REGIONAL ECONOMIC ISSUES

Working Paper Series

Census Content of Bureau of Economic Analysis Input-Output Data

Philip R. Israilevich, Randall W. Jackson and Jon Comer

FEDERAL RESERVE BANK
OF CHICAGO

WP - 1991 / 2

Census Content of Bureau of Economic Analysis Input-Output Data

Introduction

Perspectives on the nature of input-output (I-O) frameworks can be characterized by two schools of thought. One school views the input-output data as a representation of empirical reality. The values are believed to represent observations drawn from economic reality. The second school views the input-output model essentially as a concept. An input-output table is seen by this group as a qualitative representation of regional economic structure. Qualitatively accurate tables can be based in large part on expert opinion. Conversely, quantitatively accurate tables will be based strongly on empirical observation.

The ex-ante input-output framework, originally developed by Battelle Institute, is an example of the qualitative representation of economic structure.[1] These tables, which are based on expert opinion instead of observed data, can be constructed relatively quickly and inexpensively. The reliability of these tables has not been established. Ex ante IO tables have not been tested, for example, against data from the U.S. Censuses.

The input-output accounts compiled and published by the U.S. Census, Bureau of Economic Analysis (BEA), are treated widely as tables based on observation. However, close examination of the BEA manufacturing data reveals less than total consistency with the Census data on which they are based. Hence, some degree of "expert opinion", either subjective or mechanical, enters into the BEA process. Therefore, there may be no examples of purely objective quantitative IO tables. There is, more likely, a continuum of positions between purely quantitative tables based solely on observation, and tables based solely on expert opinion.

Philip R. Israilevich, Federal Reserve Bank, Chicago and University of Illinois. Randall W. Jackson and Jon Comer, Ohio State University. The support of National Science Foundation Grant SES-8822044, and of the Federal Reserve Bank of Chicago are gratefully acknowledged.

In this paper, we compare the published BEA data with published Census data to determine the degree of the bias introduced by "expert opinion" in the BEA IO table construction process.[2] If this bias is strong, then the distinction between the ex ante type and BEA type tables is overemphasized, i.e. in the process of IO tables construction both BEA and ex ante approaches rely heavily on the expert opinions. Conversely, a strong similarity between primary Census data and BEA IO data would suggest that BEA type tables primarily relies on the observed data. This would make BEA IO approach distinct from the ex ante approach.

More specifically, this paper presents the results of a preliminary analysis of the relationship between Census of Manufacturing data and BEA data. The tests reported here are part of a larger project in which the spatial variability of input-output relationships is the primary focus. Broader long-run interests include

1. providing a stronger empirical basis for modifying national input-output tables to reflect the structure of regional economies (referred to as adaptation techniques)

2. assessing variability in production within industries due to establishment size, stage of production process, capital and labor intensity, etc., and

3. providing an establishment scale basis for the development of probabilistic input-output models.

These issues involve or influence the relationships among national and regional input-output data and the production relationships of establishments. Throughout the larger project, we focus on the empirical relationships between disaggregated Bureau of Census data and aggregated Bureau of Economic Analysis data. An appropriate starting point is therefore an analysis of the relationship between aggregate national input-output data and the aggregate primary data that are the foundation of the national input-output accounts.

The following section begins by describing our understanding of the input-output tables construction process. This is a necessary basis from which to elaborate our expectations. The third section describes the data preparation and testing procedure, and the results of each step of the test. The final section summarizes our findings and raises important issues for further attention.

Accounts Construction

The procedure

Constructing and publishing the national input-output accounts is one among many responsibilities of the Bureau of Economic Analysis (BEA). The quinquennial accounts represent the interindustry and inter-sectoral dollar flows for years in which the Bureau of Census surveys manufacturers, services, trade, etc. (i.e., 1967, 1972, 1977, 1982, etc.). Since 1972, the input-output accounts have been published on a commodity-industry basis.

