Reconciling "Rows Only" and "Columns Only" Coefficients in an Input-Output Model

Shelby D. Gerking[1]

Department of Economics
Arizona State University
Tempe, Arizona 85281 U.S.A.

1. Introduction

Since Leontief's pioneering work of the 1940's, there has been a proliferation of input-output studies conducted at the regional level, largely because regional scientists have found that the input-output approach is a useful analytical device for examining a wide range of problems. Some have even claimed that for economic forecasting, input-output is superior to competing techniques. For example, Richardson [1972, 157] has stated:

> In the absence of further theoretical advances and the provision of more data the use of more complex econometric forecasting techniques is not yet practicable, . . . and input-output models are probably the most useful forecasting tool currently available.

However, even the strongest proponents of input-output would surely admit that this tool is far from perfect. Nevertheless, there has been a tendency for regional scientists to use input-output analysis without measurably contributing to a refinement of the technique. In addition, the literature contains few theoretical discussions aimed at overcoming or minimizing certain practical problems, such as coefficient estimation, reconciliation, and sample-size determination, which are faced in virtually all input-output studies that are based on a survey.

The purpose of this paper, therefore, is to conduct an in-depth examination of one of these practical problems: reconciling "rows only" and "columns only" estimates of regional coefficients in order to produce a single input-output table.[2] Specifically, this paper describes a systematic reconciliation procedure based on the econometric theory of estimating linear equations using

[1] I thank D.J. Behling, R.J. Green, R.L. Pfister, S. Pleeter, and two anonymous referees for their helpful comments. In addition, thanks are due to W. Miernyk for providing the data used in Section 4. Financial support was received from the Division of Research and Office of Research and Advanced Studies, Indiana University; research support, from the Bureau of Business and Economic Research, Arizona State University. I alone am responsible for remaining errors.

[2] Although the discussion in this paper is cast in terms of regional coefficients, all of the results can be easily extended to cover the reconciliation of technical coefficients. In constructing regional coefficients, it may be recalled, it is necessary to distinguish between "domestic" and "foreign" intersectoral transactions while the technical coefficients are constructed in order to reflect the total requirements of the output of each sector being absorbed by each other sector regardless of the region in which these inputs are produced.

Manuscript received March, 1976; revision received June, 1976.

instrumental variables. The procedure is described in detail in Section 3 and is illustrated in Section 4, using input-output data from West Virginia. Some introductory material on input-output analysis and the reconciliation problem is provided in Section 2.

2. Input-Output Analysis and the Reconciliation Problem

Typically, three assumptions are made in input-output analysis: (1) the economy can be meaningfully divided into a finite number of sectors, each of which produces a single homogeneous product; (2) there are neither economies nor diseconomies of scale in production; and (3) the level of output in each sector uniquely determines the quantity of each input that is purchased [Chenery and Clark 1959, 33-42]. Taken together, these assumptions imply that the production function for any sector may be expressed as:

$$XT'_j = \min\left[\frac{ZT'_{1j}}{\alpha'_{1j}}, \ldots, \frac{ZT'_{mj}}{\alpha'_{mj}}, \frac{ZT'_{m+1,j}}{\alpha'_{m+1,j}}, \frac{ZT'_{m+2,j}}{\alpha'_{m+2,j}}, \frac{VT'_{1j}}{\tau'_{1j}}, \ldots, \frac{VT'_{nj}}{\tau'_{nj}}\right] \tag{2.1}$$

where

XT'_j = total quantity of output in sector j.
ZT'_{ij} = total quantity of goods and services transferred from domestic sector i to sector j, $i = 1, \ldots, m$.
$ZT'_{m+1,j}$ = total quantity of a homogeneous labor service purchased by sector j from households.
$ZT'_{m+2,j}$ = total quantity of a homogeneous public service purchased by sector j from governmental agencies.
VT'_{ij} = total quantity of various types of inputs purchased by sector j from outside the geographic boundaries of the economy in question, $i = 1, \ldots, n$.
α'_{ij} = regional coefficient interpreted as the minimum quantity of output from domestic sector i required to produce one unit of output in sector j; where $\alpha'_{ij} > 0, i = 1, \ldots, m + 2, j = 1, \ldots, m$.[3]
τ'_{ij} = trade coefficient interpreted as the minimum quantity of import i required to produce one unit of output in sector j; $\tau'_{ij} > 0$, $i = 1, \ldots, n$, $j = 1, \ldots, m$.[4]
min (a, b, . . . , z) = minimum of the elements (a, b, . . . , z).

From equation (2.1), estimates of the regional coefficients can be obtained from a relation such as

$$ZT'_{ij} = \alpha'_{ij}XT'_j \tag{2.2}$$

However, data on ZT'_{ij} and XT'_j have seldom been available to input-output analysts. As a result, in empirical interindustry models, equation (2.2) is

[3] If $\alpha'_{ij} = 0$ or if $\beta'_{ij} = 0$, the appropriate ratio is understood to be deleted from equation (2.1).
[4] Problems in estimating the τ'_{ij} are ignored in this paper.

usually redefined in value terms. This implies

$$ZT_{ij} = \alpha_{ij}XT_j \tag{2.3}$$

where ZT_{ij} is the value of goods and services transferred from i to j; XT_j is total value of output in j; and α_{ij}, which is still referred to as a regional coefficient, is interpreted as the minimum value of output in domestic sector i required to produce one dollar's worth of output in sector j. Finally, the regional coefficients in equation (2.3) are generally calculated by forming the ratio $\alpha_{ij} = ZT_{ij}/XT_j$.

