ELSEVIER

Journal of International Economics 56 (2002) 55–76

Journal of INTERNATIONAL ECONOMICS

www.elsevier.com/locate/econbase

Do endowments predict the location of production? Evidence from national and international data

Jeffrey R. Bernstein[a], David E. Weinstein[b,c,*]

[a]*La Follette School of Public Affairs and Department of Agricultural and Applied Economics, University of Wisconsin—Madison, Madison, WI, USA*

[b]*Department of Economics, Columbia University, 420 W. 118th St. MC 3308, New York, NY 10027, USA*

[c]*NBER, Washington, DC, USA*

Received 5 May 1999; received in revised form 22 June 2000; accepted 25 September 2000

Abstract

Examining the relationship between factor endowments and production patterns using data from Japanese prefectures and from OECD nations, we find evidence of substantial production indeterminacy. Regressions of outputs on endowments yield prediction errors six to 30 times larger for goods traded relatively freely than for non-traded goods. We argue that a compelling explanation for these results is the existence of more goods than factors in the presence of trade costs. If so, regressions of trade or output on endowments have weak theoretical foundations. Furthermore, since errors are largest in data sets where trade costs are small, we explain why the common methodology of imputing trade barriers from regression residuals has produced counterintuitive results.

Keywords: Factor endowments; Heckscher–Ohlin–Vanek; Production indeterminacy; Trade barriers

JEL classification: F1; F11; F13; D2

In fact, there are exactly 2,118 goods and 2,118 factors. You did know that, didn't you? Edward Leamer (1984)

*Corresponding author. Tel.: +1-212-854-6880; fax: +1-212-854-8059.
E-mail address: dew35@columbia.edu (D.E. Weinstein).

0022-1996/02/$ – see front matter
PII: S0022-1996(01)00108-8

1. Introduction

Trade economists regularly build models in which the number of goods exceeds, equals, or is less than the number of factors. These seemingly innocuous variations in model structure have profound implications regarding the ability of general-equilibrium models to explain production patterns. In models where the number of goods exceeds the number of factors, output and hence trade flows can no longer be determined solely on the basis of a country's factor endowments.[1] Indeed, it is precisely because of this potential indeterminacy of trade and production that many tests of the factor abundance theory have focused on the Heckscher–Ohlin–Vanek (HOV) model (e.g. Bowen et al., 1987; Trefler, 1993, 1995). This formulation posits a relationship between factor endowments and the factor services that are embodied in goods trade. According to the HOV model, countries will export the services of relatively abundant factors and import the services of relatively scarce factors.

Though the HOV model generates precise predictions of trade in factor services, more often economists are interested in using factor endowments to estimate trade flows or outputs. This task often involves an appeal to the existence of what we call the 'factor-endowments-driven' (FED) model of production, which provides the foundation for a common, one-to-one mapping of factor endowments into outputs. A necessary condition for this relationship to hold is the existence of an equal number of goods and factors. Consequently, the empirical literature has tended to rely implicitly or explicitly upon the 'even case' or 'square model,' i.e. the implausible assumption that there are equal numbers of goods and factors. This assumption has troubled empirical trade economists, even though they often adopted it for convenience. As Leamer and Levinsohn (1995) remark in their survey of the empirical trade literature, 'one rather awkward assumption that cries out for change is that of equal numbers of commodities and factors. After all, we really don't know how to count either.'

While we agree that it is not possible to determine the number of goods and factors by counting them, we argue that the observed production patterns suggest a world in which the number of goods exceeds the number of factors. Our claim is predicated on the following empirical prediction. Namely, if there are more goods than factors, then even in cases where the HOV model holds, it should not be possible to predict output on the basis of endowments — i.e. the FED model of production should fail for goods traded costlessly. We implement this test on Japanese prefectural data, analyzing whether factor endowments determine the location of production or whether outputs are indeterminate. Even for this sample of Japanese regions in which the HOV model of production holds, we find that the FED model fails for traded goods, as indicated by enormous indeterminacy in

[1] We follow the existing literature in referring to factor supplies as factor endowments.

production patterns. However, we are able to predict the output for nontraded goods. This suggests that production indeterminacy arising from the existence of more goods than factors is a major problem for predicting specialization.

When we use an international data set, much of the indeterminacy in the location of production for traded goods disappears. That is, residuals from regressions of output on factor endowments are far larger for a data set of regions with negligible to low trade costs than for a data set of countries with presumably higher and more pervasive costs of trade. We interpret this finding as evidence in support of the hypothesis that trade costs help to render international production patterns determinate. However, the fact that production patterns appear to be more predictable in the presence of trade costs undermines a major application of regressions of trade on factor endowments: attempts to identify trade barriers on the basis of prediction errors from these regressions. Indeed, this phenomenon may explain the puzzle identified by Pritchett (1996), namely the significant *negative* correlations between conventional measures of protection and the estimated trade barriers derived from regressions of trade on factor endowments.

2. Dimensionality, production, and trade: theory and tests

In this section we illustrate the relationship between production and factor endowments in a model where production is determinate and in one where it is indeterminate. The objective is to show how theory can help us distinguish between these two possible worlds. We begin by establishing some notation. Let N denote the number of goods, F represent the number of factors, and r index regions (where $r \in R$). For each region r, X^r is the $N \times 1$ vector of gross outputs, V^r the $F \times 1$ vector of factor endowments, and B^r the $F \times N$ matrix of direct factor input requirements.