The BEA table is based on data from the Manufacturing, Construction, Wholesale and Retail Censuses, on data from regulatory agencies related to financial institutions and utilities, and numerous other sources. To illustrate some of the problems faced by the BEA, consider the example of the Manufacturing Census. This Census reports expenditure shares on materials (6 - digit) consumed by each industry. Some materials are not specified by the questionnaire and fall into the "other" category. The BEA, using expert opinion, has to reapportion this "other" category into specific products and services. The Manufacturing Census does not collect any data on the consumption of services or construction, or trade, or transportation expenditures. All of these have to be estimated based on nonmanufacturing Census reports and expert opinion. The BEA has the additional chores of splitting expenditure shares into domestic and foreign, and reconciling industry sales with industry purchases and with the national income and product accounts.

As most input-output analysts are acutely aware, the process is extremely time-consuming. The accounts for a given year may be published seven or eight years later (e.g., the 1982 input-output structure is expected to be publicly available in mid-1991). From an outside user's perspective, there appear to be three main reasons for this delay. First is the sheer magnitude of the task, which will be described more fully, below. Second, like most federal agencies in the last decade, the funding for the BEA has not kept pace with the increasing demands for data provision. The result of budget tightness has led to reduction of full-time employees (primarily by attrition) devoted to the task, during a period in which the changes in the underlying structure of the U.S. economy have been complex and rapid. The third reason is due to the delay in the transfer of data from the Bureau of the Census (BOC) to the BEA. Based on a synthesis of various documents and conversations with staff of the two Bureaus, and focussing on manufacturing, our understanding of the process is as follows.

During the Census year, the BOC gathers its survey data via questionnaire. Great effort is taken to ensure that coverage is as complete as possible. Follow-up questionnaires and contacts by phone or in person are frequently undertaken. The data gathering process itself is time consuming. As the data come in, they are encoded and tested for consistency, especially with respect to reporting and coding errors. Once the data are encoded and checked, they are evaluated with respect to disclosure rules. The data are then compiled in various forms, such as that of Table 7 in the Census of Manufacturing, which lists selected materials consumed by industry. This entire process can take as long as four years.

During this time, the BEA will have been gathering primary data in other formats, including canvassing trade association experts, industry leaders, and key persons in targeted industries. At some point, perhaps four years after the quinquennial census, the BOC makes its industry summaries available to the BEA on a computer tape.

From this point on, the process becomes less clear. The input-output research community generally has operated on the assumption that BEA uses BOC data, in the form of Table 7 of Census of Manufacturing, for example, in conjunction with its own primary data, to develop a prototype input and output structure of the economy. Discrepancies among the various data sources would be arbitraged and reconciled, and the entire accounting structure would be reconciled with the national income and product accounts (NIPA's) and gross national product (GNP) for the census year.

The whole procedure is much more intricate than here suggested. To trace every procedure executed by the BEA is beyond the scope of this study. However, the activities listed, form a skeletal view of the input-output account construction process.

Expectations

Intuition suggests that inertia is characteristic of the accounts from period to period. Sources of this inertia are varied. One source is the technological inertia in the production system itself. The error and consistency checking procedure may be a second source of inertia in the accounts. Values different in the extreme from previous period values logically demand further scrutiny.

The result might well be a temporal smoothing of the data. This intentional or unintentional smoothing may, in fact, take place within both involved agencies.

We expect a significant relationship between input-output coefficients based on BOC estimates of materials consumed by industry and the final coefficients based on the BEA Use table, which lists commodities consumed by industry. We logically expect the direction, and to a significant extent the magnitude of changes, in these coefficients to correspond. Trends in the BOC data should be reflected by trends in the BEA data.