The reconciliation problem arises because there are two ways of observing the ZT_{ij}. First, the sales of firms in sector i to firms in sector j may be examined. If the ZT_{ij} are measured in this way, then equation (2.3) produces the so-called "rows only" estimate of α_{ij}. Alternatively, data may be collected concerning the purchases by firms in sector j from firms in sector i. This information, when substituted into (2.3) yields "columns only" estimates for the regional coefficients.[5]

For a given regional coefficient, it would be most improbable if these two estimates were identical. In fact, there are at least two important reasons why they might differ. For example, input-output data are sometimes obtained from a nonexhaustive sampling of the firms within each sector. In this case, there is obviously no reason why the total sales to sector j by the included firms in sector i must equal the total purchases from firms in sector i by the included firms in sector j. Furthermore, even if exhaustive samples are taken, there may be errors in the transactions data. For example, these errors may be due to sectoral classification errors by respondents, a lack of information on the part of respondents about the location of producers from whom they are purchasing, or simply slips of the pen in transcribing data. (A formal description of these errors is provided in Section 3.)

In some regional input-output studies, data on sales and purchases are collected from each firm so that both the "rows only" and "columns only" coefficients can be calculated. Clearly, this is more costly than obtaining data on sales or purchases alone. However, the additional information could be used to improve the reliability of the resulting estimates of the regional coefficients. Nevertheless, the potential gain from having both "rows only" and "columns only" coefficients probably has never been realized. Three examples drawn from the input-output studies conducted by Bourque, Miernyk, and Jensen and McGaurr will show why this is true.

First, consider the study of Washington State conducted in 1967 by Bourque and others. They stated that in many of the sectors surveyed, the discrepancy between the "rows only" and "columns only" estimates of the ZT_{ij} was significant. Hence, it was necessary to find a way to combine the two sets of data on intersectoral flows. In order to do this, they reported:

> . . . Each member of the study team met independently with each other member, compared sources, made judgments about reliability, conducted additional field work when necessary, and solved the remaining differences by trading or compromise [Bourque 1967, 6].

[5]Some regional input-output studies make use of only one type of estimate. For an example of a study using only "rows only" estimates, see Hansen and Tiebout [1963]. For an example of the exclusive use of "columns only" coefficients, see Isard and Langford [1971].

This approach to the reconciliation problem leaves much to be desired. It is unsystematic and would be virtually impossible to replicate. In fact, Isard and Langford [1971, 62] called the procedure unscientific and likened it to "... a meeting over the kitchen table." It is worth noting, though, that Isard and Langford failed to suggest how the reconciliation process might be improved.

Miernyk's study in West Virginia provides the second illustration of how the reconciliation problem has been handled in a practical setting [Miernyk 1970, 18]. As in Bourque's study, Miernyk obtained "rows only" and "columns only" estimates of the intersectoral flows. Then, for both estimates of each ZT_{ij}, Miernyk constructed what he called "reliability quotients." These were based on considerations such as the: (1) fraction of total sectoral sales accounted for by the sample; (2) homogeneity of output within the sector; (3) judgment of interviewers who collected the data; (4) "representativeness" of the sample; and (5) reliability of the sector control total. Finally, he used these quotients to make a judgement as to which of the two estimates was the more reliable for each ZT_{ij}.

Although Miernyk was the first regional scientist to incorporate a measure of reliability into the input-output reconciliation process, his contribution is of questionable value because the term "reliability" was never adequately defined and the reliability quotients were, in part, subjectively determined. As a consequence, there is a real question about what these quotients are measuring. For example, does the reliability of an input-output estimate refer to its mean, variance, its mean and variance, or to something else?

The third, and final, example dealing with input-output reconciliation is drawn from a recent paper by Jensen and McGaurr [1976]. These authors recommend a similar approach to the one Miernyk used. In particular, they suggest that the "rows only" and "columns only" measurements on the ZT_{ij} should be combined, using subjectively determined reliability weights in order to obtain a single table of intersectoral transactions. However, Jensen and McGaurr recognized that if the reconciled estimates were determined in this way, these two accounting identities will almost certainly be violated:

$$XT_i = \Sigma_j ZT_{ij} + FT_j \tag{2.4}$$

where FT_j denotes final demand in sector j and

$$XT_j = \Sigma_i ZT_{ij} + \Sigma_i VT_{ij}. \tag{2.5}$$

Consequently, they developed an RAS-type adjustment procedure for the reconciled ZT_{ij}, to satisfy the constraints imposed by equations (2.4) and (2.5).[6]

The Jensen and McGaurr approach to the reconciliation problem is subject to the same basic criticisms as those directed at the West Virginia Study. Again, the term "reliability" was never defined and the reliability weights were determined subjectively. However, Jensen and McGaurr did allow the reliability weights to take on any value on the zero-one interval, rather than restricting them to be either zero or one. In addition, they have provided an interesting application of the RAS method for adjusting the reconciled ZT_{ij}.

[6]For a more complete discussion of the RAS adjustment method see Bacharach [1970, 27-30].

3. A New Reconciliation Procedure

The previously described problems of defining and quantifying the term "reliability" will be addressed here through the use of econometric theory. Specifically, a reconciled estimator is defined to be reliable if it has the smallest variance within the class of consistent estimators.[7] As will become apparent, this concept of reliability is applied to reconciled estimates of the regional coefficients. This should be contrasted with the approaches of the three previously discussed studies, which sought to reconcile estimates of the ZT_{ij}.

For expository purposes, this section is organized into three parts. The first one contains a review of three instrumental variables estimators that can be used to estimate the asymptotic mean and variance of the "columns only" regional coefficients. In the second part, these results will be extended to obtain the same measures for the "rows only" estimator. Part three, then, contains some remarks about an appropriate method for obtaining a reconciled estimator.

THE "COLUMNS ONLY" ESTIMATOR

To obtain the "columns only" estimator for the regional coefficients using instrumental variables, retain the standard assumptions of input-output analysis listed in Section 2. Next, assume that the following two-side conditions hold for all sectors: (1) all firms in each sector have identical production functions and (2) the inputs and outputs of each firm can be measured only with error.[8] Taken together these assumptions imply that

$$ZT_{ij}(k) = \alpha_{ij}XT_j(k) \tag{3.1}$$

and that

$$Z_{ij}(k) = ZT_{ij}(k) + \epsilon_{ij}(k) \tag{3.2}$$

$$X_j(k) = XT_j(k) + \nu_j(k), \tag{3.3}$$

where the index k refers to the k^{th} firm in sector j; $\epsilon_{ij}(k)$ and $\nu_j(k)$ are each independently and identically distributed random variables with mean of zero for all k; and $Z_{ij}(k)$ and $X_j(k)$ are the measured counterparts of the true and unobservable $ZT_{ij}(k)$ and $XT_j(k)$. Substituting (3.2) and (3.3) into (3.1) produces:

$$Z_{ij}(k) = \alpha_{ij}X_j(k) + \theta_{ij}(k) \tag{3.4}$$

where $\theta_{ij}(k) = \epsilon_{ij}(k) - \alpha_{ij}\nu_j(k)$.

