ON THE ASSIMILATION OF REGIONAL SAMPLE INFORMATION INTO REGIONAL INPUT-OUTPUT MODELS RELYING ON NATIONAL INPUT-OUTPUT DATA[1]

Russel J. Cooper
Department of Economics, University of Western Sydney - Nepean, P.O. Box 10, Kingswood, NSW 2747, Australia.

ABSTRACT To the extent to which economic analyses of the regional impacts of economic change rely on national input-output data to represent the interrelationships between industry sectors, multiplier calculations are likely to involve substantial overestimation of the final effect. This is because national data contain no information on leakages from a region, other than the lower bound implied by national imports. Region-specific adjustments to nationally constructed input-output coefficients are clearly crucial, but procedures for reallocating coefficients from the inter-industry table to regional "imports" are data intensive, time consuming and prone to contamination by ad hoc reallocation assumptions. In this paper a structured approach to the integration of regional information with national input-output data is proposed. The approach is based on regional sampling followed by econometric estimation of regional input-output tables using a restricted estimator which imposes technological constraints implied by the national input-output matrix. The approach allows various additional assumptions, such as the extent to which a region is isolated and the extent to which a region may be viewed as a microcosm of the economy, as well as other region and industry-specific information, to be imposed in the estimation. The approach can be implemented with a minimal amount of regional sampling (for example, with sample information on regional sourcing of one input only), but lends itself to much more extensive use of regional sample information in a structured context.

1. INTRODUCTION

The need for regional analysis of the effects of economic change is being increasingly recognised. Whether influences on the regional economy come through government intervention or business location decisions, it is still the case that the consequences for the region are often poorly understood because of a lack of appropriate information on business interrelationships at the regional level.

---

[1] This is a revised version of a paper presented to the 15th Pacific Regional Science Conference Organisation Meeting and the 21st Australian and New Zealand Section Meeting of the Regional Science Association International, Victoria University of Wellington, Wellington, New Zealand December 8-12, 1997. I am particularly indebted to conference participants, and especially to Annette Barbetti, Keiran Donaghy, Takao Fukuchi, Geoffrey Hewings, Gennadi Kazakevitch, Tom Murphy, John Madden, Moss Madden, John Roy, Greg Walker and Guy West for a very wide range of comments, criticism and encouragement.

Many economic analyses of regional impacts of economic change rely on national input-output data to represent the interrelationships between industry sectors. While the overall size of industry sectors can be scaled down from national figures to regional levels prior to this type of analysis, it remains the case that the use of nationally constructed input-output relationships is likely seriously to overestimate the regional impact of economic change. This is because the national input-output coefficients - and consequently, implied multiplier effects - treat only national imports as leakage, not "imports" from various regions within the national economy. Consequently, many interactions occur throughout the national economy which contribute to higher multiplier effects nationally. However, at a regional level these interactions can "leak" from the region quite early in the interactive chain and only with lower probability, and lower impact, return at subsequent points in the chain.

While this problem is obvious, it is difficult at a practical level to do much about it because of the enormous costs of collecting specific input-output information at the regional level. Furthermore, where such information can be collected, it often will not be consistent with the national data in terms of sectoral classifications of interest. Additionally, since interactions with the rest of the economy are clearly going to be important, one cannot proceed with a regional input-output analysis, in any event, without dealing with the linkages with the rest of the economy. Thus, there is no escaping the need to link regional modelling with the analytical models and sectoral categories available at a national level.

The purpose of this paper is to propose a methodology for the integration of regional input-output information with national-level input-output models which recognises the role of both national and regional data. The approach incorporates the various types of information to be combined in a structured manner which makes optimal use of all the information in a statistical sense, which allows for continual updating and which provides econometric measures of reliability.

2. THE RELATIONSHIP BETWEEN NATIONAL AND REGIONAL INPUT-OUTPUT STRUCTURES

To fix ideas, suppose that a national input-output table, A, is available. Think of this as a square commodity by commodity table (of size $k \times k$, say), with each commodity representing the principal product of a nationally defined industry or sector.

Let f represent the vector of final demands for commodities at the national level, at the same level of disaggregation of commodities as in the input-output matrix, A. Let q represent the associated vector of industry total output levels. The total industry output level consists of output for intermediate use by other industries together with sales to final demand. This is typically represented as

$$q = Aq + f \tag{1}$$

from which the reduced form relationship linking industry outputs to final demand is

$$q = (I - A)^{-1} f \tag{2}$$

Now consider the case where the focus of attention is on a particular region. Denote this region as Region 1, and let the rest of the economy be denoted Region 2. Let the economy's final demand be disaggregated into final demand for the product of Region 1, f_1, and that for the rest of the economy, f_2, and let total industry output be likewise disaggregated. We have $f = f_1 + f_2$, $q = q_1 + q_2$ and the interrelationships imply

$$\begin{bmatrix} q_1 \\ q_2 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} + \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \tag{3}$$

where A_{11} represents pure intra regional input-output flows (inputs per unit of output) in Region 1, A_{22} represents the rest of the economy's internal flows, and A_{12} and A_{21} allow for interrelationships between Region 1 and the rest of the economy. The disaggregation of the final demand vector should be understood as related to the location of sale of the final product, not to the location of the final demanders. In the two regional economy there are in principle two final demands for each product.

Assume that Region 1 has access to the same technology as is available nationally, and that the region imports from abroad at the same percentage as is true nationally. These may be termed the "common technology" assumption and the "common external links" assumption respectively. While these assumptions are not crucial, they do allow the exposition to focus more on regional sourcing types of assumptions. The combined effect of these two assumptions is that the technology matrix A is still relevant to the domestic input requirements for industries in both Regions 1 and 2, but it needs to be disaggregated into intra-regional and inter-regional blocks in both cases. Specifically, the common technology and common external links assumptions imply both

$$A = A_{11} + A_{21} \text{ and } A = A_{12} + A_{22} \tag{4}$$

However, what these assumptions do not address is the A_{11}, A_{21} split on the one hand and the A_{12}, A_{22} split on the other. Various options are considered in subsequent sections.

To highlight the relevant issues, suppose that particular disaggregations (4) are chosen, and define vectors of outputs and final demands together with a disaggregated input-output matrix which keep separate account of all activities in both Region 1 and Region 2. Letting

$$q_R = \begin{bmatrix} q_1 \\ q_2 \end{bmatrix}, f_R = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \text{ and } A_R = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

the input-output identity for this "two region" economy can be represented as

$$q_R = A_R q_R + f_R \tag{5}$$

where the subscript R denotes the "regionalised" variant of the national input-output relationships. This allows regional effects to be calculated from the relationship

$$q_R = (I - A_R)^{-1} f_R \tag{6}$$

or, in partitioned form (and concentrating on Region 1), [2]

$$\begin{aligned} q_1 = {} & [(I - A_{11}) - A_{12}(I - A_{22})^{-1}A_{21}]^{-1}f_1 \\ & + (I - A_{11})^{-1}A_{12}[(I - A_{22}) - A_{21}(I - A_{11})^{-1}A_{12}]^{-1}f_2 \end{aligned} \tag{7}$$

What equation (7) makes clear is that the level of activity in Region 1 depends upon final demand both within and outside its own boundaries. The extent of this dependence is related not only to its own intra-regional input-output relationships (A_{11}) and its links to the wider economy through purchasing requirements (A_{21}), but also, generally, on interrelationships within the wider economy (A_{22}) and links back from there to Region 1 (A_{12}). These influences are clearly brought out in equation (7), and it is evident that modelling decisions on the disaggregation of A into $A_{11} + A_{21}$ on the one hand and into $A_{12} + A_{22}$ on the other will affect the predicted outcomes of the model.

At least two problems can therefore be identified in this type of regional input-output modelling which extend beyond the usual problems present in any application of the Leontief multiplier.

The problems are, firstly, that without specific information on a region there is no guidance for choice of the required splits of the A matrix and, second, any split that is chosen is a representation of extant purchasing policy, which may be ephemeral, is not technology based, and in any event will need to be continually reviewed, especially in the context of a comparative static exercise which sees some new source of supply become available within a region. That is, problems of input-output coefficient instability, which arise in any application of input-output techniques due to technological change, relative input price changes, and scale effects, are extended to a very high degree indeed by the regional split of the input-output coefficient matrix based on locational sources of supply.

Of course, the problems identified above may be present in principle but be of little empirical importance in an application of regional input-output modelling. The importance of accurately splitting the national input-output matrix into its constituent regional subcomponents may well depend upon the type of analysis to be undertaken

---

[2] Equation (7) follows from the top block (first k rows) of the $2k$ system (6) after partitioned inversion of $(I - A_R)$.

with the regionalised model. Since the seminal work of Miyazawa (1976) on the concept of internal and external multipliers in interregional analysis, there has been extensive discussion of the importance of interregional trade effects in these types of models. Views in the literature appear to be mixed. Research on the concept of the fundamental economic structure of an economy and the "inverse importance" of direct coefficients suggests that the importance of accurately constructing interregional flow coefficients will depend both upon the nature of the regions and upon the purpose of the analysis. For some purposes it may be possible to largely ignore interregional feedbacks (Gillen and Guccione (1980)) while for other purposes a careful accounting of the differences in national and regional characteristics seems to be amply repaid (Harrigan, McGilvray and McNicoll, 1980).

