NOT FOR QUOTATION
WITHOUT PERMISSION
OF THE AUTHOR

MIXED ESTIMATION OF SURVEY-BASED INPUT-OUTPUT MODELS

Dino Martellato

July 1980
WP-80-119

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

PREFACE

This paper is concerned with the estimation problems of national and regional static input-output (I-O) models. It is argued that often what are really needed in I-O analysis are the coefficients of an I-O model and not the flows of the I-O table. An econometric estimate of "columns only" coefficients is suggested as a means of obtaining unbiased estimates and a measure of their reliability. Only then is it possible to arrive at a less mechanical than that obtained using usual techniques) adjustment and updating.

Another important feature of the approach suggested is the attention given to extraneous information and judgment. Different estimators are given for the various situations that may occur.

Results of a tentative partial application of this approach to a sector of the I-O model of the Veneto region are given.

MIXED ESTIMATION OF SURVEY-BASED INPUT-OUTPUT MODELS

Dino Martellato

I. INTRODUCTION

There is increasing interest in input-output (I-O) analysis in Italy, both at the national and regional level. Short-cut methods of constructing regional I-O tables will be resected in favor of survey-based techniques; however it is becoming evident that the delay and the costs involved in full survey-based technique are often very high. The purpose of this paper is to assess a method for survey-based I-O models with reduced data requirement but with optimal properties, and to present some provisional results obtained from its application in the Veneto region in Italy.

The main characteristics of this approach are given below:

i) the I-O coefficients are estimated without, or almost before, the I-O flows;

ii) the coefficients are estimated using econometric techniques, by column only and with survey data;

iii) there is of course some prior information and judgment about technical coefficients but there is also a net separation between this and sample statistical information; and

iv) there is a two-sided reconciliation problem: one internal to the sample estimates and another between sample and prior information. The reconciliation involves some judgment but it is tackled with a technique less mechanical than the conventional rAs procedure and more akin to the nature of I-O analysis.

II. AN I-O ACCOUNT TABLE OR AN I-O MODEL?

Unless we are interested in so-called "structural analysis, it is often sufficient in I-O analysis to possess the coefficients only and not the flows.

Constructing a full survey-based I-O table is a complex matter that requires taking

i) a sample survey of firms to determine intersectoral flows disaggregated according to their geographic and sectoral origin and distinguishing features;

ii) a sample survey of the public administration, firms, and families to determine the pattern of final demand; and

iii) a census of employment or something like that for the conversion of sample flows into total flows via per-employee flows.

However, the final and most difficult step in this procedure is the reconciliation of the above three entries.

Unfortunately, often what we really need in I-O analysis are the coefficients and not the flows of the transactions table. Therefore, we maintain that the construction of an I-O table of flows is a difficult but avoidable step. In our approach we omit the transaction flows and, as a consequence, we simplify the problem of reconciliation and remove the need for the "census" of employment. In addition, we reduce the first input because the firms are surveyed on the input side of their production only.

We do not ask for their sales distribution, or for the final purchases of capital. If we omit the second input concerning families, public administration, and also the external sector, we reduce to the minimum the survey-based input of the I-O model.

III. THE VERY NATURE OF INPUT-OUTPUT COEFFICIENTS: THE NEED FOR THEIR ECONOMETRIC ESTIMATE

One of the major problems in a survey is that of non-response. It is clear, however, that a higher percentage response and a greater accuracy may be expected from the firms if fewer data are, as suggesed in Section II, requested to them.

The data requested are the input costs, the value added, the incidence of imports and the employment. Thus we do not find in our survey, data to fill the final demand vectors, nor data to build the table "by rows".

A point that must be stressed is that the coefficients cannot be estimated as a ratio between inputs and output of every sector because:

i) at our disposal we have a sample of firms, not the statistical population; and

ii) every input cost is not exactly determined by the production level in the firm, given the unknown coefficient of the sector, as advocated by its standard definition.

Actually, the cost ${}_kX_{ij}$ for input i, for the k firm, in sector j, does differ from the level that is expected from the application of the technical coefficient a_{ij} to the output of the same firm k. That coefficient is only a mean coefficient, indeed. In other words, we want to estimate the technical coefficient a_{ij} having only a sample of couples $({}_kX_{ij};\ {}_kX_j)$ in which the relation is disturbed by many factors.

These are too numerous to be listed here, but we want to remember the quality differences in the production of different firms, their uneven ability to find and keep the minimum level of output, etc.

The conclusion is that we must estimate a stochastic relationship, not a deterministic one.[1] If every sector has m cost items (value added and imports included) the relationships that are going to be estimated, in sector j, are:

$$X_{ij} = a_{ij} X_j + e_{ij} , \quad \text{for } i = 1,m . \qquad (1)$$

To choose the right estimator for these m relations we must remember that the variables X_{ij} and X_j are often affected by measurement errors and only seemingly unrelated. If we recall the constraint

$$\sum_{i=1}^{m} X_{ij} = X_j ,$$

indeed we see that only m-1 of these are independent and that X_j is dependent on the error term. An instrumental variable estimator of vector a is then in order. An application of this approach is presented in Section VIII.