As is well known, applying ordinary least squares (OLS) to equation (3.4) causes the resulting estimates of the α_{ij} to be biased and inconsistent because $X_j(k)$ will be correlated with $\theta_{ij}(k)$, even asymptotically. Elsewhere [Gerking 1976a], however, the argument has been made that consistent estimates of the

[7] This definition is clearly arbitrary as the term "reliability" has no standard statistical interpretation. However, the definition offered here does not seem to strain credibility. In addition, it will prove to be useful in deriving an appropriate reconciled estimator.

[8] A more complete discussion of appropriate estimation strategies for the "columns only" coefficients is contained in Gerking [1976a]. This paper also devotes more space to a discussion of the assumption of identical production functions among firms in each sector.

α_{ij} may be obtained by one of several instrumental variable techniques.[9] Three of these techniques are described briefly below.

The Wald-Bartlett method. This estimator for the α_{ij} may be defined as:

$$a_{ij}(WBM) = [\Sigma_k Z_{ij}(k) W_j(k) / \Sigma_k X_j(k) W_j(k)] \tag{3.5}$$

where

$$W_j(k) = \begin{cases} 1 \text{ if } X_j(k) > \text{med } X_j(k) \\ 0 \text{ if } X_j(k) = \text{med } X_j(k) \\ -1 \text{ if } X_j(k) < \text{med } X_j(k) \end{cases} \tag{3.6}$$

and where med $X_j(k)$ denotes the sample median of the observations on the variable $X_j(k)$ [Wald 1940, 284-300]. It can be shown that a_{ij} (WBM) is consistent if the values assigned to $W_j(k)$ in equation (3.6) are identical with those that would have been assigned had observations on $XT_j(k)$ been available. However, Bartlett [1949, 207-212] has demonstrated that the asymptotic variance of a_{ij} (WBM) is quite large and that its efficiency may be improved by: (1) ranking the $X_j(k)$ by size; (2) deleting the middle third of the observations from the sample; and (3) applying Wald's method to the remaining observations.

Durbin's method. Durbin [1954, 23-32] has proposed another instrumental variable estimator. When applied to the "columns only" regional coefficients, it may be expressed as:

$$a_{ij}(DM) = [\Sigma_k Z_{ij}(k) D_j(k) / \Sigma_k X_j(k) D_j(k)] \tag{3.7}$$

In equation (3.6), $D_j(k)$ equals k, assuming that the $X_j(k)$ have been ranked in ascending order by size. This estimator is consistent if the ranking of the unknown $XT_j(k)$ is identical with the ranking of $X_j(k)$. Furthermore, $a_{ij}(DM)$ generally is more efficient than $a_{ij}(WBM)$.

The two-stage least squares (2SLS) method. A third instrumental variable estimator may be obtained by estimating the α_{ij} column by column. For the jth column, the relevant system of equations is given below.

$$X_j(k) = \Sigma_i Z_{ij}(k) + RV_j(k) + WS_j(k) + PG_j(k) \tag{3.8.1}$$

$$Z_{1j}(k) = \alpha_{1j} X_j(k) + \theta_{1j}(k) \tag{3.8.2}$$

$$\vdots$$

$$Z_{mj}(k) = \alpha_{mj} X_j(k) + \theta_{mj}(k)$$

$$RV_j(k) = \alpha_{m+1,j} X_j(k) + \theta_{m+1,j}(k) \tag{3.8.3}$$

In equation system (3.8), most of the relations need no further explanation as they are identical in form to the one in (3.4). However, equations (3.8.1) and (3.8.3) deserve further comment. Equation (3.8.1) is an accounting identity stating that the measured total output for any firm must be distributed to the firms in the m endogenous sectors or to value added. As can be seen, value added has been broken down into three components: $RV_j(k)$ + $WS_j(k)$ +

[9] Admittedly, the term consistency is used here and throughout the remainder of this paper in a somewhat unconventional way since the number of observations drawn from a given sector cannot become indefinitely large. Instead, the number of observations from a sector can only approach the total number of firms in that sector.

PG(k). The first component, $RV_j(k)$, represents that part of measured value added which is determined as a residual from $X_j(k)$ and $Z_{ij}(k)$. Clearly, items falling into this category, such as profits, will be subject to measurement error. The second component, $WS_j(k)$, denotes wages and salaries. The third, $PG_j(k)$, represents certain payments to government, such as property taxes. It seems plausible to assume that the last two components can be measured without error, since firms are likely to keep accurate records of wages and salaries and tax payments, especially of wages and salaries since this variable determines federal income tax withholdings and social security contributions. Finally, the last equation (3.8.3) is included to take account of the decomposition of value added.

Since each of the last m + 1 equations in equation system (3.8) is identified (in fact, each is just identified), 2SLS may be used to estimate the regional coefficients. The 2SLS estimate of α_{ij} is:

$$a_{ij}(2SLS) = \frac{X_j^T Q_j (Q_j^T Q_j)^{-1} Q_j^T Z_{ij}}{X_j^T Q_j (Q_j^T Q_j)^{-1} Q_j^T X_j} \tag{3.9}$$

where X_j and Z_{ij} are $n_j \times 1$ vectors containing the $X_j(k)$ and the $Z_{ij}(k)$ and where Q_j is an $n_j \times 2$ matrix composed of the $WS_j(k)$ and the $PG_j(k)$. Preliminary work [Gerking, 1976a] has indicated that the asymptotic sampling variance of the 2SLS estimator for the α_{ij} tends to be somewhat smaller than for either $a_{ij}(WBM)$ or $a_{ij}(DM)$.