2.1. Testing for identical production techniques

We now make the standard assumptions about production inherent in the Heckscher–Ohlin–Vanek model. First, we assume that technology is identical across regions and exhibits constant returns to scale. We also assume that regional endowments are not so divergent as to preclude factor price equalization (FPE), that goods and factor markets are perfectly competitive, and that the number of goods is at least as large as the number of factors ($N \geq F$). If these conditions are satisfied, then production techniques will be identical across regions, i.e. $B^r = B \quad \forall\, r \in R$. Moreover, for each prefecture we can write:[2]

[2]It is important to clarify one potential source of confusion about this equation. If we let J denote Japan as a whole, BX^J must equal V^J by definition. However, as Davis et al. (1997) discuss in considerable detail, there is no guarantee that BX^r will equal V^r for *each region* within Japan.

$$BX^r = V^r \ \forall \ r \in R \qquad (\text{Dimension } F \times 1)$$

These R sets of equations can be expressed more compactly as:

$$BX = V \qquad (\text{Dimension } F \times R) \tag{1}$$

where B is the common technology matrix, and X is an $N \times R$ matrix whose columns consist of the output vectors for each region, and V is a $F \times R$ matrix whose columns are the endowment vectors for each region. The columns of the left-hand side of Eq. (1) represent the measured factor content of production for each region and the columns of the right-hand side are the actual factor endowment vectors. Hence Eq. (1) tells us that the measured factor content of production should equal the actual regional endowment.

We refer to Eq. (1) as the *Heckscher–Ohlin–Vanek (HOV) model of production*. Notice that this relationship can fail because of technological differences, increasing returns, or any other reason why factor price equalization might not obtain. If Eq. (1) holds, however, then it must be the case that any violations of the HOV model's core assumptions are insufficient to undermine the theory's posited relationship between outputs, inputs, and technology.

Following Harrigan (1995), we take prediction errors as our metric of model performance. Note that equality of the BX and V matrices in Eq. (1) implies equality for each corresponding element of the two matrices; i.e. for each factor f and region r:

$$[B^f X^r] = [V^{fr}] \qquad (\text{Dimension } 1 \times 1) \tag{1$'$}$$

where B^f denotes the fth row of the technology matrix and V^{fr} the fth element of V^r. Obviously it is too much to expect that Eq. (1′) will hold exactly. Instead we look at the percentage deviation between the predicted factor content of production and the actual factor endowment. In practice, this involves first creating an $F \times R$ matrix, D, whose elements are defined below:

$$D^{fr} = |B^f X^r / V^{fr} - 1|$$

If the errors are small, we conclude that the HOV model provides a reasonably accurate description of production structure. If there are substantial errors, we conclude that there must be substantial regional variation in unit input requirements, causing the model to be deficient in some respect.

We could in principle examine all of the elements of D at once, but it is analytically convenient to consider two types of average prediction errors. First, looking at the column of D corresponding to a given region r, we can compare the actual endowments and measured factor absorption for each region. Second, we can hold f fixed and consider the corresponding row of D, comparing the actual endowments and measured usage of each factor across all regions. In this way we distinguish how well the model fits individual regions and factors.

These tests of the HOV production model also serve as a critical control case that we can draw upon in subsequent tests of what we will call the FED model. If the HOV model of production holds for all regions, then we can conclude that our inability to predict output on the basis of factor endowments is not the result of increasing returns, technological differences, fewer goods than factors, or any other reason that might cause factor-price equalization to fail. As we argue in the next section, this significantly reduces the possible causes of problems with the FED model, and represents a major advance of this paper.

2.2. Predicting outputs from factor endowments

A fundamental question in international trade is whether production levels and trade flows are determined by factor endowments. Leamer (1984) and Harrigan (1995) were the first to use structural models to explore whether endowments could explain trade and production patterns. In this section we go over the conditions under which the Heckscher–Ohlin model guarantees such a relationship. Ultimately, we would like to write output as a linear function of factor endowments, i.e.:

$$X^r = \Omega V^r \qquad \text{(Dimension } N \times 1) \tag{2}$$

where the Ω matrix has dimension $N \times F$. Eq. (2) is what we call the *factor-endowments-driven (FED) model of production*. In general, only if $N \leq F$ can output be written as a unique function, independent of the endowments. By contrast, only if there is factor price equalization, which in turn requires that $N \geq F$, will there exist a common technology matrix B such that $BX^r = V^r \ \forall\ r \in R$. Hence an equal number of goods and factors is a necessary condition for both the HOV and FED relations to hold. An alternative way to conceptualize this issue is to note that if $BX^r = V^r$, a unique Ω will exist only if B is invertible. Invertibility in turn requires that B is of full rank and that there are an equal number of goods and factors; in this case, Ω equals B^{-1}.

Notice that we have just derived a test of whether there are an equal number of goods and factors. If the HOV model of production works and B has full rank, then the FED model of production will fail if there are more goods than factors. To demonstrate this, consider what happens if $N < F$. In this case, if country endowments are not scalar multiples of each other, then factor price equalization will not obtain in general, and the HOV model of production will be violated. In a one-good, two-factor model, for example, there is no linear relationship between endowments and output that is the same for all regions. On the other hand, if $N > F$ and the other conditions of the model are satisfied, the HOV model of production should apply; that is, all regions will use identical production techniques. However, the equilibrium output vectors are no longer unique, B is not invertible, and there is no one-for-one mapping from endowments into production as postulated in (2). Thus, if there are more goods than factors we should expect Eq. (1′) to hold but Eq. (2) to fail. This simple test based on the invertibility of B serves as our main mechanism for identifying if there are more goods than factors.

How do we assess a failure of Eq. (2)? Once again we evaluate the success of the model by focusing on percentage prediction errors. Specifically we examine the magnitudes of $|\hat{\Omega}^n V^r / X^{nr} - 1|$, where $\hat{\Omega}^n$ is our estimate of the nth row of the Ω matrix and X^{nr} the nth element of X^r. As in our tests of Eq. (1′), we generate average prediction errors across regions and across industries.[3]

Theory tells us that we should expect to see three possible outcomes from these experiments. If both the HOV and the FED models work, then we can conclude that endowments do determine the location of production. Similarly, if both models fail, then we can conclude that the world must violate a fundamental tenet of the HOV framework. The final possibility, that HOV works but the FED model fails, indicates that the basic assumptions of the HOV model hold, but there are more goods than factors.