As a conclusion of this section we turn briefly to the sample survey keeping in mind that aside of every coefficient we will now have its variance with which we measure the accuracy or reliability of its estimate. A good thing is then to arrange the sample in such a way as to reduce--with a given budget--these variances. We suggest to utilize a sample stratified by dimension of the firms and area, alloting more and more interviews to the strata in which higher is the proportion of the statistical population and minor is the homogeneity between firms.

---

[1] On this point see L. Klein (1974), Chapter 8.2.

IV. TOWARD A NEW APPROACH IN RECONCILIATION AND UPDATE OF INPUT-OUTPUT COEFFICIENTS

Having estimated column by column the coeffients of the model we have to assemble them--with all our prior information--in a coherent way because every column has been worked out independently of the others.

Usually, the problem of reconciliation is more difficult than ours because it concerns "rows only" estimates of intermediate flows, "column only" estimates of intermediate flows, direct, or--more often--indirect estimates of the final demand, value added and imports. Every input-outputter can see himself struggling with the I-O table which he is trying to fill with all these things and his prior information on certain cells or proportions on the table.

In our opinion it is necessary to exploit every prior information whose reliability we can judge, in order to reduce the need for survey-based data or in order to integrate them.

But it is also necessary to simplify the problem and keep a net separation between sample and prior information on I-O coefficients. The problem is more simple in our approach because we have "column only" estimates of coefficients (not of flows). But the crucial point is that we have the dispersion (variance) of these estimates with which, as will be made clear soon, we can reconciliate entries in a rational way quite different from that implied by commonly used techniques.[2]

The idea of the reliability of coefficients estimate is, of course, not new, but surprisingly understated in the literature. Following Miernyk, et al. (1970), who first used a "reliability quotient" in the reconciliation of row and column (dual) estimates, there were three studies dealing with this topic:[3] Jensen-McGuarr (1976), Lecomber-Allen (1971) and Gerking (1976) hereafter cited as JM, LA and G respectively. The first, JM, and the third, G, share the so-called dual approach utilizing both the row and column estimates of an I-O matrix. These two

[2] See Lecomber (1975), page 1.

[3] The author is only aware of these.

sets of coefficients r_{ij} and c_{ij} are reconciled in a final estimate which is a mean of r_{ij} and c_{ij} weighted by λ_{ij} and $(1-\lambda_{ij})$ as a measure of their reliability.

$$a_{ij} = \lambda_{ij} r_{ij} + (1-\lambda_{ij})\, c_{ij} \quad . \tag{2}$$

According to G's study λ must be chosen to minimize the variance of a_{ij} and then to maximize the precision of $A = \{a_{ij}\}$[4] of I-O coefficients, while in the study by JM, λ_{ij} is fixed on subjective basis.

The studies of JM and LA share, on the other hand, the use of a modified rAs procedure with which an initial estimate of the I-O table is constrained to given totals. There is a difference between these studies in that JM's procedure is for estimation while that of LA is an updating procedure.

Particularly interesting and general is LA's study which generalizes the rAs procedure to allow for more extraneous information. In its essence the LA modified rAs procedure is:

$$A = (Z - E) + \hat{r} A \hat{s} \quad , \tag{3}$$

where A is the final I-O matrix, Z is its initial estimate.

The formula (3) states that only the part labelled E, of Z is subjected to the rAs treatment where the control totals are measured with error. In LA's procedure there is a drastic separation between Z estimates which are thought as perfectly reliable, and E estimates which bear all the burden of rAs reconciliation. As a reconciliation technique however, rAs looks too mechanical because:

i) r multiplier works uniformly along rows and s uniformly along columns; and
ii) the only prior information subsumed in the procedure is that implied in Z matrix and control totals and that concerning reliability which is of yes-no type (yes, that of Z; no, that of E).

[4] A is the matrix equivalent of vector a.

Turning finally to JM's and G's dual approach, our contention is that it is preferable to have only c_{ij} estimates if these allow for more accuracy in the responses because one piece of accurate information is better than two pieces of inferior information.

The situation, however, which is most likely to occur is that in which both sample and prior information are available with a measure of their reliability. Prior information being an earlier matrix, or a national matrix, or a matrix built on subjective grounds with estimates of experts along the lines developed chiefly at the Battelle Memorial Institute.[5]

In this situation a suitable procedure must integrate these two pieces of information according to their reliability in every cell.

We now develop such a procedure.

V. THE FIRST KIND OF EXTRANEOUS INFORMATION AND THE CONSTRAINED ESTIMATOR OF INPUT-OUTPUT COEFFICIENTS

It is time to mention the kind of prior information which is needed in our approach and which is also thought to be generally available.

This prior information can be essentially of two kinds:

A. Firstly, a set of exact restrictions on coefficients.

It comes out from the nature of I-O models: the coefficients must be positive or zero and their sum unity. Or it comes out from current regional accounts (or from an interregional I-O model) with which the regional I-O model is requested to be compatible. In this case there will be restrictions on the sums of the rows too.