The Analysis

Generating comparable data for BEA and BOC industries is a task made more difficult by the differences in agency practices. The BEA uses its own sectoral classification, while the BOC uses the Standard Industrial Classification (SIC) code system. To conduct the tests described below, we limited the analysis to those sectors whose definitions were virtually identical. Sectoral definitions vary temporally, which further restricted the number of usable time periods. Also, because we had access only to the printed versions of these data sources, it was necessary to limit the number of sectors analyzed.

The industrial sectors for the analysis were selected by assessing the *sensitivity* of industries in the economy. The technique identifies sensitive sectors based on West's (1982) measure of *inverse* sensitivity. An input coefficient is defined to be inverse sensitive if a given change in value leads to a greater than average change in the Leontief inverse. Quantitative estimates of these sensitivities were combined on a column by column basis to determine the relative sensitivity rankings of industries. For 1977, these are BEA sectors:

14 Food and Kindred Products

17 Miscellaneous Textile Goods and Floor Coverings

59 Motor Vehicles and Equipment

These three sectors are comprised of 59 BEA sub-sectors. From these, we compiled 109 observations on materials used, for consistently defined 1972 and 1977 BEA and BOC data. The variables used in the following analysis are

B^{72}, B^{77}, denoting BEA IO tables for 1972 and 1977, and C^{72} and C^{77} denoting BOC data for 1972 and 1977.

To assess the variability in each data source, we initially regressed 1977 values on 1972 counterparts. A constant term was not included, because a zero base-level is expected.[3] We were interested in variability within each data set, and in the consistency between data sources within declining and growing input-coefficients. Therefore, we also partitioned each agencies data base. One partition corresponds to materials for which the BEA input coefficients were increasing in magnitude, and the other subset is comprised of materials for which the BEA input coefficients were decreasing. Because we expect inertia to be dominant, we expect the coefficient on prior year coefficient to be very near unity. Hence, the t-statistics and 2-tail probabilities for the prior year variable have been adjusted to indicate statistical difference of the regression coefficient from unity, rather than from zero. The results of these regressions are shown in Table 1 for BEA and Table 2 for BOC.

By regressing B^{77} on B^{72} for all observations, the coefficient for B72 is not significantly different from one. The R-squared indicates that nearly 95% of the variation in B^{77} can be explained by the variation in B^{72}. If we knew beforehand which BEA coefficients would be increasing over the period, then for these coefficients the results shown in Table 1b indicate an average increase of roughly 10% in the input coefficient value over the period, with a regression coefficient that is significantly different from one. Likewise, Table 1c suggests that for input coefficients whose values were decreasing over the period, the new value, on average, would be roughly 85% of the old, with nearly 96% of the variation in the new values explained. The results for the BOC input coefficients are very similar in form. On average, however, the observation pairs from the Census either increased more than, decreased less than, or, in the extreme, changed in the opposite direction from their BEA counterparts.

An important difference between the two data sources is that the BOC reports values in purchaser's prices, while the BEA reports values in producers prices. Ideally, we expect the foundation of the relationship to be that given in equation 1.

$$\frac{B^{77}}{B^{72}} = \frac{m^{77}C^{77}}{m^{72}C^{72}} \tag{1}$$

where B^t = BEA reported commodity use for time t divided by BEA's estimate for total output for the using industry for time t

C^t = Census reported material purchased for time t divided by the Census estimate of total value of shipments for time t

and m^t = a downward adjustment factor to purge material purchased estimates of trade and transport margins.

To simplify the expression, rewrite the equation as

$$B = mC \tag{2}$$

With the corresponding data, we can test to see how closely the equality holds in practice. By using the logarithmic form:

$$ln(B) = ln(m) + \alpha_1 \, ln(C) \tag{3}$$

we can test for a change in margin factors ln(m) as a constant, as well as for $\alpha_1 = 1$. Throughout the remainder of the paper, the following variable names are used:

α_0	- constant term in the regression equation
LBRATIO	- ln (B^{77} / B^{72}) as a proxy for ln (B)
LCRATIO	- ln (C^{77} / C^{72}) as a proxy for ln (C)

The relationship shown in equation 3 was tested using OLS regression methods. Again, the relationship was assessed for the whole set of 115 observations, and for partitions defined by increasing or decreasing BEA input-coefficients. The results are presented in Table 3. The t-statistics refer to difference from zero for the constant term and difference from unity for the coefficient on the treatment variable.