We should be a bit more careful in interpreting this last outcome, where Eq. (1) holds but Eq. (2) fails. One possibility is that there truly are fewer factors than goods. A second possibility is that our results may come from our aggregation scheme. We will deal with this possibility later in our empirical analysis, but for the moment it is worth pointing out that, in general, aggregation can resolve production indeterminacy arising from the existence of more goods than factors only in exceptional cases. One such stylized situation involves the existence of two factors and three goods (autos, brown shoes, and black shoes), with two goods (the shoe varieties) produced using identical techniques. In this case, it is not possible to predict how much of each type of shoe will be produced, but one could predict production of autos and total shoes by aggregating the two types of shoes. This kind of degenerate case is, of course, highly unlikely to prevail in practice. Particularly in our sample, where industries are defined at a relatively high degree of aggregation, the odds of any two industries employing exactly the same production techniques are remote, at best.[4] Aside from implausible scenarios, the literature addressing the problem has typically concluded that aggregation does not

[3] We also implemented a second test, making use of the fact that somewhat more structure can be placed on the Ω matrix when the HOV model holds. Premultiplying Eq. (2) by B, we obtain $BX = B\Omega V$. Assuming that Eq. (1) holds, we can then substitute for BX to yield:

$$V = B\Omega V \Rightarrow B\Omega = I \qquad \text{(Dimension } F \times F\text{)} \tag{3}$$

where I is an $F \times F$ identity matrix. This imposes F^2 restrictions on NF parameters, which is permissible in our data set where $N > F$. We tested Eq. (3) to see if the estimated parameters of the Ω matrix are related to the B matrix in the manner dictated by theory.

[4] We checked for this possibility by verifying that no two columns in the B matrix were identical.

offer a viable way of obtaining one-to-one mappings from factor endowments into production when $N > F$.[5]

A third possibility is that we may have omitted some important factors from our production specification. Suppose, for example, that we are in a two-good, two-factor model but have data on only one factor. If one has the row of the B matrix corresponding to that factor and all the other conditions of the model are satisfied, then the HOV model would hold even with an omitted factor. On the other hand, the FED model would fail, as it would be impossible to predict production structure without the missing factor. Since we do have not a complete list of potential factors, we must be open to the possibility that some factor endowments are not included in our analysis. Hence, if we find that Eq. (1) holds but Eq. (2) fails, our analysis must address the question of whether there are truly more goods than factors, or whether we have inadvertently omitted important factors. We return to this issue in the empirical implementation. First, however, we discuss the possible role of trade costs in eliminating production indeterminacy arising from the existence of more goods than factors.

2.3. The role of trade costs in reducing production indeterminacy

There are costs to trading certain commodities across regional or national boundaries. Broadly defined, such costs encompass not only the direct costs of transportation, tariffs, and other trade barriers, but all other transactions costs of trade, including information costs. To see how these transactions costs of trade can reduce production indeterminacy in a world with $N > F$, consider the case where $N^* < N$ goods are traded at some cost and $N - N^* > F$. In this world it is possible to obtain factor price equalization, and with FPE there will be no trade in any of the goods with positive transactions costs of trade, as each region will produce the amount of the non-traded goods that exactly satisfies its domestic demand. Given the assumption of homothetic preferences, demand for each good will be a linear function of endowments. This guarantees that the output of the N^* non-traded goods will also be a linear function of endowments. However, given that there are

[5] More generally, it can be demonstrated that it is impossible to aggregate while maintaining reasonable industry definitions. Chang (1979) and Leamer (1994) have thought about this problem more systematically. Suppose that no two goods are produced with the same technology so that we are not in a degenerate case. Then arbitrarily choose a set of F goods, X_1, and separate the technology matrix into B_1, an $F \times F$ matrix corresponding to X_1, and B_2, a matrix of dimension $M \times (N - F)$. The HOV production equation is now $B_1X_1 + B_2X_2 = V$. Since B_1 is invertible, this can be written as $X^* = X_1 + B_1^{-1}B_2X_2 = B_1^{-1}V$. X^* is a set of F aggregated goods, but one cannot give this aggregate an interpretation since every element will be a linear combination (with, in general, some negative weights) of $N - F + 1$ goods. Furthermore, even if one tried to forge ahead by regressing X_1 on V and treating $B_1^{-1}B_2X_2$ as part of the error term, the fact that X_2 is correlated with V means that one's estimates of B_1^{-1} will be biased and inconsistent.

$N - N^* > F$ goods traded costlessly, we will not be able to predict the output of these goods. In other words, endowments should be able to predict the output of goods traded at some cost even if they cannot predict the output of goods traded costlessly.

Leamer (1984) and Xu (1993) provide more elaborate models of how trade costs can eliminate indeterminacy in worlds in which all goods have small positive transactions costs of trade. The basic intuition behind their models is that large trade costs can upset FPE and hence one-to-one relationships between endowments and production; however, small trade costs that leave FPE essentially unaffected will preserve these one-to-one relationships while serving to eliminate production indeterminacy.

The upshot of this analysis is that trade costs may represent an important mechanism for reducing or eliminating production determinacy in a Heckscher–Ohlin model with more goods than factors. On this basis, one expectation is that factor endowments should provide more accurate predictions of output in circumstances where the costs of trade are relatively high. As we discuss later, this is the opposite of what has been conventionally assumed in empirical work.

3. Empirical results

3.1. Data description

This paper makes use of two data sets. The first, covering Japanese prefectures, is a slightly modified version of the data set constructed in Davis et al. (1997). We obtained our technology matrix, B, from that paper but augmented the data to obtain information on 47 Japanese prefectures rather than ten aggregated Japanese regions. For the earlier paper, it was crucial that households consumed in the same regions in which they produced, so that regional trade vectors could be constructed. Since our analysis does not require any information on consumption, we used data from all 47 Japanese prefectures rather than the ten aggregated regions from Davis et al. (1997). Our X vector contained data on gross output for 29 sectors for each prefecture, and our V vector included three factors: workers with less than a college education, college-educated workers, and capital.

The second is an international data set. Information on production comes from the COMTAP data set provided by James Harrigan and available in Feenstra et al. (1997). The endowment data come from Davis and Weinstein (1999), supplemented by education data from Barro and Lee (1993).

3.2. The HOV model of production: tests of Eq. (1′) on regional data

We begin our analysis by verifying that the Heckscher–Ohlin–Vanek model of production is valid for our sample of Japanese regions. In particular, we confirm that $BX = V$, or that each region employs the same production techniques ($B^r = B$ $\forall\, r \in R$). As we mentioned earlier, this relationship could fail to hold as a result of increasing returns, Ricardian technical differences across regions, fewer goods than factors, or any other reason that would cause factor price equalization to fail. In this sense, our test is identical to that used by Davis et al. (1997), but performed on a larger number of observations.