If this is the case, we have enough side information to constrain the matrix of I-O coefficients. To make this point clear, we use some notation. With k = 2 sectors we have:

---

[5] See Fisher-Chilton (1972), Fisher (1975) and Streit (1979).

intersectoral flows; aggregated final demand; disposable resources; value added; imports

X_{11}	X_{12}	X_1	F_1	W_1
X_{21}	X_{22}	X_2	F_2	W_2
X_1	X_2	X	F	W
V_1	V_2	V		V
M_1	M_2	M		M
W_1	W_2	W	F	T

(4)

The coefficients which we want to estimate are in general: $(m \cdot n-2)$ with $(m = k+2)$ and $(n = k+1)$.

row constraints; column constraints

$\frac{X_{11}}{W_1}$ (a_{11})	$\frac{X_{12}}{W_2}$ (a_{12})	$\frac{F_1}{W}$	$\frac{W_1}{W}$
$\frac{X_{21}}{W_1}$ (a_{21})	$\frac{X_{22}}{W_2}$ (a_{22})	$\frac{F_2}{W}$	$\frac{W_2}{W}$
$\frac{V_1}{W_1}$ (a_{31})	$\frac{V_2}{W_2}$ (a_{32})		$\frac{V}{W}$
$\frac{M_1}{W_1}$ (a_{41})	$\frac{M_2}{W_2}$ (a_{42})		$\frac{M}{W}$
1	1	$\frac{F}{W}$	1

(5)

We suppose then to know:[6]

i) the vector of present distribution of disposable resources in the sectors $\frac{W}{i'W}$;

ii) the vector of ratio between total intermediate sales of the sectors and the total of disposable resources[7] $\frac{Xi}{i'W}$;

iii) the ratio between total value added and total disposable resources $\frac{i'V}{i'W}$; and

iv) the ratio between total imports and disposable resources $\frac{i'M}{i'W}$.

For the moment we will not estimate the final demand coefficients because the survey is devoted to cost analysis only.

With n sectors the I-O coefficients to be constrained are then $(m \cdot n - 2 - k)$.

The number of independent constraints in the vector r are $(m + n - 1 - 1)$. The matrix of $(m \cdot n - 2 - k)$ coefficients suitably transformed in a column vector, is indicated as a, while the matrix of weights R has $(m + n - 1 - 1)$ rows and $(m \cdot n - 2 - k)$ columns. The matrix constraint is as follows:

$$\ll Ra = r \gg \; . \qquad (7)$$

---

[6]Vectors are column vectors, the prime denotes transposition, and i is the unity vector.

[7]It is apparent that it is equivalent to know $\frac{Xi}{i'W}$ and $\frac{F}{i'W}$.

$$\begin{bmatrix} \frac{W_1}{W} & \frac{W_2}{W} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{W_1}{W} & \frac{W_2}{W} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{W_1}{W} & \frac{W_2}{W} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{W_1}{W} & \frac{W_2}{W} \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix} * \begin{bmatrix} \frac{X_{11}}{W_1} \\ \frac{X_{12}}{W_2} \\ \frac{X_{21}}{W_1} \\ \frac{X_{22}}{W_2} \\ \frac{V_1}{W_1} \\ \frac{V_2}{W_2} \\ \frac{M_1}{W_1} \\ \frac{M_2}{W_2} \end{bmatrix} = \begin{bmatrix} \frac{X_1}{W} \\ \frac{X_2}{W} \\ \frac{V}{W} \\ \frac{M}{W} \\ 1 \end{bmatrix} \qquad (7)$$

$[(m + n - 1 - 1)(m \cdot n - 2 - k)]\,[(m \cdot n - 2 - k) \cdot 1]\,[(m + n - 1 - 1) \cdot 1]$

$$\underset{(5.8)}{R} \quad \underset{(8.1)}{a} = \underset{(5.1)}{r}$$

weights → coefficients; row and column independent constraints

From the TSLS estimation of every column of m coefficients we obtain a matrix of order (m · m) of covariances for them. For a (m = 4) model they look like this

COLUMN 1 VARIANCE

$$\begin{bmatrix} \mathrm{var}\,(a_{11}) & \mathrm{cov}\,\left(\begin{smallmatrix} a_{11} \\ a_{21} \end{smallmatrix}\right) & \mathrm{cov}\,\left(\begin{smallmatrix} a_{11} \\ a_{31} \end{smallmatrix}\right) & \mathrm{cov}\,\left(\begin{smallmatrix} a_{11} \\ a_{41} \end{smallmatrix}\right) \\ & \mathrm{var}\,(a_{21}) & \mathrm{cov}\,\left(\begin{smallmatrix} a_{21} \\ a_{31} \end{smallmatrix}\right) & \mathrm{cov}\,\left(\begin{smallmatrix} a_{21} \\ a_{41} \end{smallmatrix}\right) \\ & & \mathrm{var}\,(a_{31}) & \mathrm{cov}\,\left(\begin{smallmatrix} a_{31} \\ a_{41} \end{smallmatrix}\right) \\ & & & \mathrm{var}\,(a_{41}) \end{bmatrix} \quad (8)$$

We can finally build up the matrix of covariances, denoted by C, for the column vector "a", starting from two basic hypothesis:

i) As we have a column-only matrix of estimated coefficients where there has not been any possibility - at all - for compensation of the error along the rows. This means that the covariance between coefficients belonging to different columns is zero and there is not room for compensation (direct compensation, actually) or accommodation along the rows.

ii) If a row sum, however, doesn't equal the constraint every coefficient in the row is to be corrected accordingly (see the principal diagonal of the C matrix). All this implies that also column coefficients are going to vary according to the covariances (C rows).