Although there has been no shortage of clever suggestions for avoiding the more expensive, survey-based, approaches to generating input-output coefficients for multi-regional models, there seems to be also an accumulating body of evidence which cautions against extreme assumptions in place of real data. Issues associated with this survey versus non-survey debate are covered, for example, in Hewings (1974), Miernyk (1976), Hewings (1977), Hewings and Janson (1980) and West (1981). The consensus of this line of research seems to be that a cautious compromise ought to be possible. However, there remains the difficulty of determining a priori where the acceptable point of compromise will be. For example, recent results such as those based on the findings of Israilevich, Hewings, Schindler and Mahidhara (1996) suggest that there will certainly be occasions where the choice of input-output subcomponent tables within regional models will be crucial to the predictions of the model. Taking this view seriously, the current paper investigates an approach to using detailed input-output data at the regional level in an integrated fashion with national data. In a sense, it is an alternative to a line of research which has concentrated upon the use of assumptions which allow the development of simplifying formulae aimed at avoiding the need for construction of a full regional input-output table. Drake (1976) and Katz and Burford (1985) are examples of this less data intensive tradition. There is a good deal of evidence from a variety of research results, however, that analysis based on a higher proportion of assumptions to "hard" data is prone to greater error.

In proposing a methodology which allows statistical estimation of key components of the regional input-output matrix, the approach of this paper offers the potential to contribute to the literature on measuring the accuracy of input-output models, an area of investigation which has received a good deal of attention from a variety of perspectives. Jensen (1980), in introducing the concepts of holistic and partitive accuracy, points out methodological issues with the concept of accuracy, especially in isolation from the intended use of a model, and emphasises the difficulties, cost ineffectiveness and possibly the elusiveness in attempting to achieve partitive accuracy in the context of regional input-output modelling. Nevertheless, although this caveat needs to be kept in mind, many innovative approaches to tracking the inverse importance of coefficients and/or to describing aspects of direct coefficient structure which may have particular relationships to the multiplier structure of an economy continue to be developed. An example of an approach which has potential for application to the interregional trade issue is the "fields of influence" technique of Sonis and Hewings (1989). This approach generalises the tracing of the inverse importance of individual coefficients to groups of coefficients and as such could be applied to the determination of the importance of the off-diagonal interregional trade blocks of a multi-regional model.

The approach of Sonis and Hewings is deterministic. West (1986), on the other hand, demonstrates the value of a stochastic approach, provided that the distribution of the direct coefficients is known. Because the current proposal involves a method of estimation of direct coefficients econometrically, standard errors can be obtained as a by-product of the approach. Use of statistically calculated standard errors for development of measures of model accuracy is an area of potential future application of the approach of the current paper which points to the prospect of a reconciliation of many of the views and approaches which have been suggested in the literature.

A related proposal has been suggested and implemented by Gerking (1976a, 1976b). Gerking's approach assumed that individual input-output coefficients were regression coefficients from separate simple regression models in which an errors in variables problem necessitated use of an instrumental variables estimation technique. By contrast, the current proposal aims to estimate each row of the input-output table as a set of regression coefficients from a multiple regression model in which each row of the absorption matrix acts as a set of observations on a particular dependent variable (an input) and each column of the make matrix acts as a set of observations on one of a number of explanators of what is essentially taken to be an "average" firm's input demand function for intermediate inputs (explained, essentially, by the variety of outputs of the "average" firm). In addition, there are cross-equation (cross-input) restrictions, the most simple of which are adding up accounting identities which hold automatically under the technique employed. The approach generalises to encompass more complex within and across equation restrictions.

The remainder of the paper is set out as follows. Section 3 discusses a top down approach to construction of the various sub-matrices in (7). Section 4 then outlines the econometrically based bottom up approach. Section 5 integrates the top down and bottom up approaches through the use of a restricted seemingly unrelated regression estimator. It also exhibits a useful updating formula which may be applied as new sample data become available.

3. REGIONAL MODEL CONSTRUCTION: WORKING FROM THE TOP DOWN

3.1 Top Down Assumptions

The type of information which is needed on a regional level to determine appropriate splits of the input-output coefficient matrix into components, as in (4), is the extent to which businesses purchase their inputs from within or without the region in which they are located. The "top down" approach uses national information which in the first instance, is treated as informative of the structure of the regional economy. There are several options for choosing the regional sub-matrices whilst enforcing (4) within this approach. At one extreme, for example, one could have

$$A_{11} = A_{22} = A \text{ with } A_{21} = A_{12} = 0 \tag{8}$$

and this is effectively the very extreme assumption made in using national input-output coefficients, unadjusted, in regional analyses. Refer to this as the "isolated region" assumption.

For ease of subsequent exposition, it will be convenient to treat this very extreme assumption as a kind of base case assumption which could be applied initially to generate a regionalised input-output matrix. Other assumptions, to be discussed below, can then be conceptualised as being applied subsequently in order to modify the extreme implications of the isolated region assumption. With this in mind, equation (8) may be rewritten as

$$A_R^0 = I_2 \otimes A \tag{8'}$$

where the superscript 0 indicates that this is an initial variant of the regionalised input-output matrix.[3]

An alternative, less extreme but also unrealistic, assumption would be to presume that each row of A_{11} and A_{12} is proportional to the size of the relevant industry total in q_1 relative to q, while corresponding industry rows in A_{21} and A_{22} are proportional to industry totals in q_2 relative to q. Refer to this as the "microcosm of the economy" assumption. Under this assumption,[4]

$$A_{11} = \hat{q}_1 \hat{q}^{-1} A \text{ and } A_{21} = \hat{q}_2 \hat{q}^{-1} A \tag{9a}$$

describe the input structure for Region 1, while

$$A_{12} = \hat{q}_1 \hat{q}^{-1} A \text{ and } A_{22} = \hat{q}_2 \hat{q}^{-1} A \tag{9b}$$

describe the input structure for Region 2. Thus the microcosm of the economy assumption relates the linkages between regions to their relative economic size in the national economy. More generally, the diagonal matrices of row modifiers $\hat{q}_1 \hat{q}^{-1}$ and $\hat{q}_2 \hat{q}^{-1}$ could be varied to reflect changing marketability conditions for Region 1 vis a` vis the rest of the economy. In particular, the row modifier vectors in (9b) need not be related to those in (9a). If w_1 represents market retention weights for Region 1 and w_2 represents market attraction weights from Region 1 to Region 2, the more general form of (9) is

$$A_{11} = \hat{w}_1 A, \quad A_{21} = (I - \hat{w}_1) A, \quad A_{12} = \hat{w}_2 A, \quad A_{22} = (I - \hat{w}_2) A \tag{9'}$$

3 I_2 is a 2 x 2 identity matrix and ⊗ denotes the Kronecker product.

4 The ^ symbol denotes a diagonal matrix formed from the vector operated upon.

Other extreme and generally unrealistic allocations of the national input-output information to the regionalised model include the "unrepresentative region" (from the point of view of Region 1) assumption

$$A_{11} = A_{12} = 0 \text{ and } A_{21} = A_{22} = A \tag{10}$$

and, at the opposite end of the spectrum the "powerhouse region" assumption

$$A_{11} = A_{12} = A \text{ and } A_{21} = A_{22} = 0 \tag{11}$$

The unrepresentative region assumption (10) is assuming that Region 1 cannot meet its own intermediate needs, while the powerhouse region assumption (11) assumes that Region 1 meets these needs not only for itself but also for Region 2.

More realistically, a mix of all these assumptions could be applied. To illustrate, suppose that the common technology and common external links assumptions (4), apply to all products but the other assumptions apply only to selected inputs. Let J_i be an $i \times k$ matrix selecting those inputs (rows) of A for which the isolated region assumption is to be applied. Then (8) may be condensed to $J_i A_{11} = J_i A$, $J_i A_{12} = 0$, with the relevant restrictions on A_{22} and A_{21} holding automatically by virtue of (4). Similarly, letting J_g select those g rows for which the generalised form (9') of the microcosm of the economy assumption is to be applied, J_u select those u inputs to be covered in Region 1 by the unrepresentative region assumption and J_p select the p inputs for which the powerhouse region assumption is relevant to Region 1, the mix of restrictions (4), (8)-(11) may be represented as[5]

$$\begin{bmatrix} I_k & I_k \\ J_i & 0 \\ J_g & 0 \\ J_u & 0 \\ J_p & 0 \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} I_k & I_k \\ J_i & 0 \\ J_g \hat{w}_1 & J_g \hat{w}_2 \\ 0 & 0 \\ J_p & J_p \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix} \tag{12}$$

The types of assumptions which have been highlighted in (12) are rather extreme. As such, even if they are used selectively for certain inputs only, it may be the case that they are only applicable as a description of the input structure of certain activities in Regions 1 or 2, not for all activities. These assumptions can, however, be applied selectively to activities by the use of column operators applied to a modified form of (12). This allows the restrictions in (12) to be applied only to particular columns of A_R.