Panels A and B of Table 1 present the average prediction errors of the HOV

Table 1
Deviations from the HOV and FED models of production for Japanese regional data (three factors, 1/SQRT(GDP) weighting)

Panel A: Prefectural average prediction errors

Prefecture	HOV (%) $\left\vert \frac{BX}{V} - 1 \right\vert$	FED (%) $\left\vert \frac{\Omega V}{X} - 1 \right\vert$	Prefecture	HOV (%) $\left\vert \frac{BX}{V} - 1 \right\vert$	FED (%) $\left\vert \frac{\Omega V}{X} - 1 \right\vert$
Hokkaido	9	503	Shiga	7	120
Aomori	17	575	Kyoto	5	287
Iwate	13	303	Osaka	20	112
Miyagi	7	115	Hyogo	5	42
Akita	16	161	Nara	31	226
Yamagata	14	167	Wakayama	11	160
Fukushima	18	75	Tottori	9	423
Ibaraki	12	50	Shimane	17	279
Tochigi	14	90	Okayama	6	50
Gumma	11	116	Hiroshima	6	158
Saitama	20	122	Yamaguchi	8	187
Chiba	21	124	Tokushima	14	426
Tokyo	33	271	Kagawa	10	50
Kanagawa	11	156	Ehime	7	279
Niigata	19	110	Kochi	10	469
Toyama	7	70	Fukuoka	6	151
Ishikawa	12	332	Saga	12	384
Fukui	14	241	Nagasaki	11	2131
Yamanashi	10	270	Kumamoto	8	376
Nagano	14	183	Oita	8	68
Gifu	6	165	Miyazaki	11	523
Shizuoka	16	60	Kagoshima	11	1007
Aichi	15	49	Okinawa	27	2120
Mie	11	53	*Total average*	13	310

Panel B: Average prediction errors for HOV model across factors

Non-college	17
College	14
Capital	7
Total average	13

Table 1. Continued

Panel C: Average prediction errors for FED model across industries

Industry	$\left\| \frac{\Omega V}{X} - 1 \right\|$ (%)	R^2
Agriculture/fishery	56	0.511
Mining	120	0.290
Construction	9	0.882
Processed food	37	0.566
Textiles	294	0.006
Apparel	93	0.014
Lumber and wood	45	0.226
Furniture	102	0.223
Paper & pulp	117	0.236
Publishing	91	0.629
Chemicals	406	0.282
Petroleum & coal	2165	0.123
Rubber	419	0.250
Leather & footwear	539	0.360
Ceramics & Glass	49	0.105
Iron & steel	312	0.264
Non-ferrous metals	945	0.278
Metal products	82	0.393
General machinery	336	0.446
Electrical machinery	516	0.331
Transport machinery	556	0.251
Precision instrument	1266	0.140
Other manufacturing	141	0.429
Transportation/communication	20	0.852
Electricity/gas/water	38	0.338
Wholesale/retail	26	0.723
Finance/insurance/real estate	15	0.857
Other services	23	0.788
Public administration	12	0.796
Total average	310	0.386

Panel D: Average prediction errors for FED model, tradable vs. non-tradable sectors

Industry	$\left\| \frac{\Omega V}{X} - 1 \right\|$ (%)
Tradables	395
Nontradables[a]	20

[a] The following sectors were classified as nontradables: construction, transportation/communication, electricity/gas/water, wholesale/retail, finance/insurance/real estate, other services, and public administration.

model of production, applied to Japanese prefectural data. The prediction errors are generally quite small, averaging 13% across all observations (where each observation is prefecture-factor specific). As explained earlier, we also calculate average errors over each prefecture and over each factor. There are few outliers among the prefectures; only three have average errors in excess of 25%, and none has an average error greater than 33%. Among the factors, the model works best for capital and worst for non-college-educated labor. The small magnitude of the prediction errors indicates that, to the extent that economies of scale or technological differences exist, they are not significant enough to invalidate the production relationship specified by the HOV model. This evidence corroborates the findings of Davis et al. (1997) for more aggregated regions in Japan.

3.3. Assuming the even case: tests of Eq. (2) on regional data

Thus far we have established that the Heckscher–Ohlin–Vanek production model describes the regional data quite well. However, the links between endowments and production that we have explored so far have been loose relationships between endowments and the factor content of production. The question examined in this section is whether the data support a stronger relationship between factor endowments and production. In particular, is there a linear, one-to-one functional relationship between factor endowments and outputs, or are 'correlations' between these two variables the most that we can expect to obtain? As discussed earlier, this amounts to asking whether there are an equal number of goods and factors.

To examine this question, we regressed output on factor endowments. For each industry i, we estimated the following equation:

$$X_{ir} = \gamma_{0i} + \sum_{f \in F} \gamma_{if} V_{fr} + \epsilon_{ir}$$

where the γ values are parameters to be estimated and ϵ values correspond to the errors.

Our analysis is similar to that of Harrigan (1995), but with a major difference: Harrigan used only international data. As a result, it is not clear whether the errors he obtained in fitting the FED model were the result of a violation of a core assumption of the model, or some other problem that would prevent the model from fitting an international sample. The latter includes measurement error, missing factors, government policy, and the failure of FPE in the some of the countries. By using regional data where the HOV model fits well, we can rule out many of the potential reasons why the FED model might fail.

Before turning to our results, we need to address several econometric issues. First, our dependent variable is truncated at zero, and even though only nine observations were zeros, we decided to use a Tobit procedure to correct for a bias in these industries.[6] Second, larger regions are likely to have larger errors, so it is important to correct for heteroskedasticity. We followed the common practice in the literature of deflating all observations by the square root of prefectural GDP.[7] Third is the issue of whether to include a constant term. In a model with equal numbers of goods and factors, a constant would not be necessary unless the error did not have a mean of zero. If we assume that the error incorporates omitted factors, then there is good reason for not forcing the error term to have a mean of zero. In a world with more goods than factors, the equation is misspecified with or without the constant. Since there seemed to be good arguments both for including and excluding the constant, we ran the model both ways. We report only the results from the regressions with a constant, because the results without the constant were similar although the fits were somewhat worse. Finally, if regional demand shocks produce factor supply changes, our coefficient estimates may be inconsistent. Unfortunately, there are no obvious instruments that would allow us to correct for this possibility, and so we are forced to assume that factors are supplied inelastically.