The matrix has (m·n - 2 - k) rows and columns.

$$\begin{bmatrix}
\mathrm{var}\binom{a_{11}}{a_{11}} & 0 & \mathrm{cov}\binom{a_{11}}{a_{21}} & 0 & \mathrm{cov}\binom{a_{11}}{a_{31}} & 0 & \mathrm{cov}\binom{a_{11}}{a_{41}} & 0 \\
 & \mathrm{var}\binom{a_{12}}{a_{12}} & 0 & \mathrm{cov}\binom{a_{12}}{a_{22}} & 0 & \mathrm{cov}\binom{a_{12}}{a_{32}} & 0 & \mathrm{cov}\binom{a_{12}}{a_{42}} \\
 & & \mathrm{var}\binom{a_{21}}{a_{21}} & 0 & \mathrm{cov}\binom{a_{21}}{a_{31}} & 0 & \mathrm{cov}\binom{a_{21}}{a_{41}} & 0 \\
 & & & \mathrm{var}\binom{a_{22}}{a_{22}} & 0 & \mathrm{cov}\binom{a_{22}}{a_{32}} & 0 & \mathrm{cov}\binom{a_{22}}{a_{42}} \\
 & & & & \mathrm{var}\binom{a_{31}}{a_{31}} & 0 & \mathrm{cov}\binom{a_{31}}{a_{41}} & 0 \\
 & & & & & \mathrm{var}\binom{a_{32}}{a_{32}} & 0 & \mathrm{var}\binom{a_{32}}{a_{42}} \\
 & & & & & & \mathrm{var}\binom{a_{41}}{a_{41}} & 0 \\
 & & & & & & & \mathrm{var}\binom{a_{42}}{a_{42}}
\end{bmatrix} \tag{9}$$

The basic idea is that the reconciliation burden measured as the difference between the costrains r and the preliminary TSLS estimate of vector a, (r - Ra), is to be attached to less precise estimates in the vector a, which will be reduced or increased according to the following formula.[8]

$$\hat{a} = a + c^{-1}\ R'\ (RC^{-1}\ R')^{-1}\ (r - Ra) \tag{10}$$

[8]This is a constrained GLS estimator see Theil (1971) page 285. For its application and derivation in an I-O Table see Martellato (1978).

The resulting coefficients estimates are now perfectly compatible with row and column constraints and have the property of being unbiased and the most precise within all linear estimates of I-O coefficients.

VI. THE SECOND KIND OF EXTRANEOUS INFORMATION

B. A second kind of extraneous information on I-O coefficients we want estimated is obtained from "sector experts" and from earlier statistical estimates of these coefficients.

This extraneous information doesn't necessarily take the form of a constraint for coefficients, but rather that of a point estimate to which the experts attach a probability in the form of a standard error or - equivalently - the form of a confidence interval.

We have in this case a second vector α - containing the most probable values attached to the parameters and a second diagonal matrix Λ with unity weights. It's worth noting that r contains sums of coefficients, α contains individual coefficients.

We have, moreover, a vector ν of errors of which we know the covariance matrix T in which we convey all the confidence on α estimates displayed by experts or earlier estimates.

The experts must then be able to bind the α vector or, what is the same, to define T_{11}, T_{22} and so on because:

$$|\alpha| \pm 2 \left|\begin{matrix} T_{11} \\ T_{22} \end{matrix}\right|^{1/2} \qquad (11)$$

We can now try to put together:

i) The survey data necessary to estimate by TSLS the $k \cdot m$ relations of type (1) which must be arranged now in a suitable way. A matrix Z of (i) rows and (mn - 2) columns contains cost flows of i firms included in the survey. All values are taken as reciprocal. A column vector y is built up making before an LS estimate an instrumental variable of disposable resources in every firm and then summing up m times its reciprocal value. The error time vector is then accordingly modified and labelled as f.

$${}_iW_{lj} = a_{lj}\ {}_iX_j + {}_ie_{lj} \tag{11}$$

becomes

$$\frac{1}{{}_i\hat{\hat{W}}_j} = a_{lj} \frac{1}{{}_iX_{lj}} + {}_if_{lj} \tag{12}$$

${}_i\hat{W}_j$ being the instrumental variable substitute for $({}_iW_j)$.

