A Method for Constructing Commodity by Industry Flow Matrices

By

Randall W. Jackson, Walter R. Schwarm, Yasuide Okuyama, and Samia Islam

RESEARCH PAPER 2004-5

Regional Research Institute, West Virginia University
Morgantown, WV 26506-6825, USA
Email: Randall.Jackson@mail.wvu.edu

Paper presented at the:
2004 Southern Regional Science Association Conference, New Orleans, LA

Abstract: This paper describes the method used to construct an interregional Commodity by Industry Flow matrix for the United States. The interregional flow matrix method involves the construction of single-state (and DC) SAMs using data from IMPLAN. Once complete, the interregional flows connecting states are estimated using a method based on the Commodity Flow Survey data published by the Bureau of Transportations Statistics. The estimated interregional SAM is then adjusted to insure the integrity of intraregional and system-wide accounts. The procedures have been designed with the goal of ease of replicability, so that updates and extensions of the database can be generated efficiently and at much lower cost as new data are released. The resulting US interregional framework describes flows within and among the 50 states and the District of Colombia, and will provide a valuable database for a broad range of analysis on regions, interregional relationships and policy research.

Introduction

As an economy develops and evolves, the economic interactions among industries, governments, and households become more closely tied and complex. Trends in deregulation and structural change in market-based economic systems within an interregional domestic context have received far less attention than those that effect national and international relations. Recent studies have found in both the US and Japan that interregional trade within a country is growing more rapidly than intra-regional and international trade, and that regions have become tied very closely together (for example, Hewings *et al*., 1998, and Hitomi *et al*., 2000). In fact, according to the Commodity

Flow Survey (CFS) by the Bureau of Transportation Statistics, US industries shipped approximately $7 trillion worth of goods in 1997 using the nation's highways, railroads, waterways, pipelines, and aviation systems (11 billion tons and 2.7 trillion ton-miles). This volume has increased 18.8 percent (up 14.5 percent and 9.9 percent for tons and ton-miles, respectively) since 1993. Not only has the volume of interregional trade increased, but the trading patterns also have become more complex.

It is also critically important for regions/states to understand the nature of their economic interdependencies and to analyze the public policy implications arising there from. In particular, investigating these economic relationships in further detail, identifying, for example, which industries in one state have the strongest and the closest relationships with a given industry in another, can provide a better understanding of how policy changes in one region (state) create impacts other regions (states).

As social accounting matrices and methods find increasing interest and use, a wide body of literature has developed around them. Most SAMs are constructed for nations or individual regions, and although work on interregional SAMs has been evolving over the last decade, very few attempts to generate these models for US regions have been reported in the literature. This paper describes the method used to construct an interregional Commodity by Industry Flow matrix for the United States. It presents an export distribution estimation method, and describes the steps necessary to generate the interregional trade flow portions of the ISAM, and to insure the consistency of both the individual SAM accounts and the system as a whole. After problem and data definition we examine generating single-region social accounting matrices, the estimation of interregional trade characteristics by commodity, how we apportion aggregate interregional commodity flow estimates, and finally adjustments to foreign trade to insure the integrity of the intra-regional and system wide accounts.

Organization and Data:

The interregional Social Accounting Matrix (SAM) framework is employed in this project to analyze the interactions among economic agents (industries, governments, households, etc.) within and across states. The SAM framework describing the full circular flow of income, establishing separate accounts for production, consumption, and transaction with other regions, was originally pioneered by Stone (1961), and applied at the regional and interregional level by Pyatt and Round (1983), Round (1985), and Bell *et al.* (1982).

Traditionally, the interest in SAM has stemmed from their application in a developing country context (Pyatte and Thorbecke, 1976; Pyatt and Round, 1985, for example) where income distribution and poverty issues are of paramount concern. More recent empirical research in developed countries using this framework includes the construction of social accounts for community level (Cole, 1993), the construction of SAM for European nations (Round, 1995), the analysis of interstate capital flows in the US (Kilkenny and Rose, 1995), and the analysis of US rural-urban interdependencies (Kilkenny, 1995).

The empirical application most similar to ours for US states was done during the 70's, initiated at the Harvard Economic Research Project and developed by Polenske (1972). In the most detailed form, it was a 51-region multiregional input-output (MRIO) model for 1963 (50 states and Washington, D.C.) with 79 sectors in each region (see Polenske 1980 for a complete description of the model and its construction). The staff at the Jack Faucett Associates, Inc. assembled the 1977 version of the US MRIO model for 51 regions and 120 industries (Faucett Associates, 1983). These US MRIO models have brought many new research opportunities for the detailed analysis of economic structure and policy analysis and were employed in numerous research projects in subsequent years. Recent applications of these models include Miller and Shao (1990) comparing 1963 and 1977 models to examine the sectoral and spatial aggregations, the US Army Corps of Engineers (1994) creating their own Red River MRIO model based on the 1977 model to evaluate the Red River Water Project, and Horiba (2000) examining interregional trade in comparison to interregional migration in the US using the 1977 model.