Let $\hat{v}_1$ denote a diagonal matrix of weights representing the extent to which particular columns of the national input-output matrix A are to be utilised as

---

[5] The subscripting of I and J matrices indicates the dimensionality (of I) and number of rows (of J).

restrictions on corresponding columns of the Region 1 interindustry block A_{11}. Thus, a weight of unity in position *j* allocates the isolated region assumption to activity *j* in Region 1, a weight of zero allocates the unrepresentative region assumption to Region 1 from the point of view of this activity, and so on. Let $\hat{v}_2$ denote a diagonal matrix of weights which will be applied similarly to the intraregional block A_{12}. Here a weight of unity in position *j* allocates the powerhouse region assumption to Region 1 from the point of view of activity *j* conducted in Region 2, while a weight of zero would allocate the isolated region assumption to this activity in Region 2.

Let C_1 select those columns of $\begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}$ which are to be restricted by assumptions such as those discussed above, and let C_2 select those columns of $\begin{bmatrix} A_{12} \\ A_{22} \end{bmatrix}$ which are to be restricted. Generally, the column restrictions may be represented as

$$A_{11}C_1 = A\hat{v}_1C_1, \quad A_{21}C_1 = A(I - \hat{v}_1)C_1, \quad A_{12}C_2 = A\hat{v}_2C_2, \quad A_{22}C_2 = A(I - \hat{v}_2)C_2,$$

that is

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}\begin{bmatrix} C_1 & 0 \\ 0 & C_2 \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix}\begin{bmatrix} \hat{v}_1 & \hat{v}_2 \\ I - \hat{v}_1 & I - \hat{v}_2 \end{bmatrix}\begin{bmatrix} C_1 & 0 \\ 0 & C_2 \end{bmatrix} \tag{13}$$

3.2 A General Model of Top Down Restrictions

Combining sectoral selection relationships such as the column restrictions (13) with regional characteristics assumptions such as the row restrictions (12) suggests the general structure of restrictions

$$TA_rV = RA_r^0S \tag{14}$$

where $A_R^0 = I_2 \otimes A$.

Here A_R^0 is an initial (extreme) assumption on the structure of the regionalised matrix A_R. Matrices *T, V, R* and *S* are constructed to allow for more realistic assumptions on the relative importance of the regions and of the sectors within them. Equation (14) enforces the desired restrictions on relevant elements of A_R by row and column operations applied to the base case regionalised matrix A_R^0.

Equation (14) is meant to be indicative of the structure of a wide range of possible restrictions which could be employed in the top down approach. In general, *T* and *R* will both be of dimension $m \times 2k$ where *m* is the number of independent row restrictions, while *V* and *S* will both be of dimension $2k \times s$ where *s* is the number of column restrictions. In pure application of the top down approach, A_R can be constructed exactly from (14) if $m = s = 2k$ or by generalised inversion (requiring some selection criterion) if $m > 2k$ and/or $s > 2k$, such as by use of Moore-Penrose generalised inverses giving, for example, a restricted regionalised matrix, A_R^* say, constructed as

$$A_R^* = (T'T)^{-1}T'RA_R^0SV'(VV')^{-1} = R^*A_R^0S^* \tag{15}$$

The purpose in setting out the top down approach in this way is not to recommend relationships such as (15) as a stand-alone procedure for construction of the regionalised input-output matrix A_R, however, but rather to record the structural form of the relationship between A_R and A, given by (14), as indicative of a top down methodology which begins with the national input-output matrix A, expands this to a first-pass naive regionalised matrix $A_R^0 = I_2 \otimes A$ and subsequently modifies this through a series of selective assumptions to arrive at a more realistic structure for the regionalised matrix.

4. REGIONAL MODEL CONSTRUCTION: WORKING FROM THE BOTTOM UP

4.1 Basic Regional Data Requirements

While a (data intensive) regional approach which would be analogous to the comprehensive (census based) national approach could always be developed, the focus of this paper is on an approach which will be less demanding in terms of data collection, more amenable to frequent revision of data, and which will ultimately be complementary to the national statistics. For this reason a statistical (regression) based approach is proposed.[6]

Continuing to focus on one particular region (Region 1) versus the rest of the economy (Region 2), suppose that a sample of n firms throughout Region 1 is taken, and details of these firms' purchasing decisions and patterns of sales are recorded on a regional and commodity basis. Let

$$Y = \begin{bmatrix} Y_{11} \\ Y_{21} \end{bmatrix}$$

denote a $2k \times n$ absorption matrix of commodities by the sample of firms from Region 1, where Y_{11} is a $k \times n$ (commodities by firms) matrix of purchases from within the region, while Y_{21} is a similarly dimensioned matrix of purchases from outside the region (that is, flows from Region 2 to Region 1).

Correspondingly, let

$$X = [X_{11} \ X_{12}]$$

[6] In discussing this approach, it is assumed that sample information is available from one region only (Region 1). The technology of both Region 1 and Region 2 is inferred from this. The approach generalises to allow sampling from all regions. However, focusing on Region 1 in the exposition brings out most clearly both the power of the approach and the extent of dependence on strong commodity technology assumptions.

denote an $n \times 2k$ make matrix of sales by firms in Region 1 of commodities sold within the region (X_{11}) and of commodities sold outside the region (X_{12}). Both X_{11} and X_{12} are $n \times k$ (firms by commodities) matrices.

It is proposed to use a statistical procedure to split the $k \times k$ national input-output coefficient matrix into its $2k \times 2k$ regionalised variant A_R using the regional information available in Y and X. For this purpose, the sample of firms should be greater than twice the number of commodities ($n > 2k$). Additionally, the selected firms should collectively produce the full range of commodities under consideration (or alternatively, only exclude those commodities for which a simple assumption such as those discussed in the previous section would suffice). For expository purposes, it is convenient in this section to suppose that all commodities are represented as inputs and as outputs within the sample of firms.

4.2 Stochastic Commodity Technology Model

A stochastic extension of the commodity technology assumption applied to Region 1 allows observations over the sample of firms to be interpreted as information on the structure of the regionalised matrix A_R. Specifically

$$Y' = XA_R' + U \tag{16}$$

In (16) each column of Y' (row of Y) shows the purchases of a commodity as an input by the sample of n firms. (Each row of Y_{11} shows purchases of an input from within the region, while the corresponding row of Y_{21} shows purchases from outside the region. For expository purposes purchases of an input from different regions can be regarded as different "commodities".) Consider commodity j, and denote the j^{th} column of Y' by y_j. Then (16) implies

$$y_j = Xa_j + u_j \; ; \quad j = 1, \cdots, 2k \tag{17}$$

where a_j is the j^{th} column of A_R'. Interpreting (17) as a regression equation, the stochastic version of the commodity technology assumption allows the expected usage of commodity j as an input to be "explained" by the various columns of the matrix X in the sense that each of the $2k$ columns shows evidence (across the n sampled firms) on the production of a given product which in principle uses commodity j as an input. Here again outputs to different regions are treated as different commodity types. Thus, a particular input (the j^{th}) is needed to produce a range of commodities, and X contains sample evidence on this usage which can be employed to infer the size of the input-output coefficients related to input j and the full range of outputs.

The vector a_j contains the elements of the j^{th} row of A_R. These are input-output coefficients for Region 1 (j^{th} row of $\begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}$) and, by virtue of the (rather strong in this context) commodity technology assumption, input-output coefficients for Region 2

(j^{th} row of $\begin{bmatrix} A_{12} \\ A_{22} \end{bmatrix}$), inferred by implication from the pattern of production of commodities produced in Region 1 but destined for Region 2.

It should be noted here that the commodity technology assumption as employed in this context does not require firms to produce the same commodity destined variously for Regions 1 and 2 to the same specifications. What it imposes is the restriction that Region 1 firms produce commodities destined for Region 2 to the same specifications as competing firms who produce this commodity in Region 2 for sale within Region 2. This being so, it would probably be sensible to use a restricted estimator which estimates (17) subject to the restrictions (4). Such an estimator is readily available as a special case of the approach proposed in the sequel. The purpose of this section, however, is to outline the econometrically based bottom up approach in its extreme form in which information from one region is used to infer the technology for the entire economy.

4.3 The Seemingly Unrelated Regression Estimator

In (17) each of the input-output coefficient vectors (one for each commodity input) is estimable from a seemingly unrelated regression using the same explanatory variable matrix, the make matrix X. In each case, an error vector u_j allows for the fact that the sampled firms actually employ the same commodity technology only on average. Equation (17) is therefore equivalent to a system of regression equations which may be solved under a criterion such as minimisation of sum of squared errors to obtain estimates of A_R, say $\overline{A_R}$. The ordinary least squares estimator of a_j in (17) is

$$\overline{a_j} = (X'X)^{-1}X'y_j \ ; \quad j = 1, \cdots, 2k \tag{18}$$

Stacking these estimates side by side gives

$$\overline{A_R'} = (X'X)^{-1}X'Y' \tag{19}$$

and transposing this, the "two regional" input-output matrix is estimated as

$$\overline{A_R'} = YX(X'X)^{-1} \tag{20}$$

The estimates (20) are, however, unrestricted.[7] It will be noted that in this

---

[7] It will be convenient to introduce notation which anticipates later extensions accommodating restricted estimation by rewriting these equations in system form using Kronecker product notation and vec operations. In moving from matrix equations to vectorised forms of the same equations, repeated use is made of the vectorisation rule (stacking matrices by columns): $\mathrm{vec}(ABC) = (C'\otimes A)\,\mathrm{vec}\,B$. In this notation (16) becomes

$$\mathrm{vec}\,Y' = (I_{2k}\otimes X)\,\mathrm{vec}\,A_R' + u \tag{16'}$$

where $u = \mathrm{vec}U$, a disturbance vector which is assumed to have variance-covariance matrix statistically based "bottom up" approach no account is taken of national data at all. The entire economy wide "two regional" model is constructed from the regional sample data matrices Y and X.