Then defining $m\,{}_i\hat{W}_j = {}_iY_j$ we get:

$${}_iY_j = a_{lj}\ {}_iZ_{lj} + {}_if_{lj} \quad \text{for } l = 1,\ mn - 2 \tag{13}$$

which in vector notation becomes

$$\underset{[i \cdot 1]}{Y} \ \underset{\neq}{=} \ \underset{[i \cdot (mn-2)]}{Z} \ . \ \underset{[(mn-2) \cdot 1]}{a} + \underset{[i \cdot 1]}{f} \tag{14}$$

ii) The second ingredient of our new procedure is the prior information concerning both the (mk) coefficients belonging to costs and the (k) coefficients relative to final aggregated demand we have neglected till now. This is the second kind of prior information which is accommodated in the Λ matrix, in the α vector and in the T matrix; each one with (mn - 2) rows.

iii) Finally, we will use the new set of linear constraints on row and column totals (prior information of first kind) which has m + n - 1 elements because we now consider the final demand vector too.

The system which is going to be estimated - after the previous substitution of vector W with its instrumental variable estimate - looks like this:

$$\underset{[(i+mn-2) \cdot 1]}{\begin{bmatrix} Y \\ \alpha \end{bmatrix}} = \underset{[(i+mn-2)(mn-2)]}{\begin{bmatrix} Z \\ \Lambda \end{bmatrix}} \underset{[mn-2 \cdot 1)]}{a} + \underset{[(1+mn-2) \cdot 1]}{\begin{bmatrix} f \\ v \end{bmatrix}} \tag{15}$$

$$
\begin{bmatrix}
m_1\hat{W}_1^{-1} \\ m_2\hat{W}_1^{-1} \\ m_3\hat{W}_2^{-1} \\ m_4\hat{W}_2^{-1} \\ \vdots \\ m_i\hat{W}_1^{-1} \\ \hdashline \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{mn-2}
\end{bmatrix}
=
\begin{bmatrix}
{}_1X_{11}^{-1} & 0 & 0 & {}_1X_{21}^{-1} & 0 & 0 & {}_1V_1^{-1} & 0 & {}_1M_1^{-1} & 0 \\
{}_2X_{11}^{-1} & 0 & 0 & {}_2X_{21}^{-1} & 0 & 0 & {}_2V_1^{-1} & 0 & {}_2M_1^{-1} & 0 \\
0 & {}_3X_{12}^{-1} & 0 & 0 & {}_3X_{22}^{-1} & 0 & 0 & {}_3V_2^{-1} & 0 & {}_3M_2^{-1} \\
0 & {}_4X_{12}^{-1} & 0 & 0 & {}_4X_{22}^{-1} & 0 & 0 & {}_4V_2^{-1} & 0 & {}_4M_2^{-1} \\
\vdots & & & & & & & & & \vdots \\
{}_iX_{11}^{-1} & 0 & 0 & {}_iX_{21}^{-1} & 0 & 0 & {}_iV_1^{-1} & 0 & {}_iM_1^{-1} & 0 \\
\hdashline
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\vdots & & & & & & & & & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
*
\begin{bmatrix}
\frac{X_{11}}{\bar{W}_1} \\ \frac{X_{12}}{\bar{W}_2} \\ \frac{F_1}{\bar{W}} \\ \frac{X_{21}}{\bar{W}_1} \\ \frac{X_{22}}{\bar{W}_2} \\ \frac{F_2}{\bar{W}} \\ \frac{V_1}{\bar{W}_1} \\ \frac{V_2}{\bar{W}_2} \\ \frac{M_1}{\bar{W}_1} \\ \frac{M_2}{\bar{W}_2}
\end{bmatrix}
=
\begin{bmatrix}
f_1 \\ f_2 \\ f_3 \\ f_4 \\ \vdots \\ f_i \\ \hdashline v_i \\ v_2 \\ \vdots \\ v_{mn-2}
\end{bmatrix}
\qquad (15.1)
$$

subjected to the new set of constraints:

$$\underset{[(m+n-1)\ (mn-2)]}{\hat{R}} \quad \underset{[(mn-2)\cdot 1]}{a} = \underset{[(m+n-1)\cdot 1]}{\hat{r}}$$

It can be displayed as (15.1) and (16.1):

$$\begin{pmatrix} \frac{W_1}{W} & \frac{W_2}{W} & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{W_1}{W} & \frac{W_2}{W} & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{W_1}{W} & \frac{W_2}{W} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{W_1}{W} & \frac{W_2}{W} \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \end{pmatrix} * \begin{pmatrix} \frac{X_{11}}{W_1} \\ \frac{X_{12}}{W_2} \\ \frac{F_1}{W} \\ \frac{X_{21}}{W_1} \\ \frac{X_{22}}{W_2} \\ \frac{F_2}{W} \\ \frac{V_1}{W_1} \\ \frac{V_2}{W_2} \\ \frac{M_1}{W_1} \\ \frac{M_2}{W_2} \end{pmatrix} = \begin{pmatrix} \frac{W_1}{W} \\ \frac{W_2}{W} \\ \frac{V}{W} \\ \frac{M}{W} \\ 1 \\ 1 \end{pmatrix} \qquad (16.1)$$

An important feature of Z is the absence of survey-based data on sales to final sectors (F_i). It is a consequence of our hypothesis on the sample survey that is on inputs of the firms only.

