), the close relationship between the distribution-free estimators of variance components, such as the class of locally best translation-invariant quadratic unbiased estimators (e.g., MIVQUE and MINQUE), and REML estimators indicates that the likelihood-based estimators of G, and R derived under normality assumptions are reasonable, even when the form of the error-term distribution is left unspecified.

A.2 Fixed effects estimation

The fixed effects estimator treats the country effect θ_c as a parameter to be estimated as opposed to an error component. Under this specification, $G = 0$ and $\Omega = R = diag[\sigma_c^2 I_{T_c}]$. As before, we use REML to estimate Ω, and the estimator of β is again given by equation (A8); heteroscedasticity-consistent standard errors of $\hat{\beta}$ are computed using (A11).

A.2 Between estimation

The between estimator of equation (A1) averages the country-specific data for each industry. The time average of (A1) is given by

$$s_{ic.} = \alpha_i + \beta_i' x_{ic.} + \varepsilon_{ic.} \qquad\qquad c = 1,...,C \qquad\qquad (A12)$$

where α_i is the overall intercept and

$$s_{ic.} = \frac{1}{T_{ic}} \sum_{t=1}^{T_{ic}} s_{ict}$$

$$x_{ic.} = \frac{1}{T_{ic}} \sum_{t=1}^{T_{ic}} x_{ct}$$

$$\varepsilon_{ic.} = \frac{1}{T_{ic}} \sum_{t=1}^{T_{ic}} \varepsilon_{ict}$$

We estimate equation (A12) for each industry i with weighted least squares, with weights given by $\omega_c = \sqrt{T_{ic}}$. The heteroscedasticity-consistent asymptotic covariance matrix of the weighted least squares estimator $\hat{\beta}$ is computed according to White (1980).

Appendix References

Baltagi, B. H., and Q. Li (1991): "A Transformation that will Circumvent the Problem of Autocorrelation in an Error Components Model," *Journal of Econometrics*, 48, 385-393.

Diggle, P. J., K. Y. Liang, and S. L. Zeger (1995): *Analysis of Longitudinal Data*. Clarendon Pres, Oxford, UK.

Harville, D. A. (1977): "Maximum Likelihood Approaches to Variance Components Estimation and to Related Problems," *Journal of the American Statistical Association,* 72, 320-338.

Maeshiro, A. (1979): "On the Retention of the First Observation in Serial Correlation Adjustments of Regression Models," *International Economic Review*, 20, 259-265.

Patterson, H. D., and R. Thompson (1971): "Recovery of Inter-Block Information When Block Sizes are Unequal," *Biometrika*, 58, 545-554.

White, H. (1980): "A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity," *Econometrica*, 48, 817-838.