4.4 Features of the Approach Related to its Econometric Aspects

Before extending the statistically-based model to allow incorporation of "top down" restrictions, it is worth pointing out some useful features of the basic approach. There is, of course, no reason why this approach needs to be restricted to a multi-regional context. Indeed, in its originally proposed form (Cooper (1971)) the approach was presented as a natural extension of the development of input-output tables at the national level. In that context, if Y and X are national absorption and make matrices respectively and the commodity technology assumption is written in its more usual deterministic form as $Y = AX'$ then the standard approach to constructing a commodity by commodity input-output table would be to invert the (square and nonsingular) make matrix to obtain $A = Y(X')^{-1}$. Cooper (1971) suggested the use of the Moore-Penrose generalised inverse when X is rectangular (with more "observations" - firms, establishments or activities - than "explanators" - commodities, output products or sectors) to obtain the commodity by commodity coefficients matrix as $\bar{A} = YX(X'X)^{-1}$ and, noting that this is the least squares formula, proposed a stochastic variant of the commodity technology model. The features of the econometric estimation technique which are summarised in this sub-section apply irrespective of the particular regional focus of the current paper.

Feature 1: Standard Errors

Firstly, standard errors associated with the coefficient estimates may be readily obtained. Model "predictions" of the absorption matrix are $\bar{Y} = \bar{A}_R X'$ so that, given (20), the estimated residuals are $Y - \bar{Y} = Y[I - X(X'X)^{-1}X']$. The estimated variance-covariance matrix of the input-output matrix of parameter estimates $\bar{A}_R$, is $\bar{\Sigma}\otimes(X'X)^{-1}$ where $\bar{\Sigma} = Y[I - X(X'X)^{-1}X']Y'/n$ is the matrix of maximum likelihood estimates of Σ.[8]

Since the coefficient estimates are produced from cross-sectional information on firms' uses of inputs in the production of outputs, the estimated input-output

---

$\Sigma\otimes I_n$. This allows within-firm cross-input correlation through the $2k$ by $2k$ matrix of "contemporaneous" correlations Σ, but excludes cross-firm correlation. In view of (16'), the seemingly unrelated regression estimator is

$$\begin{aligned}\text{vec}\bar{A}_R' &= [(I_{2k}\otimes X)'(\Sigma\otimes I_n)^{-1}(I_{2k}\otimes X)]^{-1}(\Sigma\otimes I_n)^{-1}(I_{2k}\otimes X)'\text{vec}\,Y' \\ &= [I_{2k}\otimes(X'X)^{-1}X']\text{vec}\,Y'\end{aligned}$$

giving (19) on reversal of the vec operation.

8 The i^{th} block diagonal component of $\bar{\Sigma}\otimes(X'X)^{-1}$ is the estimated variance-covariance matrix of the i^{th} row of the input-output coefficient matrix and the ij^{th} off-diagonal block is the estimated covariance matrix of the i^{th} and j^{th} rows. These formulae are based on the assumptions associated with model (16'), discussed in footnote 7.

coefficients represent the technology of the "average" firm and the standard errors are informative of the degree of technological variation in the sample. This opens up the potential for the use of the standard errors for a variety of purposes, such as for analyses in which the average input-output coefficients are adjusted by some fraction of the standard error to represent technological change, to provide estimated multiplier effects under alternative scenarios as to the technology employed by firms at the margin, and so on.

The calculation of standard errors in this context (typically, a system of seemingly unrelated regressions with common regressors) also opens up intriguing possibilities for the interpretation of insignificant coefficients. In single equation regressions, or in systems with different explanators for each equation, it is not always easy to determine whether insignificance of parameter estimates should denote "true" insignificance or be due to insufficient sample variation in the relevant explanator. Mechanical attempts to measure the extent to which multicollinearity may be causing the insignificance are problematic.[9] However, in the case where the same explanators appear in each equation there is more help available in avoiding Type II errors than is normally the case. In the interpretation of "zeros" in a coefficient matrix as an indicator of (lack of) structural relationships within an economy, avoidance of the error of accepting the null hypothesis when it is false is of considerable importance - rather more so than conventional tests, which concentrate on avoiding Type I errors, would allow.

Suppose, for example, that a coefficient is estimated as insignificantly different from zero. Is it acceptable to assume that the true coefficient is zero in this case (with implications for interpretation of the structure of interindustry relationships in the economy) or should caution be urged because of possible multicollinearity masking a potentially important interrelationship? The fact that all coefficients in any given column of the input-output matrix are estimated by use of the same explanator means that a method is at hand to help unravel this dilemma. Inspection of the standard errors of other coefficients in the same column can help. If any of these are significant then there is evidence that sufficient variation in the explanator, linearly independent of other explanators, is available in principle to uncover the significance of the explanator. In this context, the insignificance of a particular coefficient almost certainly means a lack of correlation between the input and the output over the sample, and so can be attributed to "true" insignificance with considerably greater certainty than would be the case in the single equation and/or different regressors case.

Turning the issue around, suppose that all coefficients in a given column are statistically insignificant. Since it would be unreasonable to suggest that different inputs are correlated with the one explanator to the same degree, this would almost certainly suggest that there is a problem of lack of sufficient independent variation

---

9 Attempts to construct statistics to "infer" the degree of multicollinearity are methodologically suspect since multicollinearity, if it is present, is a data problem, not an attribute of a population whose characteristics can be inferred from statistical procedures applied to a sample.

in the relevant explanator compared to others. This is more plausible than the alternative that the product in question requires no intermediate inputs. In this case caution should be exercised in declaring the structure to contain zero elements, since multicollinearity is likely to be present to a degree sufficient to cause the experimental situation to be regarded as one of poor design.

The above discussion points to a very valuable aspect of the proposed approach for application in cases where zeroes in the coefficient matrix are of interest, such as in analyses of structure. In this case, access to standard errors can be very valuable. They can also clearly be of great value when "important" coefficients are to be distinguished from less important ones. In particular, they offer an opportunity to avoid the mistake of classifying a coefficient as important based on either its direct size or its indirect multiplier effect when in fact the coefficient may be statistically insignificant.

Estimated covariances between coefficients are also very useful. For example, if an insignificant coefficient is set to zero, unless other coefficients in the same column are adjusted the burden of adjustment falls fully on the value added coefficient because of the adding up identity (see Feature 2 below). However, by making use of information on covariances, adjustment of other coefficients can be spread in a manner consistent with the sample evidence.

Following initial estimation and discovery of insignificant coefficients it would also be possible to make the adjustments by re-estimation, imposing zero restrictions on the insignificant coefficients. This would effectively reallocate the values associated with the insignificant coefficients to other coefficients in the matrix. This type of step-wise procedure does bring with it problems of determination of the "true" level of significance in subsequent tests. However, on a practical basis, it is a natural way to reallocate based on sample evidence.[10]

Feature 2: Automatic Adding Up

A second feature of the approach which may not be immediately obvious is that, by virtue of accounting identities in the data, each column of $\bar{A}_R$ sums to less than unity. The remaining fraction may be interpreted as the component of a dollar's worth of production paid for the use of primary factors, imports, indirect taxes and other exogenous inputs. To see this, let all of the exogenous input accounts be defined by a matrix Z, which depicts absorption of these inputs by firms in Region 1. Total inputs of all the sampled firms from Region 1 may then be defined by the identity $g' = i' \begin{bmatrix} Y \\ Z \end{bmatrix}$. However, it is also true that $g = Xi$ so that $g' = i'X'$. Now suppose that the valued added and other coefficients are estimated by the use of a similar methodology to that employed for the intermediate input coefficients. That is, the primary and other inputs demand functions are assumed to be of the form $Z = VX' + W$, mirroring the commodity technology assumption for intermediate inputs,

---

10 Some of the features associated with the calculation of standard errors are illustrated in the appendix.

(16).[11] Then, following formula (20), the value added coefficients could be expressed as $\bar{V} = ZX(X'X)^{-1}$. It follows that

$$i'\begin{bmatrix}\bar{A}_R\\ \bar{V}\end{bmatrix} = i'\begin{bmatrix}Y\\ Z\end{bmatrix}X(X'X)^{-1} = g'X(X'X)^{-1} = i'X'X(X'X)^{-1} = i'$$

that is, $i'\bar{V} = i' - i'\bar{A}_R$, and as long as the individual rows of the exogenous inputs are not required in disaggregated form, the overall total exogenous inputs (say, "value added") coefficient row can be estimated by subtraction as $i' - i'\bar{A}_R$. Technically, the variance-covariance matrix of the full system is singular and any row or single combination of rows of $\begin{bmatrix}A_R\\ V\end{bmatrix}$ may be dropped from the estimation without affecting the estimates of the remaining rows. In the current context it is natural to estimate the intermediate inputs and drop the equation for the (aggregated) value added coefficients, leaving them to be implied residually via the adding-up identity.[12]

Feature 3: Ease of Model Generalisation Incorporating More Theory and Data

A third feature of the approach is that it lends itself to embedding within a more general modelling context. Specifically, the assumption of a Leontief technology can easily be generalised. The estimation technique extends naturally to allowing for variable (either deterministic or random) parameters. This simply requires the addition to (16) of an auxiliary assumption on the process generating the variation in the average input-output coefficients. To illustrate, suppose that $A_R = BCD$ where B and D are matrices of data. If the objective is to model interregional trade as price responsive, the data in B and D could consist of differential prices or transport costs for inputs purchased from different regions. Alternatively, if production functions are to be modelled as non-homogeneous, the data in B and D could consist of indicators of sectoral activity levels. In this illustration, the matrix C would be a collection of "deep" parameters which may be smaller in size than A_R (a more parsimonious parameterisation) or larger in size than A_R (available data on explanators permitting).