On final demand coefficidnts F_i/W we then only have prior information. Every row of Z contains the reciprocals of costs of a specific firm included in the sample, while every element of y is a function of the value of its disposable resources estimated by instrumental variable technique and substituted for the empirical value.

It is also remarkable that the Z matrix should have displayed data of m-1 costs only in every sector because only m-1 relations or type (1) are independent in it.[9] We have in fact K constraints $\sum_i x_{ij} = x_j$ for j = 1,k which make independent only K(m-1) coefficients of the mk we want to estimate with sample data.

It is interesting to note that a negative covariance (α_i, α_j) is expected if an increase in α_i is likely accompanied by a decrease of α_j.

The exploitation of the extraneous information can be done in two different ways. If we utilize the Bayesian approach to inference, we can pool sample and prior information of the second kind to obtain posterior estimates of I-O coefficients which will modify our old prior information from that moment.

We think, however, that almost at the regional level and until the I-O modeling will produce a reliable background, today's prior information is too uncertain to be preferred to that collected with a good survey.

In our opinion it is then more convenient to use the classical approach. The method of estimation and reconciliation utilizes the principles of mixed estimation developed by Theil and Goldberger.[10]

Our approach utilizes extraneous information for control of sample information and for a rational reconciliation, but stress is on survey information.

---

[9] In this case, every element of the vector y is equal to $(m-1)\ \hat{W}^{-1}$.

[10] See Theil (1971), page 346.

VII. THE MIXED ESTIMATOR AND THE CONSTRAINED MIXED ESTIMATOR OF INPUT-OUTPUT COEFFICIENTS

Before turning to the solution for vector a we must underline the basic assumption of our approach.

If we put together every piece of available information for which reliability we are able to make a judgement, the result will be--in a sense--optimal because every I-O coefficient will become questionable, and in this case discarded or updated, only when fresh and comparable information will be available.

In our view, new information must be devoted to a projection, and not to a substitution for the old coefficients[11] and then be integrated with the old one. The latter will acting with a prior information in a process of the unending updating of coefficients.

We now go on with our problem assuming that the sample data have not been mended according to prior judgement and information; then follows a null covariance between errors in sample data and those of prior information as expected. We have then a diagonal inverse matrix of error covariances:

$$\left(\begin{bmatrix} f \\ v \end{bmatrix} [f \;\; v]'\right)^{-1} = \begin{bmatrix} \frac{1}{\sigma 2} V^{-1} & \\ & T^{-1} \end{bmatrix} \tag{17}$$

in which only T is certainly known.

A straightforward application of GLS to our system (15) then gives the following unbiased estimator of a vector of I-O coefficients:

$$\hat{a} = \left(\frac{1}{\hat{\sigma}^2} Z' \hat{V}^{-1} Z + \Lambda' \hat{T}^{-1} \Lambda\right)^{-1} \left(\frac{1}{\hat{\sigma}^2} Z' \hat{V}^{-1} y + \Lambda' \hat{T}^{-1} \alpha\right) , \tag{18}$$

$\hat{\sigma}^2$ and $\hat{V}$ are approximations, of course, of their unknown counterparts.

---

[11] Too often, input-outputers behave with their tables as if they were trying to fill a bottomless bucket.

In (18), σ^2 has been estimated from the LS residuals of y regressed on Z. It is not likely that the V matrix is the unity matrix because heteroscedasticity. A rather conventional hypothesis is to assume that the variance is equal to the square of output, up to the constant σ^2:

$$\sigma^2 \; V \cong \hat{\sigma}^2 \; \hat{V} \cong \hat{\sigma}^2 \; (x^i)^2 \quad .$$

The estimate $\hat{a}$ fails however, to satisfy the constraints (16); we must then resort to a GLS constrained estimator. This is an easy task indeed if we substitute in (10) the new mixed estimator $\hat{a}$ for the old one a and its variance:

$$\left(\frac{1}{\hat{\sigma}^2} \; Z' \; \hat{V}^{-1} \; Z + \Lambda' \; \hat{T}^{-1} \; \Lambda\right)^{-1} = \delta \quad ,$$

for its counterpart C^{-1}. We must further slightly modify the Ra = r relation because the new vector $\hat{\hat{a}}$ that is going to be estimated now contains k more coefficients for the presence of final aggregated demand. Substituting then (16.1), (18) and (19) for (10) we obtain:

$$\hat{\hat{a}} = \hat{a} + \delta \; \hat{R}' \; (\hat{R} \; \delta \; \hat{R}')^{-1} \; (\hat{r} - \hat{R}\hat{a}) \quad . \tag{20}$$

A few final comments are necessary: $\hat{\alpha}$ is the mixed but unconstrained GLS estimator of the (mn-2) I-O coefficients whose variance is equal to δ. This estimator involves the TSLS estimator of k(m-1) independent cost coefficients and its integration with the prior information on all the (mn-2) coefficients.

The factor $(\hat{r} - \hat{R}\hat{a})$ is the discrepancy with the extraneous information a constraint defined as $\hat{r}$ that is going to be demonstrated by the factor

$$\delta \; \hat{R}' \; (\hat{R} \; \delta \; \hat{R}') \quad , \tag{21}$$

which is a linear function of the covariance of the mixed estimator $\hat{a}$. If T = 0 the estimator (20) collapses to (10).