The results of regressing X on V are presented in panels A and C of Table 1.[8] We have calculated average deviations for each prefecture (across all industries) and for each industry (across all prefectures). Strikingly, the average error is more than an order of magnitude larger than that obtained when we compared BX and V. The typical error for the FED model exceeds 300% — almost 25 times larger than the 13% average prediction error for the HOV model. What makes this enormous discrepancy even more astonishing is the fact that our B matrix is given as data, while our Ω matrix was estimated in a way designed to minimize the residuals.[9]

[6] The industries with zeros in them were rubber, leather, and non-ferrous metals.

[7] We used three alternative specifications to adjust for heteroskedasticity: (1) unweighted OLS with Huber–White robust standard errors; (2) weighted OLS with prefectural GDP as the deflator; and (3) endogenous weights, where weights were chosen by assuming that the variance of the error term is proportional to GDP raised to some power. The use of alternative weighting schemes did not qualitatively affect our results.

[8] Coefficient estimates and standard errors can be found in Bernstein and Weinstein (1998). We dropped the nine observations with zero reported output, for which this measure is undefined. We also acknowledge a potential problem with this measure: it may generate exceptionally large errors for observations in which actual production is close to zero and predicted production is negative. Although this possibility exists, it does not explain why we obtain such large average errors. Relatively few points (19 of 1354) fall into this category, and the average error for these observations is 348%, which is not much larger than the average error for the entire sample.

[9] The data also fail the statistical test for the FED model specified in Footnote 3. There we argued that if $BX = V$ (as in our data set), then satisfaction of the FED model ($X = \Omega V$) requires that $B\Omega = I$. In order to test this specified relationship between input requirements (the elements of B) and the coefficients obtained by regressing output on factor endowments (the elements of Ω), we regressed industry outputs on factor endowments and imposed the nine linear constraints implied by the relationship $B\Omega = I$. The appropriate test uses the Wald criterion, which has a $\chi 2$ distribution with nine degrees of freedom. We used two estimating methods, iterative and non-iterative seemingly unrelated regression. The critical value (1% level) of the Wald statistic was 22, but we obtained 1403 and 1053, for the iterative and non-iterative SUR regressions, respectively. The data clearly reject the hypothesis that our estimated coefficients are related to unit input requirements in the manner required by theory.

One striking feature of the data that cannot be perceived directly from Table 1 is the existence of much more extreme outliers in the production data than in the factor endowment data. The average maximum value in the production data was eleven times larger than the median value, while the maximum/median ratio for the endowment data was only about two. However, it is not clear how to interpret this finding. One possible reason for this difference in variances is the existence of Jones magnification effects. Alternatively, far greater variation in production patterns than in factor endowments is also consistent with production indeterminacy. Whatever the reason for these extreme outliers in virtually every industry, they cannot fully account for the poor predictions of the FED model. When we used a least absolute deviations estimator, the results remained qualitatively similar: high prediction errors overall, and especially so for the manufacturing goods sectors. One implication of the large outliers, though, is that the R^2 values tend to overstate the ability of the FED regressions to fit the typical point.

Thus, our results using regional data demonstrate that the FED model fails to hold. Why is this so? Since previous studies have used international data, they have been unable to distinguish among the many potential reasons that the FED model of production might fail. However, by establishing that HOV holds in our regional data, we have eliminated virtually all of the problems that made the Harrigan (1995) results difficult to interpret. Technological differences (e.g. increasing returns), lumpy regions, regional industrial policy, or any other reason that might cause factor price equalization to fail within Japan would cause *both* $BX = V$ and $X = \Omega V$ to fail.[10] Similarly, if there were fewer goods than factors, we would either have found both relationships failing or, with factor mobility, both working. Since the first relationship holds and the second does not, either the B matrix is not invertible due to the existence of more goods than factors, or our analysis has omitted some important factors.

3.4. The possibility of missing factors

Since it is impossible to be certain that all relevant factors have been included, there is no way to prove that the structure of production is indeterminate. For example, college graduates who studied engineering may represent a different type of labor than those who majored in English. However, there is a danger in using factors that are excessively disaggregated, since they may generate tautological relationships between, say, agricultural workers and agricultural output. Moreover, if our tests fail because of the unavailability of data on finely specified factors, then the theory itself is ultimately not very useful, since researchers cannot access this type of information in practice. We therefore focus on the set of factors that can be found in national or international data sets to see if these additional variables can improve the fit of the FED model.

[10] See Courant and Deardorff (1992) for a discussion of problems arising from lumpiness.

Previous studies using international data suggest that additional, measurable factors are unlikely to provide great improvements in the fits of cross-sectional regressions of trade on factor endowments. When Leamer (1988b) looked at the question of how many factors should be included in a model predicting international trade flows on the basis of factor endowments, he found that he could reject models using more than nine (out of eleven potential) factors. Many of the factors that Leamer considered, including illiterate workers, tropical land, desert land, coal, oil and gas, and minerals, are either non-existent in Japan or present in extremely small amounts, so it is doubtful that these factors significantly explain Japanese production patterns.[11] Even before we begin the search for missing factors, then, there is reason to be skeptical that adding more factors will greatly improve the fits.

The most obvious missing factor is land. In principle, we could have used eight different land factors, but many of the land categories just represent different types of fields or construction, so we decided to use two aggregated land variables. The first category was usable urban and agricultural land, and the second was undeveloped mountain and forest land. To consider the effects of using finer measures of human capital, we decomposed our two labor measures into four educational classes: 4-year college and above, 2-year college, high school, and less than high school.