THE "ROWS ONLY" ESTIMATOR

The three instrumental variable techniques just given for estimating the "columns only" coefficients can be modified to obtain estimates for the "rows only" counterparts. A description of these modifications is given next. The discussion will focus on a method of estimating the α_{ij} when intersectoral transactions are measured by observing the sales to sector j by firms in i. The resulting regression equations will be shown to have much the same form as equation (3.4). As in the previous section, it will be argued that in each of the regression equations, the explanatory variable is likely to be correlated with the disturbance term. Hence, instrumental variable techniques are recommended.

To begin the derivation of a "rows only" analogue for equation (3.4), equation (3.1) must be summed over all firms in sector j and divided by $\sum_{k}^{N_j} XT_j(k)$ to obtain:

$$\alpha_{ij} = \left[\sum_{k}^{N_j} ZT_{ij}(k) \Big/ \sum_{k}^{N_j} XT_j(k) \right] \tag{3.10}$$

where N_j denotes the total number of firms in sector j. Next, recall that $ZT_{ij}(k)$ is interpreted as the true purchases of the kth firm in sector j from firms in sector i. This implies that

$$\sum_{k}^{N_j} ZT_{ij}(k) = \sum_{k}^{N_i} ST_{ij}(k) \tag{3.11}$$

where $ST_{ij}(k)$ represents the true sales by the k^{th} firm in sector i to the firms in sector j. Hence,

$$\alpha_{ij} = \left[\sum_{k}^{N_i} ST_{ij}(k) \Big/ \sum_{k}^{N_j} XT_j(k) \right]. \tag{3.12}$$

Next, multiplying both sides of (3.12) by the ratio of the true total output in sector j to the true total output in sector i yields

$$\left[\sum_{k}^{N_i} ST_{ij}(k) \Big/ \sum_{k}^{N_i} XT_i(k) \right] = \alpha_{ij} \left[\sum_{k}^{N_j} XT_j(k) \Big/ \sum_{k}^{N_i} XT_i(k) \right] \tag{3.13}$$

Therefore, if it is assumed that each firm in i obeys equation (3.13), $ST_{ij}(k)$ can be expressed as:

$$ST_{ij}(k) = \alpha_{ij} \left[\sum_{k}^{N_j} XT_j(k) \Big/ \sum_{k}^{N_i} XT_i(k) \right] XT_i(k) \tag{3.14}$$

This equation specifies that all firms in sector i sell a constant fraction of their output to firms in sector j. As is reasonable, this fraction varies directly with both the i, j^{th} regional coefficient and the level of output in sector j. The fraction also varies inversely with the level of output of firms in sector i.

In order to obtain an estimating equation from (3.14), suppose that the sales of any firm are subject to measurement error, according to:

$$S_{ij}(k) = ST_{ij}(k) + \xi_{ij}(k) \tag{3.15}$$

where $S_{ij}(k)$ represents the observed sales of the k^{th} firm in sector i to firms in sector j and $\xi_{ij}(k)$ is a random disturbance term that is independently and identically distributed with the zero mean for all k. Substituting (3.3) and (3.15) into (3.14) produces:

$$S_{ij}(k) = \alpha_{ij} \left[\sum_{k}^{N_j} XT_j(k) \Big/ \sum_{k}^{N_i} XT_i(k) \right] X_i(k) + \eta_{ij}(k) \tag{3.16}$$

where $\eta_{ij}(k) = \xi_{ij}(k) - \alpha_{ij} \left[\sum_{k}^{N_j} XT_j(k) \Big/ \sum_{k}^{N_i} XT_i(k) \right] v_i(k)$.

Finally, since the quantity $\left[\sum_{k}^{N_j} XT_j(k) \Big/ \sum_{k}^{N_i} XT_i(k) \right]$ is a constant, define

$$\beta_{ij} = \alpha_{ij} \left[\sum_{k}^{N_j} XT_j(k) \Big/ \sum_{k}^{N_i} XT_i(k) \right] \tag{3.17}$$

so that

$$S_{ij}(k) = \beta_{ij} X_i(k) + \eta_{ij}(k) \tag{3.18}$$

Equation (3.18), then, implies that estimates of the α_{ij} can be obtained if the β_{ij} can be estimated and if the ratio $\left[\sum_{k}^{N_j} XT_j(k) \Big/ \sum_{k}^{N_i} XT_i(k) \right]$ is known.

As will be demonstrated, the first of these obstacles is not difficult to overcome. In fact, the β_{ij} can be estimated by employing instrumental variable methods similar to those used in obtaining the "columns only" estimates. However, the second condition, that the ratio $\left[\sum_{k}^{N_j} XT_j(k) \Big/ \sum_{k}^{N_i} XT_i(k)\right]$ must be known, requires further discussion. This is momentarily postponed though, so that the estimation procedures for the β_{ij} may be fully explained.

In equation (3.18), OLS will produce inconsistent estimates of the β_{ij}, since $X_i(k)$ and $\eta_{ij}(k)$ are likely to be asymptotically correlated. Nevertheless, it is possible to obtain consistent estimators for these coefficients by using instrumental variable methods. For example, the Wald-Bartlett and Durbin methods could be applied. Also, consistent estimators for the β_{ij} may be obtained by 2SLS although some additional explanation is required. To show how this can be done, consider the equation system (3.19):

$$X_i(k) = \Sigma_j S_{ij}(k) + SH_i(k) + SG_i(k) + RF_i(k) \tag{3.19.1}$$

$$S_{i1}(k) = \beta_{i1} X_i(k) + \eta_{i1}(k) \tag{3.19.2}$$

$$\vdots$$

$$S_{im}(k) = \beta_{im} X_i(k) + \eta_{im}(k).$$

As can be seen, the equations in system (3.19) are similar in form to those in system (3.8). Equation (3.19.1) is an accounting identity stating that total output of firm k in sector i must be sold to firms in the m endogenous sectors or to final demand. Final demand is written as the sum of three components: $SH_i(k) + SG_i(k) + RF_i(k)$, where $SH_i(k)$ equals sales to households, $SG_i(k)$ equals sales to government, and $RF_i(k)$ equals all remaining sales to final demand including sales to other firms on capital account and exports. The remaining m equations in system (3.19) are identical in form to those in system (3.18).