Our procedure produces a similar and current database for interindustry activities between regions but also generates more extensive and complete database for the US state economies. Moreover, the interregional SAM model described in this paper includes fully specified interregional relationships, more comparable to Isard's (1951) interregional input-output framework, providing more detailed information regarding economic interactions across regions than the multiregional framework Polenske's model provided.

Single Region SAMs

Constructing IMPLAN single-region SAMs generates data according to the partitions and format listed in Table 1. Data partitions for a single region SAM, with imports treated separately (import ridden as opposed to import laden) are organized as shown in Figure 1. The IMPLAN SAM data are reported in this format to assist GAMS users in constructing single region CGE models from IMPLAN data. Industry sectors were defined in such a way as to correspond closely with the commodity codes used by the US Bureau of Transportation Statistics. The modeled framework encompasses fifty-one regions and 54 industry and commodity sectors, along with four factors of production sectors and 18 institutional sectors.

Table 1. IMPLAN SAM Partitions

2X1	Domestic use of commodities by industries
3X1	Factor incomes
7X1	Industry foreign import use
8X1	Industry domestic import use
1X2	Domestic industry make
4X2	Domestic institutional make
4X3	Factor distributions
5X3	Foreign factor imports
6X3	Domestic factor imports
2X4	Domestic institutional use
4X4	Interinstitutional transfers
7X4	Institutional foreign import use
8X4	Institutional domestic import use
1X7	Industry foreign export make
4X7	Institutional foreign export make
5X5	Foreign transhipments
1X8	Industry domestic export make
4X8	Institutional domestic export make
Each file contains three columns.	
Column 1: Institution Receipts or the row code;	
Column 2: Institution Payments or the column code;	
Column 3: The value in millions of dollars.	

The general structure of the *interregional* SAM is shown in Figure 2, which depicts a 3-region SAM, but which generalizes straightforwardly to our 51-region case. The challenge in constructing the interregional SAM lies in the estimation of values for the shaded and labeled partitions of the off-diagonal blocks in the diagram in Figure 2, and the necessary adjustments to other sectors to ensure a balanced table consistent with the accounting identities of the SAM. This is accomplished using the procedure described in the remainder of this paper.

Figure 1. Single-Region, Import Ridden SAM

	1 Industry	2 Commodity	3 Factors	4 Institutions	5 Foreign Trade	6 Domestic Trade
1-Industry		1x2			1x7	1x8
2-Commodity	2x1			2x4		
3-Factors	3x1					
4-Institutions		4x2	4x3	4x4	4x5	4x8
5-Foreign Trade	7x1		5x3	7x4	5x5	
6-Domestic Trade	8x1		6x3	8x4		

Figure 2. General Structure of the Interregional SAM

		Region 1				Region 2				Region 3				
		Ind	Com	Fac	Inst	Ind	Com	Fac	Inst	Ind	Com	Fac	Inst	ROW
R1	Industry		r011x2				r01021x8				r01031x8			r011x7
	Commodity	r012x1			r012x4	r01028x1			r01028x4	r01038x1			r01038x4	
	Factors	r013x1												
	Institutions		r014x2	r014x3	r014x4		r01024x8				r01034x8			r014x7
R2	Industry		r02011x8				r021x2				r02031x8			r021x7
	Commodity	r02018x1			r02018x4	r022x1			r022x4	r02038x1			r02038x4	
	Factors					r023x1								
	Institutions		r02014x8				r024x2	r024x3	r024x4		r02034x8			r024x7
R3	Industry		r03011x8				r03021x8				r031x2			r031x7
	Commodity	r03018x1			r03018x4	r03028x1			r03028x4	r032x1			r032x4	
	Factors									r033x1				
	Institutions		r03014x8				r03024x8				r034x2	r034x3	r034x4	r034x7
	Foreign Trade	r017x1			r017x4	r012x1			r0217x4	r037x1			r037x4	
	For Fac Imports			r015x3				r025x3				r035x3		
	Dom Fac Imports			r016x3				r026x3				r036x3		
		TIO	TCO	Total Fac. Pmts.	Total Inst. Exp.									