Now since the auxiliary assumptions $A_R = BCD$ imply that $\text{vec}A_R' = (B \otimes D')\text{vec}\,C'$, it follows that (16') generalises to

$$\text{vec}\,Y' = (B \otimes XD')\text{vec}\,C' + u \tag{16''}$$

[11] This is not necessarily a restrictive assumption. There is no reason why V need be constant (although if A is constant then $i'V$ would be constant). See Feature 3 below for a discussion of related issues for intermediate input demands.

[12] The invariance of the econometric estimates to which equation is dropped from the estimation is a general feature of econometric estimation of systems of equations in which cross-equation adding-up ("budget") constraints hold in the data. See, for example, McLaren (1990).

Since B need not be block diagonal, the system of equations may no longer be seemingly unrelated, but (16") is still a straightforward regression estimation problem.

The illustration demonstrates a natural way to introduce either behavioural or technical assumptions (or both) into an extended form of input-output analysis. The analyst can still work with the matrix of average input-output coefficients, but treat it as variable and update it as necessary by the economic modelling of firms' average intermediate input decisions.

Another possible generalisation is a random coefficients approach in which the auxiliary assumption is $A_R = A_R^* + W$, where A_R^* is the expected input-output coefficient matrix and W is a matrix of zero mean random variations around the "average" technology. This leads to natural heteroscedasticity in the model and would require GLS estimation.[13]

Of course, the variable (data dependent) and random parameter approaches could be combined in various ways. For example, the auxiliary assumptions could be that $A_R = BCD$ with C a random matrix further parameterised as $C = C^* + W$ where the * indicates expected values of the random parameters. The essential point is that, by appropriate choice of auxiliary assumptions, a reasonable compromise can be found, if necessary by experimentation, between theory and data in construction of the model.

Given that the approach presented here is designed for use with cross-sectional data, there may of course be rather limited options for accumulation of appropriate data which could explain variation in the average input-output coefficients. Nevertheless, the point should be made that this type of extension is available in principle. It would be especially attractive if time series data on prices were available. With cross-sectional data in a multi-regional context, the approaches to model extension outlined above also suggest an avenue for modelling interregional trade coefficients as a function of transport costs.[14]

---

[13] In this case the interpretation of the standard errors involves other issues in addition to those discussed above. In particular, they contain evidence on different technologies, not just due to other factors omitted from the explanation of input demands, but due to different inputs required per unit of output under, say, an efficient technology compared to a less efficient one, all other factors held equal. This opens up prospects for additional use of the information contained in the standard errors to represent the potential for future diffusion of the better technologies and to allow options for updating the average input-output coefficients based on this evidence.

[14] The "data" matrices B and D, whether they consist of "hard" data or indicator variables (say to represent qualitative influences), control the extent to which the "deep" parameters" in C influence components of A_R. It is possible in principle to use them, therefore, to control the extent to which firms make locational decisions on purchases. Explanators could include transport costs, location of head office (within or outside a region), partnership commitments, and so on. An additional aspect of the value of this type of extension to the model is that it opens up the potential for interesting comparative static exercises based on greater or lesser regional participation of firms.

4.5 Working From the Bottom Up with Limited Information

In sub-section 4.3 it was assumed that the bottom up information would enable inference on all input-output relationships. In reality, it is likely that sample information will be inadequate for this purpose for at least two reasons. Firstly, sampling may not cover the full range of activities in the region. This could be due to a deliberate decision to reduce the costs of sampling by concentrating on certain key sectors for the provision of primary data with the intention of employing top down techniques to fill out the missing information. Second, any one region may not be representative enough to yield the required range of national input-output information, even if a generous sample is available. Again, having concentrated attention on what the sampling allows, it will be necessary to fill out the input-output structure via the addition of top down techniques.

In the next section an approach to the integration of top down and bottom up information is outlined. However, to allow for the limited sampling options outlined above, it is first necessary to modify the notation for the bottom up approach to recognise the reality that only certain components of the full regionalised input-output structure are likely to be able to be freely estimated from sample data in the manner described in sub-section 4.3. The purpose of this sub-section is to set up the necessary notation to enable a limited sampling option to be integrated into a structured approach to full model construction.

Recall that the $2k \times n$ absorption matrix Y represents absorption of commodities bought both from Region 1 (Y_{11}) and from Region 2 (Y_{21}) by a sample of firms in Region 1. Since the list of $2k$ "commodities" conforms to the k commodity categories contained in the national input-output accounts, it needs to be recognised that the sample of firms taken from Region 1 may not yield records of purchases of all $2k$ commodities, or may yield data which is considered unreliable for the purposes of the inferential procedure proposed in the previous sub-section. Define a selection matrix P, say, which selects only the useable rows of Y, eliminating the uninformative rows (that is, the zero or unreliable rows of Y_{11} and Y_{21} and/or those rows for which a deliberate decision is made not to obtain and use sample information). For completeness, let $\tilde{P}$ denote a matrix which selects the "uninformative" (in view of the sampling decision and/or results) rows of Y. Let

$$P = \begin{bmatrix} P_1 & 0 \\ 0 & P_2 \end{bmatrix} \text{ and } \tilde{P} = \begin{bmatrix} \tilde{P}_1 & 0 \\ 0 & \tilde{P}_2 \end{bmatrix}$$

where P_1 selects the useable rows of Y_{11}, P_2 does the same for Y_{21}, while the non-useable rows are selected respectively by $\tilde{P}_1$ and $\tilde{P}_2$. The bottom up approach must now be understood as applying only to those rows of commodity inputs for which sample information is contained in the restricted absorption matrix PY.

The row permutation matrix $E = \begin{bmatrix} P \\ \tilde{P} \end{bmatrix}$ rearranges the rows of Y, moving to the top those rows for which sample based statistical estimation is deemed appropriate and moving to the bottom those rows for which this approach is deemed inappropriate.[15]

Correspondingly, define a selection matrix Q which selects only the informative columns of the $n \times 2k$ make matrix X, eliminating those columns of X_{11} and X_{12} for which no sample information on sales is available. Clearly, input-output coefficients estimated as parameters of output constrained input demand functions cannot be obtained if no sample information on particular outputs is available. The lack of such sample information could be related to the structure of the region or to deliberate cost-cutting decisions, but the precise reasons are not relevant here. Instead, the objective is to deal with the restricted sampling issue for whatever reason it might arise. The preceding analysis needs now to be modified so that it may be thought of as applying to the restricted sample make matrix XQ. Let $\tilde{Q}$ select the zero (or uninformative, or deliberately unsampled) columns of X corresponding to commodities for which no (or insufficient) output information is available from the sample. Let

$$Q = \begin{bmatrix} Q_1 & 0 \\ 0 & Q_2 \end{bmatrix} \text{ and } \tilde{Q} = \begin{bmatrix} \tilde{Q}_1 & 0 \\ 0 & \tilde{Q}_2 \end{bmatrix}$$

so that Q_1 and Q_2 select the relevant columns of X_{11} and X_{12} respectively, while $\tilde{Q}_1$ and $\tilde{Q}_2$ select the remaining columns.

The column permutation matrix $F = [Q \ \tilde{Q}]$ rearranges the columns of X, moving to the left those columns for which sufficient information is available to enable them to act as explanators in the estimation of input-output coefficients.[16]

The two-region input-output matrix may now be written in terms of the subcomponents for which relevant sample information is and is not available as

$$A_R = E' \begin{bmatrix} P \\ \tilde{P} \end{bmatrix} A_R [Q \ \tilde{Q}] F' = E' \begin{bmatrix} PA_R Q & PA_R \tilde{Q} \\ \tilde{P} A_R Q & \tilde{P} A_R \tilde{Q} \end{bmatrix} F' \tag{21}$$

and $PA_R Q$ is that subcomponent to which the bottom up approach is to be addressed.

---

[15] It is worth noting for later use that E is an orthogonal matrix, so that $E^{-1} = E' = [P' \ \tilde{P}']$ and hence the following useful results hold: $P'P + \tilde{P}'\tilde{P} = I$ and $\begin{bmatrix} P'P & P\tilde{P}' \\ \tilde{P}P' & \tilde{P}\tilde{P}' \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$.