The system of equations in (3.19) will be useful in estimating the β_{ij} if at least one of the three components of final demand can be measured without error—a condition that may hold in some sectors. For example, due to various reporting requirements, a firm's sales to government at the state and federal levels may be measured exactly. In addition, sales to households for a particular firm may be accurately computed from its data on tax collections from retail sales. Error-free measurements on exports and sales to other firms on capital account, though, will probably be much more difficult to obtain. However, this problem is similar to the one that arose in connection with measuring value added in (3.8), and can be dealt with by adding another equation to (3.19) to take this decomposition into account.

These arguments concerning the quality of measurement on the three components of final demand are not intended as general statements about all input-output sectors. For example, some firms making sales to households may sell goods and services that are exempt from retail sales tax. In addition, collection on government sales contracts may lag behind the actual transfer of goods. These and other measurement problems must be handled sector by sector. Finally, it is also possible that the observed firms in some input-output sectors do not make sales in any of the final-demand categories. If so, 2SLS estimation of the β_{ij} would not be feasible. Another of the instrumental variable techniques would have to be used.[10]

In any event, if two components of final demand can be measured free of error, say $SH_i(k)$ and $SG_i(k)$, the β_{ij} can be estimated by 2SLS since each of the estimable relations is identified. The 2SLS estimate of β_{ij} is given by

$$b_{ij}(2SLS) = \frac{X_i^T M_i (M_i^T M_i)^{-1} M_i^T S_{ij}}{X_i^T M_i (M_i^T M_i)^{-1} M_i^T X_i} \tag{3.20}$$

where X_i and S_{ij} are $n_i \times 1$ vectors containing the n_i observations on $X_i(k)$ and $S_{ij}(k)$ while M_i is an $n_i \times 2$ matrix containing the observations on $SH_i(k)$ and $SG_i(k)$, the two variables which are assumed to be measured exactly.

Now that an estimator for the β_{ij} has been obtained, the corresponding estimates for the α_{ij} must be derived. To do this, recall that from the discussion preceding (3.17), α_{ij} can be expressed as

$$\alpha_{ij} = \beta_{ij}\left[\sum_{k}^{N_i} XT_i(k) \Big/ \sum_{k}^{N_j} XT_j(k)\right] = \beta_{ij}\gamma_{ij} \tag{3.21}$$

where

$$\gamma_{ij} = \left[\sum_{k}^{N_i} XT_i(k) \Big/ \sum_{k}^{N_j} XT_j(k)\right].$$

As a result, the derivation in question hinges on whether or not γ_{ij} is known.

Strictly speaking, the ratio γ_{ij} probably will not be known exactly. However, observed analogues for its two components are always calculated in input-output studies. In fact, these analogues, usually called control totals, appear in the margin of an input-output table. Obviously, these values cannot be obtained from observing the output of firms included in a nonexhaustive sample. Instead, they are usually computed from outside sources. For example, in his study, Miernyk [1970, 17] obtained the control totals from the West Virginia State Tax Commission data, together with information provided by the West Virginia Department of Employment Security.

These measured control totals may be used to approximate the γ_{ij}. To see why, note that the control totals must be measured by summing the observed total output of each firm in each sector. Further, for the sake of simplicity, assume that these measurements are obtained according to equation (3.3).[11] Denoting the observed counterpart of γ_{ij} by c_{ij} then,

$$\begin{aligned} c_{ij} &= \left[\sum_{k}^{N_i} X_i(k) \Big/ \sum_{k}^{N_j} X_j(k)\right] \\ &= \sum_{k}^{N_i} [XT_i(k) + \nu_i(k)] \Big/ \sum_{k}^{N_j} [XT_j(k) + \nu_j(k)] \end{aligned} \tag{3.22}$$

[10]This problem actually arose among the observed firms in several of the West Virginia input-output sectors.
[11]Obviously, these measurements may not be obtained according to equation (3.3). However, making this explicit would require more notation and would not affect the results.

Therefore, if N_i and N_j are of even moderate size, c_{ij} will be approximately equal to the constant γ_{ij}, because the sum of the observation errors should be small in relation to the true total output for both sectors. Based on this view, then, a consistent "rows only" estimator for α_{ij} constructed by 2SLS becomes

$$a_{ij}(2SLS) = c_{ij}b_{ij}(2SLS) \tag{3.23}$$

Further, the variance of this estimator may be approximated by

$$VAR[a_{ij}(2SLS)] = c_{ij}^2 \cdot VAR[b_{ij}(2SLS)] \tag{3.24}$$

THE RECONCILED ESTIMATOR

In the previous parts of this section, three instrumental variable methods were described for obtaining estimates of both the "columns only" and the "rows only" regional coefficients. As a result, it only remains to derive an appropriate reconciliation procedure. Actually, this task is straightforward if the criterion of minimum variance is used. In particular, assume that consistent "rows only" and "columns only" estimates of the α_{ij} have been constructed using the instrumental variable technique which has the smallest asymptotic sampling variance. (As noted previously, the preliminary work on estimating the "columns only" coefficients indicates that this technique is likely to be 2SLS.) Then, denote the "rows only" and "columns only" estimates of the ij^{th} regional coefficient as a_r and a_c. Since a_r and a_c are consistent estimators of α_{ij}, a consistent and reconciled estimator may be found from

$$a_R = qa_r + (1 - q)a_c \tag{3.25}$$

where $0 \leq q \leq 1$ and where a_R represents the reconciled estimator. By an elementary theorem on the variance of a linear combination of random variables.