We considered three specifications: (1) the basic three-factor model, with unskilled labor, skilled labor, and capital; (2) a five-factor model, with the original three factors plus two types of land; and (3) a seven-factor model, with four categories of labor, two land variables, and capital. Adding factors to the model yields some improvement in the adjusted R^2 values of the industry regressions, especially in land-intensive sectors like agriculture; however, average prediction errors do not change much. Even with seven factors, the FED production model yields prediction errors that are around 25 times larger than those obtained from the HOV model of production. Our evidence suggests that missing factors, or at least those for which we can obtain data, cannot explain the poor fit of the FED production model.[12] Moreover, Table 2 shows that the basic three-factor model is the preferred specification as evaluated by the Schwarz criterion.

3.5. Is industry aggregation the solution?

Until this point we have refrained from aggregating industries because there is not a sound theoretical justification for this. Nonetheless, given the highly

[11] Japan has no deserts or tropical land. Leamer (1984) reports that less than 1% of the labor force is illiterate in Japan. Furthermore, only 0.1% of the labor force is employed in mining.

[12] We also examined whether differences in factor productivities across prefectures could account for the poor predictive power of the FED model. Even when we permitted non-neutral factor-augmenting technology differences across prefectures, however, our findings remained substantively unchanged.

Table 2
Comparison of models using the Schwarz criterion[a]

Model	Log-likelihood	Schwarz criterion
Three-factor	−11,502	−11,921
Five-factor	−11,430	−12,058
Seven-factor	−11,367	−12,205

[a] The Schwarz criterion = log(likelihood) − $p/2$*log(n), where p is the number of estimated parameters and n is the number of observations. Higher values are preferred. Three-factor model: non-college-educated labor, college-educated labor, and capital; five-factor model: same as above, plus usable land and undeveloped mountain/forest land; seven-factor model: capital, the two land variables, and four labor variables corresponding to the level of education (below high school, high school, 2-year college, 4-year college).

aggregated nature of the factors we used, it is reasonable to ask whether we have stacked the results in our favor by allowing there to be 29 industries but only seven factors. In order to test this, we redid the analysis using the seven factors from before and seven industry aggregates. Six of the industry aggregates were groups of manufactured products used in Davis and Weinstein (1999), where the individual industries were classified according to the relative skilled to unskilled labor ratio of their factor input requirements.[13] The remaining aggregate consisted of the non-tradable sectors, as identified in panel D of Table 1. The results of this exercise, shown in Table 3, are qualitatively similar to those obtained in Table 1. Though smaller, the average prediction error is 123%, an order of magnitude larger than the 13% error obtained from tests of the HOV model.

We also observe an interesting pattern about which more will be discussed below: the average error for the tradable aggregates vastly exceeds that of the non-tradable aggregates. Indeed, even the best fitting sector, aggregate 1, has an average error almost double that of non-tradables. This good fit of aggregate 1 is due to the almost tautological relationship between agricultural output and arable and urban land. If we drop agriculture, the average prediction error for aggregate 1 rises to 76%. Even if we do not discount the relatively good fit of aggregate 1, the average prediction errors for the aggregated tradables sectors are over ten times the size of that for the non-traded goods sectors.

3.6. Evidence on trade costs and indeterminacy at the regional level

We argued earlier that trade costs may help resolve production indeterminacy arising from the existence of more goods than factors. In particular, we hypothesized that in a data set where the number of traded goods exceeds the number of

[13] Agriculture and mining were added to the appropriate aggregates based on the same criterion.

Table 3
Output prediction errors for the FED model using aggregated industries (seven-factors, seven aggregate industries, 1/SQRT(GDP) weighting)

	$\left|\frac{\Omega V}{X}-1\right|$
Aggregate 1 (Agriculture, apparel, lumber & wood, textiles)	22
Aggregate 2 (Ceramics, iron and steel, mining)	74
Aggregate 3 (Food, metal products, paper, rubber)	30
Aggregate 4 (Transport equip., other manufacturing)	266
Aggregate 5 (Electrical machinery, general machinery, leather, nonferrous metals, precision instruments)	294
Aggregate 6 (Chemicals, petroleum, printing & publishing)	164
Manufacturing aggregate subtotal	142
Aggregate 7 *Non-tradable aggregate subtotal* [a]	13
Total average	123

[a] Sectors included in the non-tradable aggregate are construction, transportation/communication, electricity/gas/water, wholesale/retail, finance/insurance/real estate, other services, and public administration.

factors and where certain sectors are traded at cost while other sectors are freely traded, the FED model should work better for non-traded goods than for tradables.

As noted in the previous section, this is precisely what we observed using seven aggregated industries and seven factors. This result is independent of the aggregation scheme, as the same pattern appears in the results using three factors and our original 29 sectors. Panel D of Table 1 gives the average prediction error for tradable sectors and for non-tradable sectors: the average prediction error for non-tradable sectors is only 20%, which is quite close in magnitude to the 13% average error obtained from tests of the HOV model of production. By contrast, the average prediction error for tradables is nearly 400%, or 20 times that for non-tradables. This huge divergence between the performance of the FED model in tradable and non-tradable sectors is consistent with a world in which $N > F$, and some goods N^* have positive trade costs, while the remaining goods ($N - N^* > F$) do not.

3.7. Testing the HOV and FED models on international data

We now examine how the international data set performs on the same battery of tests applied to the Japanese prefectural data. There are two principal differences between the cross-country and regional data. First, technology differences, measurement errors, and other problems are likely to plague the HOV framework when applying it to international as opposed to regional data. One might reasonably expect these problems to result in worse fits for *both* the HOV and the FED models of production. The second important difference, higher costs of trade at the international level, has differential effects on the two models. To the extent that trade costs mitigate indeterminacy without distorting patterns of specialization, the FED model may perform *better* on international than on regional data, even if the HOV model fails at the international level. Whether this is actually the case is an empirical question to which we now turn.

We begin by considering how well the HOV model of production describes the international data. In Table 4, we calculate for a sample of countries the average percentage deviations between *BX* and *V* using Japan's *B* matrix, just as we did for the regional sample in Table 1.[14] As we expected, the fit of the HOV model is far worse at the international level. The average prediction error is 81% for international data, or about six times larger than the 13% obtained using regional data. Even the smallest error for the cross-country sample (22% for Finland) exceeds the average error for the regional data.