[16] As with E, the matrix F is orthogonal, so $F^{-1} = F' = \begin{bmatrix} Q' \\ \tilde{Q}' \end{bmatrix}$ and we note for later use the following implications: $QQ' + \tilde{Q}\tilde{Q}' = I$ and $\begin{bmatrix} Q'Q & Q'\tilde{Q} \\ \tilde{Q}'Q & \tilde{Q}'\tilde{Q} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$.

The stochastic commodity technology assumption (16) may now be applied to the statistically reliable portions of the sample absorption and make matrices PY and XQ in order to estimate the PA_RQ subcomponent of the full regionalised input-output matrix (21). Effectively, the model now is

$$(PY)' = XQ(PA_RQ)' + U^* \tag{22}$$

where (22) may be interpreted as intermediate input demand functions corresponding to the useable input information, modelled as dependent upon the recorded outputs.

Following the reasoning in the derivation of (20) from (16), the subcomponent two-regional input-output matrix estimated by the bottom up approach is[17]

$$\overline{PA_RQ} = PYXQ(Q'X'XQ)^{-1} \tag{23}$$

The remaining subcomponents of the two-regional input-output matrix (21) must clearly be constructed in some other fashion. For illustrative purposes, it is assumed here that top down restrictions such as those implied by (15) are applied to these subcomponents. The full two-regional matrix may therefore be represented as

$$\bar{A}_R = E'\begin{bmatrix} \overline{PA_RQ} & PA_R^*\tilde{Q} \\ \tilde{P}A_R^*Q & \tilde{P}A_R^*\tilde{Q} \end{bmatrix}F' \tag{24}$$

where the subcomponents come from (23) and relationships such as (15).

5. ON THE INTEGRATION OF THE "TOP DOWN" AND THE "BOTTOM UP" APPROACHES

5.1 Isolating Relevant Top Down Restrictions

The top down approach makes very little use of regional data. Regional information is only employed in selection of the types of constraints to be imposed in (14). The proposed bottom up approach (20), on the other hand, makes too extensive use of regional data since it constructs the entire A_R matrix from sampled firms in Region 1. Even in the modified bottom up approach (23), in which it is

---

17 Following the notation introduced in footnote 7, (22) may be written

$$\text{vec}(PY)' = (I_p \otimes XQ)\text{vec}(PA_RQ)' + u^* \tag{22'}$$

This makes use of the vectorisation rule $\text{vec}(ABC) = (C' \otimes A)\text{vec}B$ where $A = XQ$, $B = (PA_RQ)' = Q'A_R'P'$, $C = I_p$, $u^* = \text{vec}\,U^*$ and the identity matrix I_p is of dimension equal to the number of rows in P. The seemingly unrelated regression estimator is

$$\begin{aligned}\text{vec}(\overline{PA_RQ})' &= [(I_p \otimes XQ)'(\Sigma \otimes I_n)^{-1}(I_p \otimes XQ)]^{-1}(\Sigma \otimes I_n)^{-1}(I_p \otimes XQ)'\text{vec}(PY)' \\ &= [I_p \otimes (Q'X'XQ)^{-1}Q'X']\text{vec}(PY)'\end{aligned}$$

giving (23) on reversal of the vec operation and transposition.

explicitly recognised that only a subcomponent of the two-region input-output matrix can be constructed from regional sample information, the regional data is still being asked to do too much since inferences on the technology of firms in Region 2 - and inferences on the extent to which firms from Region 2 utilise products from Region 1 - are being drawn from sample information on Region 1 firms only.[18]

While (24) represents one possible approach to combining top down and bottom up information in the estimation of the full regionalised matrix A_R it is somewhat arbitrary in that top down assumptions are used only for those sub-matrices where bottom up estimation is not possible. For the sub-matrix PA_RQ, where bottom up estimation is possible, this approach ignores any top down information. However, it is possible to integrate some top down information into the estimation of PA_RQ by using a restricted estimation technique instead of the unrestricted estimator (23) outlined in the previous section.

To exploit fully the opportunity for restricted estimation of the sub-matrix PA_RQ it is necessary firstly to develop notation to represent restrictions which apply specifically to this sub-matrix. To isolate these restrictions, define a row-restrictions permutation matrix $E_T = \begin{bmatrix} P_T \\ \tilde{P}_T \end{bmatrix}$ is defined, which permutes the rows of T as follows: P_T selects the rows of T which apply restrictions to those rows of A_R which are selected by P. All other restrictions on rows of A_R, which by definition are not relevant to the estimation procedure, are selected by $\tilde{P}_T$.

By their construction, the selection sub-matrices P_T and $\tilde{P}_T$ have properties such that $P_T T\tilde{P}' = 0$ and $\tilde{P}_T TP' = 0$.[19] It follows that there exists a matrix $T_P = P_T TP'$ with the property that $T_P P = P_T T$ and there further exists a matrix $T_{\tilde{P}} = \tilde{P}_T T\tilde{P}'$ with the property that $T_{\tilde{P}}\tilde{P} = \tilde{P}_T T$.[20]

The matrix T_P contains all the row restrictions which are relevant for estimation of the sub-matrix PA_RQ. All other row restrictions, which are not relevant for the sampling based statistical estimation of PA_RQ, are collected within $T_{\tilde{P}}$.

---

18 There is, of course, no reason why Region 2 firms could not also be sampled (funds permitting). However, the basic point still remains that weak statistical estimates could be strengthened by imposing (reasonable) top down restrictions.

19 To justify the first of these in some detail, note that P_T is designed to select restrictions in T which apply to those rows of the input-output coefficient matrix which are to be estimated from sample data but for which restrictions (other than full exclusion restrictions) are to be applied in the estimation. On the other hand, $\tilde{P}$ is designed to select rows which are not to be estimated at all. Thus $P_T T$ cannot by its design apply to the rows of the coefficient matrix which are to be selected by $\tilde{P}$. By similar reasoning, it is apparent that neither can $\tilde{P}_T T$ select the rows selected by P. This structural (lack of) relationship is used in the proof of the result in footnote 20.

20 To demonstrate this property for T_P, note that

$$T_P P = P_T TP'P = P_T T(I - \tilde{P}'\tilde{P}) = P_T T - P_T T\tilde{P}'\tilde{P} = P_T T$$

where the first equality follows from the definition of T_P, the second from the results implied by the orthogonality of E, discussed in footnote 15, and the final equality follows by the constructed nature of $P_T T$ in contrast to the nature of $\tilde{P}$, as discussed in footnote 19.

Now define a composite row restriction-selection permutation matrix $T_E = E_T T E^{/}$ and observe that

$$T_E = \begin{bmatrix} P_T \\ \tilde{P}_T \end{bmatrix} T \begin{bmatrix} P^{/} & \tilde{P}^{/} \end{bmatrix} = \begin{bmatrix} P_T T P^{/} & P_T T \tilde{P}^{/} \\ \tilde{P}_T T P^{/} & \tilde{P}_T T \tilde{P}^{/} \end{bmatrix} = \begin{bmatrix} T_P P P^{/} & T_P P \tilde{P}^{/} \\ T_{\tilde{P}} \tilde{P} P^{/} & T_{\tilde{P}} \tilde{P} \tilde{P}^{/} \end{bmatrix} = \begin{bmatrix} T_P & 0 \\ 0 & T_{\tilde{P}} \end{bmatrix} \quad (25)$$

that is, T_E has a block diagonal structure.[21]

Similarly, define a column-restrictions permutation matrix $F_V = [Q_V \ \tilde{Q}_V]$ which permutes the columns of V such that the sub-matrices Q_V and $\tilde{Q}_V$ separate out (for purposes of concentration on the sub-matrix $PA_R Q$) the relevant and irrelevant column restrictions on A_R. Then, by construction, $Q^{/} V \tilde{Q}_V = 0$ and $\tilde{Q}^{/} V Q_V = 0$,[22] and there exist matrices $V_Q = Q^{/} V Q_V$ and $V_{\tilde{Q}} = \tilde{Q}^{/} V \tilde{Q}_V$ with the respective properties $Q V_Q = V Q_V$ and $\tilde{Q} V_{\tilde{Q}} = V \tilde{Q}_V$.[23]

Now define a composite column restriction-selection permutation matrix $V_F = F^{/} V F_V$. It follows that

$$V_F = \begin{bmatrix} Q^{/} \\ \tilde{Q}^{/} \end{bmatrix} V \begin{bmatrix} Q_V & \tilde{Q}_v \end{bmatrix} = \begin{bmatrix} Q^{/} V Q_V & Q^{/} V \tilde{Q}_V \\ \tilde{Q}^{/} V Q_V & \tilde{Q}^{/} V \tilde{Q}_V \end{bmatrix} = \begin{bmatrix} Q^{/} Q V_Q & Q^{/} \tilde{Q} V_{\tilde{Q}} \\ \tilde{Q}^{/} Q V_Q & \tilde{Q}^{/} \tilde{Q} V_{\tilde{Q}} \end{bmatrix} = \begin{bmatrix} V_Q & 0 \\ 0 & V_{\tilde{Q}} \end{bmatrix} \quad (26)$$

a block diagonal structure as in (25).[24]

Now, using the permutation matrices E_T and F_V, the general restrictions (14) may be rearranged into the permuted form

$$E_T T A_R V F_V = E_T R A_R^0 S F_V \quad (27)$$

and, given (21), (25) and (26), the left hand side of (27) may be set out as

$$E_T T E^{/} E A_R F F^{/} V F_V = T_E (E A_R F) V_F = \begin{bmatrix} T_P & 0 \\ 0 & T_{\tilde{P}} \end{bmatrix} \begin{bmatrix} P A_R Q & P A_R \tilde{Q} \\ \tilde{P} A_R Q & \tilde{P} A_R \tilde{Q} \end{bmatrix} \begin{bmatrix} V_Q & 0 \\ 0 & V_{\tilde{Q}} \end{bmatrix} \quad (28)$$

It follows that the relevant restrictions to be applied in the estimation of $PA_R Q$ may be read off directly from the upper left block of (28) and (27) as

---

21 The first two equalities in (25) are definitional. The third equality follows from the properties of T_P and $T_{\tilde{P}}$, discussed above and demonstrated in footnote 20. The final equality uses results given in footnote 15, based on the orthogonality of E.