Table 3a shows the regression results for the entire sample. The constant is not statistically different from zero, so on average, the ratio of margins is not different from one. The adjusted R^2 of .15 shows that only 15% of the variation in the logged BEA ratios can be accounted for by corresponding variation in the logged Census coefficient data. Influential variables must be missing from the model specification. BEA's decisions and other data manipulations are responsible for the majority of the variation in published input-output data. The coefficient on the independent variable is the elasticity of BEA coefficients to Census coefficients. On average, a 1% change in Census coefficients changes the corresponding BEA coefficient by .57%. However, this elasticity is not sufficiently informative, because it does not separate coefficients increasing in time from those that are decreasing.

Had we had prior information on which BEA coefficients would increase and which would decrease, Table 3b and Table 3c would provide the relevant information. Partitioning the data set reveals the significance of the change in margins in determining the change in BEA data. However, in both cases the magnitude of the coefficient on the LCRATIO falls, which indicates an even weaker relationship between the BEA data and the BOC data upon which they are thought to be based.

Discussion

The results of this limited analysis might suggest cause for great alarm on first reading. However, there are several points that should be noted, before any call to arms. First among these is small sample size. The BEA's use table contains 534 x 537 transactions (286,758) values. Many of these are zero values. Our analysis used only 109 observations from among the larger set. This criticism is partially balanced by recalling our focus on coefficients from industrial sectors with greater than average analytical significance. These sectors are in one sense, key economic sectors. Hence, results for these sectors alone are of interest.

There also are different views concerning the appropriate foundations of input-output tables. There are those who believe that input-output tables must be quantitatively precise. Many economists, for example, treat BEA IO coefficients as unbiased, observed data, and attempt to "explain" these coefficients with input prices or other variables (see Hudson and Jorgenson, 1975, for a well known example). There are others, however, who are convinced that for the input-output framework to be useful, the structure must only be qualitatively accurate. For them, input-output tables are, conceptually, qualitative representations.

Based on the limited findings presented, herein, those who treat the BEA input-output data as grounded on empirical observation might be tempted to question the degree of confidence that can be placed on BEA's published input-output data. However, if BEA IO accurately reflect the qualitative character of regional (national) economic structure, such input-output tables should play a valuable role in economic analysis.

Lacking, however, is a more complete communication of the process by which the BEA makes its final estimates of inter-industry and inter-sectoral transactions. Input-output data increasingly provide the bases for estimates of regional economic impacts of public capital expenditures, and for forecasts of regional economic activity. The data are used in generating regional input-output models, and in integrated input-output econometric forecasting models. The accounts at the national level are used by industry in assessing markets and in forecasting demand. Researchers and analysts at all levels need to know that the data on which they base their work is reliable. To this end, the input-output table construction process must be more publicly available. This effort may simply involve the disclosure of the extent to which there is heavy reliance on expert opinion. Finally, if expert opinion is playing a significant role in the BEA's table construction process, researchers may do well to re-assess the merits of ex ante approaches to IO table generation.

Footnotes

[1] Ex ante IO was developed by the Battelle Memorial Institute and is based on the data developed from experts testifying on the use of energy, materials and services by different industries, see H.W. Fisher (1975).

[2] Bias, in the content of IO table construction, should not be interpreted negatively. Problems of primary data collection clearly may justify modification based on expert opinion.

[3] There are no data pairs for which one element is zero. However, the use of a constant would reflect an expected base-level use of an input in 1977 when that input was not used in 1972. The number of new non-zero elements in a table is a very low percentage of all prior-year zero elements. Separate regressions including a constant term produced a coefficient not statistically different from zero.