Repeated applications of the estimation procedure expounded here clearly gives use to a process "with memory" in which the last couple a and δ is a background (or statistical prior information) which following survey-based data can easily update.

VIII. A TENTATIVE AND PARTIAL APPLICATION TO THE VENETO REGION

The approach presented in Section III has been applied to the wood furniture sector in the Veneto region for 1976. The traditional problems encountered in the implementation of an I-O survey-based model are the determination of the statistical population of firms, the response rate and the quality or reliability of responses. We faced all these problems in our firm survey conducted during 1977 according to usual lines of I-O analysis.

The characteristics of the statistical population of that sector, in which the Veneto region is highly specialized, forced us to use a stratifical sample; the strata being the dimensional class of firms and area. The firms with less than 10 employees were not surveyed.

The problem of nonresponse proved deeply connected with the difficulty of the questionnaire. During the survey, we faced a trade-off between a larger response rate and a greater reliability and completeness of responses because every interview requested one day of one person. We decided then to pursue a greater reliability of responses in order to get a good test of the questionnaire in view of subsequent survey.

As a consequence, we got an ex post sample (of 32 very well compiled questionnaires), different in dimension and stratification from the ex ante sample.

With these data, we must estimate I-O column coefficient. According to the traditional deterministic approach we would have to compute per-employee coefficients within each sample strata and then we would have to multiply them by employees in the strata of the statistical population to obtain total flows. But actually, we do not need these flows, neither do we have the statistical population.

We want to use the stochastic approach because we have a sample after all and because I-O sector coefficients come out to be a good approximation for the sector, but not for individual firms [see (11)].

We cannot, however, use the LS estimator because the production of ${}_jX$ in firms and error terms ${}_je_i$ are correlated: X is measured with error and $\sum_i {}_jX_i = {}_jX$ (see Section III) from which follows [substitute (1)]: ${}_jX - \sum_i a_i {}_jX = \sum_i {}_je_i$ and then ${}_jX\ [1 - \sum_i a_i] = \sum_i {}_je_i$. From this it is apparent that only if ${}_iX$ is stochastic, we have $\sum_i {}_je_i \neq 0$ which means that error terms are independent of each other.

It is then necessary to choose a suitable estimator for a_i. We can use the instrumental variable technique which imply the substitution for ${}_jX$ of an estimate some new variables uncorrelated (almost in the limit) with ${}_je_i$.

We can do this if we think about the way in which the questionnaire is filled out by firms. The turnover is of course their basic starting point. This variable is then assumed as measured without error. From an economic point of view sales are a function of many other demand variables, but in this context we consider it as exogenous. It follows then that it is fixed.

The basic information for the I-O table is production, however. This concept is not exactly familiar to firms. But they can easily calculate it after an evaluation of their inventories variation.

With the production ${}_jX$, every production cost is now determined; depreciation and profit included. It is clear that the evaluation of depreciation, profit, inventories and production is interlinked and quite uncertain.

These four variables are then not only endogenous but also affected by errors.

We can now write down our complete simultaneous model[12] for the column of I-O flows of the sector as follows:

[12] It is quite similar to that used by Gerking (1976) but with a quite different hypothesis on the error structure.

$$\begin{cases} {}_jX_1 = a_1 \; {}_jX + {}_je_1 \\ \cdot \\ \cdot \\ \cdot \\ {}_jX_K = a_K \; {}_jX + {}_je_K \\ {}_jX_{K+1} = a_{K+1} \; {}_jX + {}_je_{K+1} \\ {}_jX_{K+2} = a_{K+2} \; {}_jX + {}_je_{K+2} \\ {}_jX - \sum_{i=1}^{K+2} {}_jX_i = a_{K+3} \; {}_jX + {}_je_{K+3} \\ {}_jX = {}_jS + {}_jI \\ {}_jI = f \left[{}_jL; \; {}_jX_{K+1}; \; ({}_jX - \sum_{i=1}^{K+2} {}_jX_i) \right] \end{cases} \tag{22}$$

${}_jX_i$ = cost of input i, $i = 1,K$;

${}_jX_{K+1}$ = depreciation;

${}_jX_{K+2}$ = labor cost;

${}_jX_{K+3}$ = gross profit;

${}_jS$ = sales;

${}_jI$ = inventories

${}_jX$ = production

One of the first K+3 cost equations is linearly dependent from the other.

If we assume as rigid the number of employees L (almost in the short run), as well as measured without error, it can be considered a predetermined and fixed variable. We can then use, ${}_jL$ and ${}_jS$ as instrumental variables for ${}_jX$.

The following table shows the results obtained in the traditional way (col. 1) and with the TSLS estimator (col. 2) of the first K+2 equations of system (22). This estimator uses as

Table 1. Wood Furniture Production Percentage Costs, Veneto, 1976.