$$\begin{aligned} VAR(a_R) = {} & q^2\, VAR(a_r) + (1 - q)^2\, VAR(a_c) \\ & + 2(1 - q)\, q\, COVAR(a_r a_c). \end{aligned} \tag{3.26}$$

Therefore, to find the value of q that minimizes $VAR(a_R)$, set the derivative of (3.26) with respect to q equal to zero and solve for q^*.

$$\begin{aligned} q^* = {} & [VAR(a_c) - COVAR(a_r a_c)]/ \\ & [VAR(a_r) + VAR(a_c) - 2\, COVAR(a_r a_c)] \end{aligned} \tag{3.27}$$

At least two features of equation (3.27) deserve further comment. First, there is an important distinction between choosing q^* for off-diagonal versus diagonal elements in the matrix of regional coefficients. In the case of diagonal elements, it is evident that equations (3.8) and (3.19) would be estimated from observations on the same firms. Hence, it would be reasonable to expect that $COVAR(a_r a_c)$ does not equal 0. On the other hand, for off-diagonal elements, $COVAR(a_r a_c)$ is likely to equal 0, since in this case (3.8) and (3.19) are estimated from different sets of sample information. More specifically, $COVAR(a_r a_c)$ will equal zero for the off-diagonal regional coefficients if measurement errors are uncorrelated between firms in different sectors.

A second important feature of equation (3.27) concerns the interpretation of q*. For off-diagonal coefficients, q* depends exclusively on the ratio $[VAR(a_c)]/[VAR(a_c) + VAR(a_r)]$. Hence, in constructing a_R, equation (3.25) weights a_r and a_c inversely according to the amount of misinformation each is likely to provide. Therefore, in Jensen and McGaurr's nomenclature, q* might be interpreted as an objectively determined reliability weight.

4. An Example of Reconciliation

In this section, the reconciliation of a set of "rows only" and "columns only" regional coefficients will be illustrated by using an instrumental variable method. The 2SLS estimation technique will be applied to survey data obtained by Miernyk [1970] from 29 input-output sectors of the West Virginia economy, an especially interesting exercise since Miernyk's study of West Virginia is one of the best regional input-output investigations conducted to date. The discussion will be confined to a presentation of empirical results. A description of the data and the collection methods used is available elsewhere [Gerking 1976b, Miernyk 1970].

To illustrate the reconciliation procedure using 2SLS, estimates of the regional coefficients must be obtained for all 29 sectors, according to the methods described in Section 3. As a practical matter, this amounted to: (1) estimating the α_{ij} and the β_{ij} together with the variances for both sets of coefficients; (2) converting estimates of the β_{ij} to estimates of α_{ij}, according to a_{ij} (2SLS) equals $c_{ij} b_{ij}$ (2SLS); (3) testing the regression residuals for heteroscedasticity, using the Goldfeld-Quandt test; (4) adjusting the data as needed to correct this problem; and (5) revising both parameter and coefficient variance estimates in those cases where heteroscedasticity was found to be present. This exercise produced a large quantity of estimates, and no attempt will be made to report them all. However, a subset of these results is presented in Tables 2, 3, and 4.

The "rows only" estimates, the "columns only" estimates, and the minimum-variance, reconciled estimates are presented for three rows of the West Virginia input-output table.[12] These rows correspond to the following sectors: (1) Logging and Sawmills, (2) Printing and Publishing, and (3) All Other Retail Trade. These sectors were chosen because the assumptions regarding the exact measurement of the variables $SH_j(k)$ and $SG(k)$ appeared to be satisfied the best. In fact, most of the firms in these sectors that indicated they made sales to households provided data on their sales tax collections.

In Tables 2, 3, and 4 the first column gives the row and column index for the regional coefficient under consideration. These indices are taken from Table 1. As a result, the values in, say, the first line of Table 2 pertain to the parameter $\alpha_{14,2}$, which represents the minimum value of the output from firms in the Logging and Sawmills sector required to produce a dollar's worth of output in the Underground Coal Mining sector. The second and third columns provide the "rows only" and "columns only" estimates of the regional

[12] Only those coefficients for which both the "rows only" and "columns only" estimates could be calculated are reported. An expanded set of tables is available from the author on request.

TABLE 1
WEST VIRGINIA INPUT-OUTPUT SECTORS

Sector number	Sector name	Sector number	Sector name
1	Agriculture	25	Transportation equipment
2	Coal mines (underground)	26	Instruments and related products
3	Coal mines (strip and auger)	27	All other manufacturing
4	Petroleum and natural gas	28	Eating and drinking establishments
5	All other mining	29	Wholesale trade
6	Building construction	30	Retail food stores
7	Nonbuilding construction	31	Retail gasoline stations
8	Special construction	32	All other retail
9	Food products (n.e.c.)	33	Banking
10	Food products (dairies)	34	Other finance
11	Food products (bakeries)	35	Insurance agents and brokers
12	Food products (beverages)	36	Real estate
13	Apparel and accessories	37	All other FIRE
14	Logging and sawmills	38	Hotels and lodgings
15	Furniture and wood fabrication	39	Medical and legal services
		40	Educational services
16	Printing and publishing	41	All other services
17	Chemicals	42	Railroads
18	Petroleum	43	Trucking and warehousing
19	Glass	44	All other transportation
20	Stone and clay products	45	Communications
21	Primary metals	46	Electric companies and systems
22	Fabricated metals		
23	Machinery (except electric)	47	Gas companies and services
24	Electrical machinery	48	Water and sanitary services

Source: Miernyk [1970, 10].