22 Reasoning is similar to that in footnote 19.

23 These results follow in a similar manner to those discussed in footnote 20 with the obvious modifications. For example, results from footnote 16 replace those from footnote 15, drawing on the orthogonality of F.

24 These results follow in a similar manner as for (25), with the use of the analogous intermediate results.

$$T_P(PA_RQ)V_Q = R_P A_R^0 S_Q \tag{29}$$

where $R_P = P_T R$ and $S_Q = SQ_V$.

5.2 The Restricted Seemingly Unrelated Regression Estimator

To impose restrictions (29) in the context of the modified bottom up model (22), recall from (22') (footnote 17) that the model may be written

$$\text{vec}(PY)' = (I_P \otimes XQ)\text{vec}(PA_RQ)' + u^* \tag{30}$$

Now, using similar notation, the top down restrictions (29) may be written as[25]

$$(T_P \otimes V_Q')\text{vec}(PA_RQ)' = \text{vec}(R_P A_R^0 S_Q)' \tag{31}$$

so that restricted least squares estimation of $\text{vec}(PA_RQ)'$ in (30) subject to the linear restrictions (31) implies[26]

$$\begin{aligned}\text{vec}(\overline{\overline{PA_RQ}})' &= \text{vec}(\overline{PA_RQ})' - \{T_P'(T_PT_P')^{-1} \otimes (Q'X'XQ)^{-1}V_Q[V_Q'(Q'X'XQ)^{-1}V_Q]^{-1}\} \\ &\quad \times \{(T_P \otimes V_Q')\text{vec}(\overline{PA_RQ})' - \text{vec}(R_P A_R^0 S_Q)'\}\end{aligned} \tag{32}$$

where $\text{vec}(\overline{PA_RQ})'$ is the ordinary least squares estimator of (30), that is, the vec of the transpose of (22). Effectively, this implies that the restricted estimator of PA_RQ is[27]

$$\begin{aligned}\overline{\overline{PA_RQ}} &= \overline{PA_RQ} + T_P'(T_PT_P')^{-1}[R_P A_R^0 S_Q - T_P(\overline{PA_RQ})V_Q] \\ &\quad \times [V_Q'(Q'X'XQ)^{-1}V_Q]^{-1}V_Q'(Q'X'XQ)^{-1}\end{aligned} \tag{33}$$

where $\overline{PA_RQ}$ is given by

$$\overline{PA_RQ} = PYXQ(Q'X'XQ)^{-1} \tag{34}$$

Equations (33) and (34) represent a structured approach to construction of the subcomponent PA_RQ of the regionalised input-output matrix A_R. The subcomponent is estimated econometrically from the (limited) absorption matrix PY on a (seemingly unrelated) row by row basis using the (limited) make matrix XQ as the set of explanators. Although the formulae are not displayed here, it is clearly possible in this statistical approach to obtain standard errors for all the estimated coefficients and overall goodness of fit measures for each row of the regionalised coefficient

25 This follows on application of the vectorisation rule to the transpose of (29).

26 This is the standard restricted least squares estimator of (30) subject to (31).

27 To obtain (33) from (32) the vectorisation rule is inverted and the result transposed.

matrix.[28] This opens up the potential for a structured approach to the measurement of model accuracy, such as the calculation of implied confidence intervals for the elements of the Leontief inverse and the option to conduct sensitivity analyses on predictions based on variation of coefficient values within statistically realistic bounds.

5.3 An Integrated Approach to Updating the Input-Output Matrix

Suppose that an initial version of A_R is available, say $\underline{A}_R$, which already satisfies the top down restrictions (14), or equivalently (27). This implies that the relevant subcomponent of $\underline{A}_R$ which is amenable to construction by estimation from sample data (that is, the sub-matrix $P\underline{A}_RQ$) satisfies (29). It follows that, in restricted estimation following further data collection, $R_PA_R^0S_Q$ may be replaced by $T_PP\underline{A}_RQV_Q$ in (33). Then equation (33) suggests a computationally simple procedure for updating a regional input-output matrix A_R as new sample information becomes available. Specifically, (33) implies

$$\begin{aligned}\overline{\overline{PA_RQ}} &= \overline{PA_RQ} + T_P'(T_PT_P')^{-1}T_P(P\underline{A}_RQ - \overline{PA_RQ})V_Q \\ &\quad \times [V_Q'(Q'X'XQ)^{-1}V_Q]^{-1}V_Q'(Q'X'XQ)^{-1}\end{aligned} \tag{35}$$

and $\overline{\overline{PA_RQ}}$ has an interpretation as a matrix weighted average of $\overline{PA_RQ}$ and $P\underline{A}_RQ$, or, in an alternative interpretation, as a sub-matrix of estimates based upon the sample information, with an adjustment for the divergence between the sample information and previous information.

Using the conventions outlined above and the structure implied by (21), the full integrated estimator may be written as

---

[28] It should be pointed out that the estimation procedure does not guarantee the delivery of non-negative coefficient estimates. Although the restricted estimation procedure could be generalised to handle non-negativity restrictions, it would also be possible to use the appearance of negative coefficients as evidence of weaknesses in the data set which need to be rectified. For example, if sample information is too highly aggregated, regression coefficients estimated in the manner proposed may be negative even if, at a more fundamental level, the implied input demand structure is quite reasonable. To give a hypothetical illustration, consider the following set of three observations on each of two inputs and two outputs.

Y_1	Y_2	X_1	X_2	Y_1+Y_2	X_1+X_2
3	1	6	4	4	10
2	3	4	5	5	9
1	5	2	6	6	8

Suppose input Y_1 is used exclusively in X_1 while Y_2 is used exclusively in X_2. Clearly the disaggregated input-output coefficients are positive, and would be estimated as such at a disaggregated level (averaging 0.5 for relationship 1 and about 0.56 for relationship 2). However, when aggregated, the average input-output relationship would be estimated as negative.

$$\overline{\overline{A}}_R = E' \begin{bmatrix} \overline{PA_R Q} & P\underline{A}_R\tilde{Q} \\ \tilde{P}\underline{A}_R Q & \tilde{P}\underline{A}_R\tilde{Q} \end{bmatrix} F' \tag{36}$$

This process can be repeated whenever a new sample estimate, $\overline{PA_R Q}$, is calculated using data on regional make and absorption matrices through a process such as (34). Whenever new sample information becomes available then, since the existing $\overline{A}_R$ satisfies (14) or (27), it may be interpreted as $\underline{A}_R$ and integrated with the new sample information via (35) and finally into the full regionalised input-output matrix via (36).

6. CONCLUSION

In this paper an integrated methodology for estimation of regional input-output matrices has been proposed, based on econometric estimation of regional sample information subject to restrictions implied by nationally given technological relationships and assumptions on the various regional relativities. The restricted estimation technique lends itself to continual updating as new information becomes available. One major advantage of the approach, clearly related to its econometric basis, is its ability to deliver measures of the accuracy of the regionalised input-output coefficients through the standard errors of the regression estimates. In principle these can be exploited for sensitivity analyses, for determination of important coefficients, to make allowance for technological progress, to adjust for changed market conditions and generally to make adjustments for the parameter instability which is a major concern in this type of regional modelling.

REFERENCES

Anselin, L. and Madden, M. (1990) (eds), *New Directions in Regional Analysis*, Belhaven: London.

Anselin, L., Rey, S. and Deichmann, U. (1990) The implementation of integrated models in a multi-regional system. In Anselin, L. and Madden, M., (eds), *New Directions in Regional Analysis.* Belhaven: London., pp. 146-170.

Cooper, R.J. (1971). *The Technological Assumptions of the Open Static Input-Output Model*. B.A. Honours Thesis, Macquarie University.

Cooper, R.J. (1972). *Input-Output Estimation by Econometric Techniques*, Seminar Presentation. The Australian National University.

Dewhurst, J.H.Ll., Hewings, G.J.D. and Jensen, R.C. (1991).(eds) *Regional Input-Output Modelling: New Developments and Interpretations*. Avebury: Aldershot.