	NACE		(1) RATIO TO PRODUCTION	(2) TSLS	(3)* STANDARD ERRORS	(4)* R^2
1	45	FURNITURE, WOOD PRODUCTS	21.626	23.084	.754	.98
2	17	CHEMICALS	3.849	3.044	.348	.72
3	19	METAL PRODUCTS	4.112	5.401	.612	.82
4	41	TEXTILES	2.292	3.127	.725	.51
5	15	GLASS, MIRRORS	.716	.910	.165	.63
6	49	RUBBER, PLASTICS	1.058	1.611	.241	.76
7	47, 51	PAPER, PRINTING & PUBLISHING, OTHER MANUF.	8.342	6.521	1.305	.48
8	07	PETROLEUM	.441	.404	.014	.98
9	09	ELECTRIC ENERGY, GAS, WATER	.626	.593	.027	.96
10	--	TOTAL, RAW MATERIALS	43.062	44.695	---	---
11	73	LOCATION	.155	.261	.033	.81
12	57,59, 61	COMMERCE, PUBLIC SERVICES, TRANSPORTS	3.277	4.907	.509	.86
13	65, 67	OTHER SERVICES, COMMUNICATIONS	3.136	3.678	.173	.97
14	69, 71	CREDIT, INSURANCE, SERVICES TO PRODUCTION	11.010	10.789	.904	.99
15	--	TOTAL, INTERMEDIATE INPUTS	60.640	64.330	---	---
16	95,96	WAGES (LABOR COST)	20.651	17.477	.946	.94
17	95,96	SALARIES (LABOR COST)	5.738	5.493	.302	.95
18	97	COST OF CAPITAL	6.147	7.107	.399	.95
19	98	DEPRECIATION	3.745	3.298	.248	.89
20	97	REST OF VALUE ADDED	3.079	2.291	.632	.21
21	--	VALUE ADDED	39.360	36.666	---	---
22	--	PRODUCTION	100.000	99.996	---	---

*For Col. (2) only.

instrumental variables the number of employees L and the volume of sales which are predetermined and fixed variables. The three columns in this table refer to column 2.

The values of columns 1 and 2 are only probable values of unknown I-O coefficients of our sector. These estimates are unevenly reliable or precise even if their sum is unity.

If we utilize an econometric technique, as we did for column 2, we can calculate the standard error of every coefficient with which we can build a confidence interval. We will say, as usual, that the method used produces one interval which contains with a given probability of error of 0.95, the true value of the coefficient of wood e.g. in the interval 23.084 ± $t_{0.05}$.754.

If we run down columns 2 and 3 we can see how such intervals always leave out negative values. In I-O models we look for coefficients greater or equal to zero only, or equivalently, we look for standard errors not too large, comparatively, to the coefficients.

If a coefficient's estimates turn out to be insignificant, we cannot conclude that its cost is low, but that sample data are not adequate for good estimates. The estimated coefficient can be high indeed, but the true value can be very different if its confidence interval is too large.

It follows that more information is needed on the cells for which estimates are less precise and on which inferior is the homogeneity between firms.

The intensity by which the errors terms operates cell to cell or, if we prefer, the extent by which the variance in cost is explained by that of production varies from cell to cell.

We can realize this if we look at column 4 in which the R^{12} coefficients show a range from .217 to .991. As a conclusion, if these coefficients were submitted to the procedures of sections from V to VII those less precise were more heavily affected by reconciliation and prior information.

REFERENCES

Allen, W., and W. Gossling, eds. (1975) Estimating and Projecting Input-Output Coefficients. Londond: Input-Output Publishing Company.

Brody, W., and A. Carter, eds. (1972) Input-Output Techniques. Amsterdam: North-Holland Publishing Company.

Fisher, W., and C. Chilton (1972) Developing Ex-Ante Input-Output Flows and Capital Coefficients. Input-Output Techniques. Amsterdam: North-Holland Publishing Company.

Fisher, W. (1972) Ex-Ante as a Supplement or Alternative to rAs in Updating of Input-Output Coefficients. Estimating and Projecting Input-Output Coefficients. London: Input-Output Publishing Company.

Gerking, S. (1976) Estimation of Stochastic Input-Output Models. Leiden: Martimus Nijhoff.

Jensen, R.C., and D. McGaurr (1976) Reconciliation of Purchases and Sales Estimates in an Input-Output Table. Urban Studies, 13.

Klein, L. (1974) A Textbood of Econometrics. New York: Prentice Hall Inc.

Lecomber, J. (1975) A Critique of Methods of Adjusting, Updating and Projecting Matrices. Estimating and Projecting Input-Output Coefficients. London: Input-Output Publishing Company.

Martellato, D. (1978) Metodi per la stima stocastica dei modelli Input-Output. Paper presented at an Input-Output Conference held in Venice, 1976; see Costa, P. Interdispendenze industriali e programmazione regionale, F. Angeli, Milano, 1978.

Streit, M. (1979) Calculation of the Margin of Expert Errors in Input-Output Models. Paper presented at the Seventh International Conference on Input-Output Techniques, Innsbruck, Austria.

Theil, H. (1971) Principles of Econometrics. New York: J. Wiley and Sons.