We next examine the FED model by regressing output on factor endowments for a sample of OECD countries, using the same set of manufacturing industries employed in the earlier analysis of Japanese prefectures.[15] In our international regressions, we include the three factors used in our basic specification, plus arable land and mineral endowments. We add the latter two factors because they have been used in previous studies, and because they are likely to be more relevant in international comparisons.

One striking feature of the results is that the explanatory power of these regressions, as measured by the adjusted R^2, is much higher than those obtained for the regional sample. As panel C of Table 4 indicates, the regressions of output on factor endowments using international data have an average R^2 of 0.86 — more than double the average from our three-factor runs on prefectures (panel C of Table 1), and about twice that from the five-factor regional regressions reported in Bernstein and Weinstein (1998). The high adjusted R^2 values obtained in the international regressions have the same magnitude as those obtained in other studies of this type, e.g. Harrigan (1995) and Davis and Weinstein (1999), and

---

[14] We could not include Japan in our sample since it would fit by construction.

[15] Note that the international sample used in testing the FED model is different from that used to test the HOV model. The reason for the difference in sample coverage is as follows. To compare $BX = V$, we needed production data on all 29 sectors. Thus, we omitted some OECD countries for which we did not have the necessary data (e.g. Great Britain), but we included a number of non-OECD countries for which we did have this data (e.g. Israel). We could have included more nations in the regressions of X on V, but we elected to use only the OECD observations to facilitate comparisons with previous studies by Harrigan (1995) and Davis and Weinstein (1999).

Table 4
Deviations from the HOV and FED models of production for international data (three factors, 1/SQRT(GDP) weighting)

Panel A: Average prediction errors across countries

Country	HOV (%) $\left\lvert \frac{BX}{V} - 1 \right\rvert$	FED (%) $\left\lvert \frac{\Omega V}{X} - 1 \right\rvert$	Country	HOV (%) $\left\lvert \frac{BX}{V} - 1 \right\rvert$	FED (%) $\left\lvert \frac{\Omega V}{X} - 1 \right\rvert$
Argentina	91		Japan		14
Australia		73	Mexico	56	
Austria	171	26	Netherlands		86
Belgium/Luxembourg		44	New Zealand	115	76
Canada	28	33	Norway	59	114
Denmark		64	Philippines	53	
Finland	22	42	Portugal	42	97
France		43	Singapore	101	
Germany (West)	333	45	Spain	38	41
Great Britain		20	Sweden		50
Greece		66	Thailand	51	
India	72		Turkey	48	308
Indonesia	61		United States	67	7
Ireland	80	101	Yugoslavia		69
Israel	29				
Italy	110	48	*Total average*	81	67

Panel B: Average prediction errors for HOV model across factors

	Capital stock	Non-college-educated labor	College graduates	All factors
All countries (%)	54	71	119	81

Panel C: Average prediction errors and regression R^2s for FED model across industries

Industry	Error (%)	R^2	Industry	Error (%)	R^2
Processed food	25	0.911	Leather & footwear	78	0.610
Textiles	43	0.876	Ceramics & glass	25	0.896
Apparel	50	0.918	Iron & steel	50	0.881
Lumber and wood	50	0.744	Non-ferrous metals	49	0.942
Furniture	34	0.763	Metal products	35	0.915
Paper & pulp	98	0.720	General machinery	54	0.954
Publishing	42	0.869	Electrical machinery	64	0.873
Chemicals	36	0.960	Transport machinery	53	0.960
Petroleum & coal	111	0.853	Precision instruments	317	0.863
Rubber	50	0.896	*Total average*	67	0.863

provide a preliminary indication that the FED model performs better at the international level than at the regional level.

Turning to the prediction errors, which we consider to be a more telling measure of the model's accuracy, we find that the factor-endowment-driven production model fares worse than it did in previous studies of international data. The average prediction error is 67%, or more than 1.5 times the 40% average error reported by Harrigan (1995) in a time-series analysis of similar data. The cross-sectional variation in OECD output appears even harder to explain than the time-series variation. Nonetheless, like the R^2 values, the prediction errors exhibit better performance for the international sample than for the regional sample. The 67% average prediction error is only *one-sixth* as large as that obtained for the same tradable industries using regional data.[16]

This comparison of the international and regional results is harder to interpret than was our comparison of tradable and nontradable commodities at the regional level. For one thing, FPE does not appear to hold in the international context. Thus, the Leamer (1984) and Xu (1993) models are not strictly applicable here. Nonetheless, one plausible conjecture is that larger and more pervasive trade costs at the international level help to eliminate more of the overall indeterminacy in the production of tradable goods. If so, the high R^2 values obtained from international data may reveal less about the performance of the FED production model *per se* than about the interaction between the production model, the consumption model, and trade costs. We grant that this not the only possible explanation for the better performance of the FED model at the international level, and further research on the source of this result is warranted.

4. Conclusion

In this paper we argue that production indeterminacy is substantial in the type of real-world data sets typically available to empirical trade economists. Moreover, we find that the degree of production indeterminacy is greatest when trade barriers and trade costs are relatively low. Using Japanese regional data, we find that prediction errors are 20 times larger for tradable goods than for non-tradable services. Considering a common set of tradable commodities, we find that prediction errors are six times higher for Japanese prefectures than for a sample of OECD nations.

Our results have three important implications. First, in the context of the endowment and output data typically available to empirical researchers, observed

[16] These results cast further doubt on the conjecture that missing factors may be responsible for the poor performance of the FED model at the regional level. If missing factors are important in explaining the poor fits of the FED model on regional data, then these missing factors must be relatively unimportant in explaining international specialization. Since missing factors would likely matter more for international rather than intranational specialization, the possibility of missing factors driving our results seems even more remote.

production patterns appear consistent with a world in which the number of goods exceeds the number of factors. A number of economists have conjectured that there are more goods than factors, including Melvin (1968), Bhagwati (1972), Travis (1972), and Rader (1979), but we provide the first empirical evidence on this question. We acknowledge, however, the enormous difficulties inherent in determining exactly how many goods and factors there are.