TABLE 2
ESTIMATES OF THE REGIONAL COEFFICIENTS FOR THE ROW CORRESPONDING TO THE LOGGING AND SAWMILLS SECTOR

Row, column index	"Rows only" estimate	"Columns only" estimate	Reconciled estimate	q*	Miernyk's estimate
14, 2	.27064-01 (.10400-01)	.39158-02 (.23174-02)	.50108-02 (.22619-02)	.04730	.00649
14, 14	.45963-01 (.21788-01)	.13306-00 (.43201-01)	.59225-01 (.18442-01)	.74835	.15534
14, 15	.15497-00 (.12319-00)	.91322-01 (.23170-01)	.93497-01 (.22771-01)	.03417	.13416
14, 29	.13327-03 (.52066-04)	.59261-05 (.95089-04)	.10390-03 (.45668-04)	.76934	.00000

TABLE 3

ESTIMATES OF THE REGIONAL COEFFICIENTS FOR THE ROW CORRESPONDING TO THE PRINTING AND PUBLISHING SECTOR

Row, column index	"Rows only" estimate	"Columns only" estimate	Reconciled estimate	q*	Miernyk's estimate
16, 2	.33965-04 (.29910-03)	.60432-03 (.43228-03)	.21862-03 (.24596-03)	.67625	.00078
16, 6	.24366-03 (.61416-03)	.26306-02 (.75518-03)	.11939-02 (.47648-03)	.60190	.00257
16, 7	.14796-03 (.55779-03)	.44503-03 (.83414-04)	.43853-03 (.82497-04)	.02187	.00101
16, 8	.10526-04 (.43471-03)	.21344-02 (.17311-02)	.13651-03 (.42162-03)	.94068	.00051
16, 9	.35775-03 (.10471-02)	.74164-01 (.41329-02)	.48096-02 (.10150-02)	.93968	.00262
16, 11	.58619-05 (.31276-04)	.25071-03 (.69894-03)	.63512-05 (.31245-04)	.99800	.00147
16, 14	.45996-04 (.18872-02)	.56006-04 (.77628-04)	.55989-04 (.77562-04)	.00169	.00032
16, 16	.70790-04 (.74099-02)	.40595-01 (.16860-01)	.56577-02 (.69832-02)	.86213	.26199
16, 20	.20191-03 (.64192-03)	.69388-03 (.21158-02)	.24338-03 (.61427-03)	.91571	.00258
16, 22	.24822-04 (.10251-02)	.47939-02 (.23180-02)	.80495-03 (.93752-03)	.83642	.00270
16, 24	.25172-04 (.10395-02)	.55853-04 (.24747-03)	.54207-04 (.24074-03)	.05364	.00104
16, 27	.72309-03 (.20137-02)	.1813-02 (.23975-02)	.11732-02 (.15420-02)	.58635	.00250
16, 29	.20864-03 (.46897-03)	.67695-03 (.35120-03)	.50868-03 (.28111-03)	.35931	.00083
16, 31	.73247-04 (.30700-03)	.37938-03 (.33307-03)	.21387-03 (.22574-03)	.54066	.00239
16, 32	.19730-02 (.50075-02)	.24075-01 (.12112-02)	.22853-01 (.11773-02)	.65527	.02312
16, 33	.44177-03 (.11271-02)	.17330-01 (.44475-02)	.14609-02 (.10926-02)	.93965	.00547
16, 36	.25645-03 (.10200-02)	.18575-01 (.72254-02)	.61438-03 (.10100-02)	.98046	.02568
16, 38	.29675-03 (.11371-02)	.27385-03 (.40827-03)	.27647-03 (.38425-03)	.11419	.00147
16, 39	.19033-03 (.45639-03)	.34771-02 (.21848-02)	.32776-03 (.44675-03)	.95819	.00065
16, 41	.13204-03 (.45676-03)	.42843-02 (.11490-02)	.70385-03 (.42445-03)	.86354	.00489
16, 43	.11648-03 (.41448-03)	.90314-02 (.19789-02)	.49113-03 (.40568-03)	.95797	.00472
16, 47	.44989-05 (.17541-03)	.70067-03 (.54526-05)	.70000-03 (.54500-05)	.00097	.00051

coefficients indicated in the first column, according to the 2SLS methods summarized by equations (3.8) and (3.19). Standard errors are given in parentheses beneath each estimate. The fourth column, then, presents the minimum-variance, reconciled estimates. These reconciled estimates were constructed by using the values of q* in column five, in conjunction with equation (3.27). The values of q*, in turn, were calculated by assuming that COVAR $(a_r a_c)$ equals 0 for off-diagonal regional coefficients, and allowing COVAR $(a_r a_c)$ does not equal 0 for the diagonal coefficients. For comparison purposes, Miernyk's estimates of the regional coefficients are given in the last column.

TABLE 4

ESTIMATES OF THE REGIONAL COEFFICIENTS FOR THE ROW CORRESPONDING TO THE ALL OTHER RETAIL TRADE SECTOR

Row, column index	"Rows only" estimate	"Columns only" estimate	Reconciled estimate	q*	Miernyk's estimate
32, 6	.27134-02 (.76250-02)	.17581-02 (.14458-02)	.17913-02 (.14205-02)	.03471	.00044
32, 7	.42040-04 (.88885-03)	.24967-03 (.77765-03)	.15965-03 (.45723-04)	.43357	.00184
32, 8	.20688-04 (.43420-03)	.78073-02 (.76453-02)	.45723-04 (.43350-03)	.99678	.00049
32, 16	.24096-04 (.67041-03)	.54423-02 (.23333-02)	.43728-03 (.64434-03)	.92374	.00422
32, 20	.22386-05 (.62260-04)	.88932-03 (.13733-02)	.40581-05 (.62196-04)	.99795	.00475
32, 23	.65640-05 (.18257-03)	.36500-03 (.23230-03)	.14343-03 (.14354-03)	.61817	.00100
32, 27	.18133-04 (.10316-03)	.12121-03 (.33452-03)	.27084-04 (.98579-04)	.91216	.00019
32, 31	.11435-05 (.31844-04)	.95533-03 (.72177-03)	.29972-05 (.31813-04)	.99806	.00106
32, 32	.38039-03 (.10136-02)	.20158-02 (.51091-03)	.16992-02 (.47030-03)	.19358	.00291
32, 33	.15044-03 (.14604-02)	.27677-02 (.17841-02)	.12005-02 (.11302-02)	.59879	.00286
32, 36	.20145-03 (.19276-02)	.76641-03 (.33920-03)	.14944-03 (.33407-03)	.03004	.00660
32, 38	.89713-04 (.56308-03)	.52272-02 (.75869-02)	.11786-03 (.56154-03)	.99452	.00684
32, 39	.84937-04 (.82411-03)	.28553-03 (.10220-02)	.16398-03 (.64152-03)	.60598	.00630
32, 41	.67412-05 (.65428-04)	.43989-01 (.24726-01)	.70492-05 (.65428-04)	.9999	.01079
32, 45	.75786-04 (.73425-03)	.58152-02 (.51753-02)	.18903-03 (.72697-03)	.89027	.00161
32, 47	.15959-04 (.15378-03)	.19821-02 (.15056-04)	.19634-02 (.14984-04)	.00949	.00046