Drake, R.L (1976) A short-cut to estimates of regional input-output multipliers: methodology and evaluation. *International Regional Science Review*, 1, pp. 1-17.

Gerking, S.D. (1976a) Reconciling 'rows only' and 'columns only' coefficients in an input-output model. *International Regional Science Review*, 1, pp. 30-46.

Gerking, S.D. (1976b) Input-output as a simple econometric model. *The Review of Economics and Statistics*, LVIII, pp. 274-282.

Gillen, W.J. and Guccione, A. (1980) Interregional feedbacks in input-output models: some formal results. *Journal of Regional Science*, 20, pp. 477-482.

Guccione, A., Gillen, W.J., Blair, P.D. and Miller, R.E. (1988) Interregional feedbacks in input-output models: the least upper bound. *Journal of Regional Science*, 28, pp. 397-404.

Harrigan, F.J., McGilvray, J.W. and McNicoll, I.H. (1980) Simulating the structure of a regional economy. *Environment and Planning A*, 12, pp. 927-936.

Harrigan, F. and McGregor, P.G. (1988) (eds) Recent advances in regional economic modelling. *London Papers in Regional Science*, 19, Pion: London.

Harrigan, F. and McNicoll, I. (1986) Data use and the simulation of regional input-output matrices., *Environment and Planning A*, 18, pp. 1061-1076.

Hewings, G.J.D. (1974) The effect of aggregation on the empirical identification of key sectors in a regional economy: a partial evaluation of alternative techniques. *Environment and Planning A*, 6, pp. 439-453.

Hewings, G.J.D. (1977) Evaluating the possibilities for exchanging regional input-output coefficients. *Environment and Planning A*, 9, pp. 927-944.

Hewings, G.J.D. (1984) The role of prior information in updating regional input-output models. *Socio-Economic Planning Sciences*, 18, pp. 319-336.

Hewings, G.J.D. and Janson, B.N. (1980) Exchanging regional input-output coefficients: a reply and further comments. *Environment and Planning A*, 12, pp. 843-854.

Israilevich, P.R., Hewings, G.J.D., Schindler, G.R. and Mahidhara, R. (1996) The choice of an input-output table embedded in regional econometric input-output models. *Papers in Regional Science*, 75, pp. 103-119.

Jackson, R.W. and West, G.R. (1989) Perspectives on probabilistic input-output analysis. In Miller, R.E., Polenske, K.R. and Rose, A.Z. (eds), *Frontiers of Input-Output Analysis*. OUP: New York, pp. 209-221.

Jensen, R.C. (1980) The concept of accuracy in input-output. *International Regional Science Review*, 5, pp. 139-154.

Jensen, R.C. and West, G.R. (1980) The effect of relative coefficient size on input-output multipliers. *Environment and Planning A*, 12, pp. 659-670.

Katz, J.L. and Burford, R.L. (1985) Shortcut formulas for output, income and employment multipliers. *The Annals of Regional Science*, 19, pp. 61-76.

McLaren, K.R. (1990) A variant on the arguments for the invariance of estimators in a singular system of equations. *Econometric Reviews*, 9, pp. 91-102.

Miernyk, W.K. (1976) Comments on recent developments in regional input-output analysis. *International Regional Science Review*, 1, pp. 47-55.

Miller, R.E. (1986) Upper bounds on the sizes of interregional feedbacks in multiregional input-output models. *Journal of Regional Science*, 26, pp. 285-306.

Miller, R.E., Polenske, K.R. and Rose, A.Z. (1989) (eds), *Frontiers of Input-Output Analysis.* OUP: New York.

Miyazawa, K. (1976) Input-output analysis and the structure of income distribution. *Lecture Notes in Economics and Mathematical Systems*, 116. Springer-Verlag: New York.

Pleeter, S. (1980) (ed.), Economic impact analysis: methodology and applications.

Studies in Applied Regional Science, 19. Martinus Nijhoff: London.
Round, J.I. (1983) Nonsurvey techniques: a critical review of the theory and the evidence. *International Regional Science Review*, 8, pp. 189-212.
Sonis, M. and Hewings, G.J.D. (1989) Error and sensitivity input-output analysis: a new approach. In Miller, R.E., Polenske, K.R. and Rose, A.Z. (eds), *Frontiers of Input-Output Analysis.* OUP: New York, pp. 232-244.
Sonis, M., Hewings, G.J.D., Guo, J. and Hulu, E. (1997) Interpreting spacial economic structure: feedback loops in the Indonesian interregional economy, 1980, 1985. *Regional Science and Urban Economics*, 27 pp. 325-342.
Stevens, B.H. and Trainer, G.A. (1980) Error generation in regional input-output analysis and its implications for nonsurvey models. Pleeter, S., (ed.), *Economic Impact Analysis: Methodology and Applications. Studies in Applied Regional Science*, 19 pp. 68-84. Martinus Nijhoff: London.
Tavakoli, M., Roper, S. and Schofield, A. (1988) Some issues in the modelling of small regions: the case of Northern Ireland. In Harrigan, F. and McGregor, P.G. (eds), *Recent Advances in Regional Economic Modelling. London Papers in Regional Science*, 19, pp. 83-99. Pion: London.
West, G.R. (1981) An efficient approach to the estimation of regional input-output multipliers. *Environment and Planning A*, 13, pp. 857-867.
West, G.R. (1986) A stochastic analysis of an input-output model. *Econometrica*, 54, pp. 363-374.

APPENDIX. A Basic Illustration of the Econometric Approach

These results are provided to illustrate some of the features of the approach discussed in sub-section 4.4. They are based on Cooper (1972). In that work the published absorption matrix for Australia for 1962-63 was aggregated to a rectangular form consisting of 105 "activities" (observations) on the usage of seven "products" (input "commodities"). The make matrix, at that time unpublished and in the form of a 105 by 105 predominantly diagonal matrix, was supplied to the author by the Australian Bureau of Statistics. This matrix was also aggregated to a rectangular form, effectively producing a predominantly block diagonal structure so that "similar activities" produced a reasonably similar range of the seven "products". Econometric estimates of the implied average input-output coefficients, using an equation equivalent to (20), are given together with their standard errors in Table A1.

There is at least one highly significant coefficient in each column. Therefore, insignificance of other coefficients in each column cannot be attributed to multicollinearity. In the case of column 1, the (5,1) element (Chemical and Mineral Products) is the only significant coefficient at conventional test levels. In Cooper (1972) covariances were not calculated, but to illustrate the application the above table is presented after adjustment in which each element which is insignificant (using a t-statistic of 2 as a rule of thumb) is set to zero and the released values of the coefficients are spread (roughly) proportionally among the significant coefficients. The procedure demonstates the potential to provide a more parsimonious variant of

Table A1. Coefficient Matrix from Unrestricted Estimation

Sector	1	2	3	4	5	6	7
1. Agricultural	.013	.623	.061	.079	.002	.002	.001
Products	(.019)	(.022)	(.044)	(.050)	(.042)	(.018)	(.006)
2. Processed Foods	.015	.036	.007	.002	.010	.001	.001
	(.009)	(.011)	(.021)	(.025)	(.021)	(.009)	(.003)
3. Clothing, Textiles,	.008	.003	.396	.005	.004	.003	.003
etc.	(.008)	(.009)	(.017)	(.020)	(.017)	(.007)	(.002)
4. Light Manufactures	.009	.021	.013	.208	.018	.010	.046
	(.024)	(.028)	(.054)	(.062)	(.052)	(.023)	(.007)
5. Chemical/Mineral	.067	.009	.012	.025	.395	.064	.031
Products	(.020)	(.023)	(.044)	(.051)	(.043)	(.019)	(.006)
6. Heavy Manufactures	.054	.028	.035	.033	.054	.233	.122
	(.042)	(.048)	(.095)	(.109)	(.092)	(.040)	(.013)
7. Services	.075	.115	.064	.108	.120	.119	.211
	(.059)	(.067)	(.132)	(.152)	(.128)	(.056)	(.018)
Total Intermediate Inputs	.241	.836	.588	.459	.605	.431	.417
Value Added	.759	.164	.412	.541	.395	.569	.583

Table A2. Coefficient Matrix after Reallocation of Insignificant Coefficients

Sector	1	2	3	4	5	6	7
1. Agricultural Products		.7					
2. Processed Foods		.1					
3. Clothing, Textiles, etc.			.5				
4. Light Manufactures				.4			.05
5. Chemical/Mineral Products	.1				.5	.07	.035
6. Heavy Manufactures						.24	.125
7. Services						.12	.215
Total Intermediate Inputs	.1	.8	.5	.4	.5	.43	.425
Value Added	.9	.2	.5	.6	.5	.57	.585

the coefficient matrix by virtue of its concentration on (direct, statistically significant) important coefficients. The results for the current illustration are given in Table A2.

Clearly there is substantial advantage to be gained from simplifying the structure where statistical insignificance allows. One of the valuable features of the proposal discussed in the body of this paper is that the restricted estimator allows the possibility of reallocating insignificant coefficients in accordance with the sample evidence by imposing zero restrictions in re-estimation where insignificant coefficients are found on a first estimation. In the interregional modelling context which is the subject of the current paper, the approach offers the potential to identify significant interregional trade coefficients.