A second implication of our results is that one should exercise great caution in interpreting regressions of production (or commodity trade) on factor endowments. A primary application of such regressions has been the estimation of trade barriers, with large residuals interpreted as indications of high trade barriers. Three major users of this methodology are authors seeking to test trade models; those attempting to identify trade barriers or industrial policy interventions (e.g. Leamer, 1988a,b; Noland, 1993; Saxonhouse, 1983, 1986, 1989); and consumers of the preceding literature, who employ their estimates of trade barriers in studies of trade policy and economic growth (e.g. Edwards, 1992; Gould and Gruben, 1996; Levine and Renelt, 1992). Our results suggest that the underlying assumption behind this methodology is flawed, because observed residuals are far larger in contexts where trade barriers and other trade costs are low. Thus, our paper can account for the disconcerting finding of Pritchett (1996): the observed *negative* correlation between the Leamer (1988b) measures of openness and conventional measures of tariffs, non-tariff barriers, and price distortions.

We also found that the standard, factor proportions model of trade does an even worse job of explaining production patterns for tradable commodities than previous researchers had surmised. Thus, a third implication is that, to obtain better descriptions of commodity production patterns, empirical researchers should try to incorporate technological differences, trade costs, and other sources of specialization into their analyses.

Acknowledgements

We wish to thank, without implicating, Richard Bagozzi, Alan Deardorff, Peter Debaere, James Harrigan, Gary Saxonhouse, Robert Stern, two anonymous referees, and especially Donald Davis. Matt Fleming provided excellent research assistance.

References

Bernstein, J.R., Weinstein, D.E., 1998. Do endowments determine the location of production? Evidence from national and international data. NBER Working Paper #6815.

Barro, R., Lee, J.-W., 1993. International comparisons of educational attainment. NBER Working Paper #4349.

Bhagwati, J., 1972. The Heckscher–Ohlin theorem in the multi-commodity case. Journal of Political Economy 80 (5), 1052–1055.

Bowen, H.P., Leamer, E.E., Sveikauskas, L., 1987. Multifactor, multicountry tests of the factor abundance theory. American Economic Review 77 (December), 791–809.

Chang, W., 1979. Some theorems of trade and general equilibrium with many goods and factors. Econometrica 47 (May), 709–726.

Courant, P.N., Deardorff, A.V., 1992. International trade with lumpy countries. Journal of Political Economy 100 (February), 198–210.

Davis, D.R., Weinstein, D.E., 1999. Market access, economic geography, and comparative advantage: an empirical assessment. European Working Review 43 (February), 379–407.

Davis, D.R., Weinstein, D.E., Bradford, S.C., Shimpo, K., 1997. Using international and Japanese regional data to determine when the factor abundance theory of trade works. American Economic Review 87 (June), 421–446.

Edwards, S., 1992. Trade orientation, distortions, and growth in developing countries. Journal of Development Economics 39, 31–57.

Feenstra, R.C., Lipsey, R.E., Bowen, H.P., 1997. World trade flows, 1970–1992, with production and tariff data. NBER Working paper #5910.

Gould, D.M., Gruben, W.C., 1996. The role of intellectual property rights in economic growth. Journal of Development Economics 48, 323–350.

Harrigan, J., 1995. Factor endowments and the international location of production: econometric evidence for the OECD, 1970–1985. Journal of International Economics 39 (August), 123–141.

Leamer, E.E., 1984. Sources of International Comparative Advantage: Theory and Evidence. MIT Press, Cambridge.

Leamer, E.E., 1988a. Cross-section estimation of the effects of trade barriers. In: Feenstra, R.C. (Ed.), Empirical Methods For International Trade. MIT Press, Cambridge, pp. 51–82.

Leamer, E.E., 1988b. Measures of openness. In: Baldwin, R.E. (Ed.), Trade Policy Issues and Empirical Analysis. University of Chicago Press, Chicago, pp. 147–200.

Leamer, E.E., 1994. Trade, wages and revolving door ideas. NBER Working Paper #4716.

Leamer, E.E., Levinsohn, J., 1995. International trade theory: the evidence. In: Grossman, G., Rogoff, K. (Eds.). Handbook of International Economics, Vol. 3. North-Holland, Amsterdam, pp. 1339–1394.

Levine, R., Renelt, D., 1992. A sensitivity analysis of cross-country growth regressions. American Economic Review 82, 942–963.

Melvin, J.R., 1968. Production and trade with two factors and three goods. American Economic Review 58 (December), 1249–1268.

Noland, M., 1993. The impact of industrial policy on Japan's trade specialization. Review of Economics and Statistics 75, 241–247.

Pritchett, L., 1996. Measuring outward-orientation in developing countries: can it be done? Journal of Development Economics 49, 307–335.

Rader, T., 1979. Factor price equalization with more industries than factors. In: Green, J.R., Scheinkman, J.A. (Eds.), General Equilibrium, Growth, And Trade: Essays in Honor of Lionel McKenzie. Academic Press, New York, pp. 347–354.

Saxonhouse, G.R., 1983. The micro- and macroeconomics of foreign sales to Japan. In: Cline, W.R. (Ed.), Trade Policy in the 1980s. Institute for International Economics, Washington, pp. 259–304.

Saxonhouse, G.R., 1986. What's wrong with Japanese trade structure? Pacific Economic Papers 137 (July), 1–36.

Saxonhouse, G.R., 1989. Differentiated products, economies of scale, and access to the Japanese market. In: Feenstra, R.C. (Ed.), Trade Policies For International Competitiveness. University of Chicago Press, Chicago, pp. 145–174.

Travis, W.P., 1972. Production, trade, and protection when there are many commodities and two factors. American Economic Review 62, 87–106.

Trefler, D., 1993. International factor price differences: Leontief was right!. Journal of Political Economy 101 (December), 961–987.

Trefler, D., 1995. The case of the missing trade and other mysteries. American Economic Review 85 (December), 1029–1046.

Xu, Y., 1993. A general model of comparative advantage with two factors and a continuum of goods. International Economic Review 34 (2), 365–385.