References

Fisher, H.W. (1975). *Ex Ante* as a Supplement or Alternative to RAS in Updating Input-Output Coefficients, in *Estimating and Projecting Input-Output Coefficients*, edited by R. Allen and W. Gossling, Input-Output Publishing Co., 1975.

Hudson, E., and D.W. Jorgenson (1974). US Energy Policy and Economic Growth, 1975-2000, *Bell Journal of Economics and Management Science*, Volume 5, pp 461-514.

U.S. Department of the Census, *Census of Manufacturing*, 1977.

U.S. Department of Commerce, Bureau of Economic Analysis, *Detailed Input-Output Structure of the U.S. Economy*, 1977.

West, Guy R. "Approximating the Moments and Distributions of Input-Output Multipliers," Working Paper No. 6, University of Queensland, Queensland, Australia, 1982.

Table 1
Regression Results for Bureau of Economic Analysis*

a) Dependent Variable is B^{77}
N = 109 Sample includes all coefficients

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
B^{72}	0.9877550	0.0195678	.626	>0.35
R-squared	0.947483	Mean of dependent var		0.067165
Adjusted R-squared	0.947483	S.D. of dependent var		0.124958

b) Dependent Variable is B^{77}
N = 53 Increasing BEA input coefficient partition

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
B^{72}	1.0974697	0.0219744	4.436	0.000
R-squared	0.971970	Mean of dependent var		0.087723
Adjusted R-squared	0.971970	S.D. of dependent var		0.145092

c) Dependent Variable is B^{77}
N = 56 Decreasing BEA input coefficient partition

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
B^{72}	0.8419614	0.0216109	7.313	0.000
R-squared	0.956900	Mean of dependent var		0.047708
Adjusted R-squared	0.956900	S.D. of dependent var		0.099825

*All T-statistics reported are absolute values.

Table 2
Regression Results for Bureau of Census Data

a) Dependent Variable is C^{77}
N = 109 Sample includes all observations

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C^{72}	0.9892841	0.0130227	.823	>0.3
R-squared	0.975604	Mean of dependent var		0.074076
Adjusted R-squared	0.975604	S.D. of dependent var		0.129948

b) Dependent Variable is C^{77}
N = 56 Decreasing BEA input coefficient partition

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C^{72}	0.9267161	0.0172175	4.256	0.000
R-squared	0.976653	Mean of dependent var		0.050503
Adjusted R-squared	0.976653	S.D. of dependent var		0.101289

c) Dependent Variable is C^{77}
N = 53 Increasing BEA input coefficient partition

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C^{72}	1.0204701	0.0177561	1.153	0.26
R-squared	0.977768	Mean of dependent var		0.098984
Adjusted R-squared	0.977768	S.D. of dependent var		0.151623

Table 3
Inter-agency relationships

a) Dependent Variable is LBRATIO
N = 109 Sample includes all observations

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
α_0	-0.0526087	0.0642354	0.8189987	0.415
LCRATIO	0.6480273	0.1419332	2.479	0.001
R-squared	0.163054	Mean of dependent var		-0.117065
Adjusted R-squared	0.155232	S.D. of dependent var		0.711817

b) Dependent Variable is LBRATIO
N = 56 Decreasing BEA coefficient partition

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
α_0	-0.4529642	0.1276762	3.5477574	0.001
LCRATIO	0.1820999	0.2345767	3.487	0.001
R-squared	0.011037	Mean of dependent var		-0.509457
Adjusted R-squared	-0.007278	S.D. of dependent var		0.782208

c) Dependent Variable is LBRATIO
N = 53 Increasing BEA coefficient partition

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
α_0	0.2424457	0.0318115	7.6213158	0.000
LCRATIO	0.4470740	0.0966971	5.718	0.000
R-squared	0.295349	Mean of dependent var		0.297537
Adjusted R-squared	0.281532	S.D. of dependent var		0.253333