To the extent that they are typical, the results in Tables 2, 3, and 4 indicate that the reconciliation problem in input-output analysis should not be taken lightly. Even a casual examination of columns 2 and 3 of these tables reveals a substantial difference between the "rows only" and "columns only" estimates for many of the regional coefficients. In addition, the standard errors for the two types of estimates often differ markedly, and do not show a pronounced tendency to be lower for one type of estimate than for another. In comparing columns 4 and 6 of each of the three tables, perhaps most importantly, there are often substantial discrepancies between the minimum-variance, reconciled estimates and those constructed by Miernyk. For example, the minimum variance estimates for $\alpha_{16,11}$, $\alpha_{32,20}$, and $\alpha_{32,41}$ are more than a hundred times smaller than Miernyk's estimates. Thus, it is an understatement to say that the final table of regional coefficients may be greatly affected by the way in which the discrepancies between the "rows only" and "columns only" estimates are reconciled.

5. Conclusion

The purpose of this paper is not to condemn Miernyk's estimates of the West Virginia regional coefficients. The reader should not interpret the minimum-variance, reconciled estimates presented in Tables 2, 3, and 4 as "correct," or regard Miernyk's estimates as "incorrect." No evidence has been presented to support such a conclusion. Instead, the minimum-variance estimates are intended only to illustrate an alternative to past approaches concerning the reconciliation problem in input-output analysis.

The minimum-variance procedure does have at least two theoretical advantages over its competitors. First, in constructing a given reconciled coefficient, "rows only" and "columns only" estimates are weighted inversely according to the amount of misinformation each is likely to provide. Second, the standard errors for the minimum-variance, reconciled estimates will always be no larger than the corresponding measure for either the "rows only" or "columns only" counterparts. This last point may be illustrated by comparing standard errors of the coefficients in the second, third, and fourth columns of Tables 2, 3, and 4.

Since this paper has presented no proof that the minimum-variance procedure is superior to other methods, further research is needed on the reconciliation problem. Three research avenues seem promising.

First, and most obvious, would be experimentation with alternate econometric techniques for estimating the "rows only" and "columns only" coefficients. The three instrumental-variable methods discussed here—the Wald-Bartlett, Durbin, and 2SLS—were presented mainly because of the ease with which they can be applied in a practical setting. However, it is not difficult to imagine that other estimators, and perhaps equally simple ones, may prove to be superior on certain statistical grounds.

Second, a minimum-variance approach to reconciliation should be applied to other, and possibly more comprehensive, sets of input-output data. This exercise would be of interest for two related reasons. Further comparisons of the "rows only" and "columns only" coefficient estimates would be useful. Owing to budget restrictions, many input-output analysts may implement their studies by collecting either sales or purchases data, rather than both. Therefore, such a comparison might help these investigators decide which type of data to collect. Also, estimates of the variance of the reconciled estimates could be compared with the corresponding measure for, say, the "rows only" ("columns only") estimates in order to determine the incremental value of collecting purchase (sales) data on interindustry transactions.

A third, and final, research suggestion concerns the use of *a priori* information in the reconciliation process. The reconciliation strategy described in this paper uses no information of this type. However, input-output analysts may have a great deal of pertinent information about the final table of regional coefficients that is not fully captured by the data they have collected. This information, if correct, could be used to further improve the efficiency of the reconciled estimates. As a consequence, it may be valuable to develop constrained estimation procedures by which to take *a priori* information into account.

References

Bacharach, M. 1970. *Biproportional Matrices and Input-Output Change*. Cambridge University Press.

Bartlett, M.S. 1949. Fitting a straight line when both variables are subject to error. *Biometrics 5*.

Bourque, P.J. *et al.* 1967. *The Washington Economy: An Input-Output Study*, University of Washington Press.

Chenery, H.B., and P.G. Clark. 1959. *Interindustry Economics*. Wiley.

Dorfman, R., P.A. Samuelson, and R.M. Solow. 1958. *Linear Programming and Economic Analysis*. McGraw-Hill.

Durbin, J. 1954. Errors in variables. *Review of the International Statistics Institute: 23*.

Gerking, S.D.

1976a. Input-output as a simple econometric model. *The Review of Economics and Statistics 58*.

1976b. *Estimation of Stochastic Input-Output Models: Some Statistical Problems*. Martinus Nijhoff Social Sciences Division.

Goldfeld, S.M., and R.E. Quandt. 1965. Some tests for homoscedasticity. *Journal of the American Statistical Association 40*.

Hansen, W.L., and C.M. Tiebout. 1963. An intersectoral flows analysis of the California economy. *The Review of Economics and Statistics 45*.

Isard, W., and T. Langford. 1971. *Regional Input-Output Study*. MIT Press.

Jensen, R.C., and D. McGaurr. 1976. Reconciliation of purchases and sales estimates in an input-output table. *Urban Studies 13*.

Miernyk, W.H. *et al.* 1970. *Simulating Regional Economic Development*. D.C. Heath.

Richardson, H. 1972. *Input-Output and Regional Economics*. Wiley.

Tilanus, C.B. 1966. *Input-Output Experiments: The Netherlands, 1948-1961*. Rotterdam University Press.

Wald, A. 1940. The fitting of straight lines if both variables are subject to error. *Annals of Mathematical Statistics 11*.