Row and Column Totals

Industry Row - Total Regional Industrial Output (make)
Industry Column - Total Regional Industry Input (use) (Output)
Commodity Row - Total Regional Commodity Supply (Disposition)
Commodity Column - Total Regional Commodity Supply all sources
Factor Row - Total factor receipts (payments to factors) of production
Institutions Row - Total Institutional Receipts (payments to institutions)
Factor Column - Total factor payments to institutions (and trade)
Institutions Column - Total Regional Institutions Expenditures (use)

Export Distributions

The US Bureau of Transportation Statistics collects data through its commodity flow survey (CFS). Although these state-to-state commodity flow estimates are published and available from the BTS, their usefulness is limited for a number of reasons. Foremost among these reasons is that for almost all listed commodities, state-to-state origin-destination tables are dominated by disclosure codes or annotations of one sort or another. The most common of these codes indicates that the estimate is not published due to an unacceptably high statistical variability, and thus, little confidence in the estimate. A second problem for model construction is that the CFS data report *shipment* origin and destination rather than manufacturing origin. Hence, we develop an alternative approach which has the effect of generalizing the distance-volume relationships embedded in the BTS data, smoothing out irregularities observed in the more specific origin-destination commodity-specific shipments data, and enabling application to regions whose boundaries do not coincide with states.

Estimates of intra-regional flows, total domestic imports and total domestic exports were generated in the construction of the single-region SAMs. Because the SAMs are estimated sequentially rather than simultaneously, complete *mutual* consistency is not a certainty (and indeed is unlikely). We could choose to modify all values in the IMPLAN-generated SAMs. Instead, however, we make the simplifying assumption that the intra-regional trade estimates from the IMPLAN-generated single-region SAMS are correct. The task, therefore, is to estimate only the interregional commodity flow distributions and to modify the foreign trade portion of each regional SAM in such a way as to retain or restore internal and external consistency. Individual SAM identities must hold, and because the SAMs exhaust the entire US, the total amount imported by all regions from all other regions also must equal the total amount exported by all regions to all other regions. That is, for all regions combined, domestic imports must equal domestic exports.

We need, therefore, an estimating equation to generate the distribution of known regional domestic exports (given by the single-region SAMs) from each region to each of the other domestic regions in the model. We assume that the *distribution of exports* from one region to all others is fixed, while *export levels* vary with regional production. Hence, our estimating equation need only be a function of transportation costs (as measured by interregional distances) and region-specific commodity demand. To this end, we have applied the following formulation.

For each commodity *i*, let the predicted value of the flow from region *m* to region *n* be computed as

$$\hat{y}_i^{mn} = \frac{\left(w_i^n\right)^{\beta_i} \exp(-\lambda_i d^{mn})}{\sum_n \left(w_i^n\right)^{\beta_i} \exp(-\lambda_i d^{mn})} y_i^{m\bullet} \qquad (2.1)$$

where $\left(w_i^n\right)$ is a weight reflecting region *n*'s demand for imports of commodity *i*,

d^{mn} is the distance separating region *m* from region *n*,

$y_i^{m\bullet} = \sum_{n \neq m} y_i^{mn}$ is total domestic commodity *i* exports from region m.

Where the y_i^{mn}, ideally, are actual shipments derived from observed values published in the 1997 BTS Commodity Flow Survey (CFS). λ_i and β_i are elasticities on distance and commodity demand, respectively. Commodities with larger β values are more sensitive to demand variations, while those with smaller values for λ are more sensitive to shipment distances.

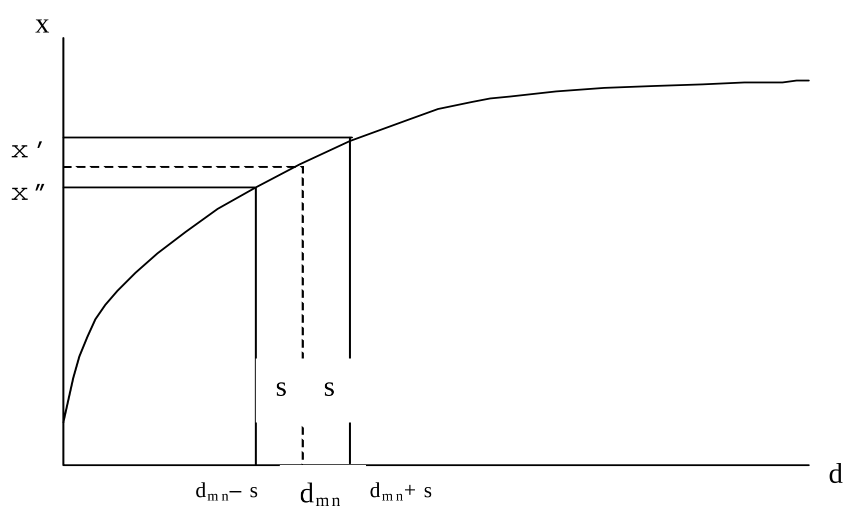
Ideally, to estimate the values of the elasticities for each commodity, λ_i and β_i would be selected to minimize the absolute difference between estimated and observed flows, or $\min Z = \left|\hat{y}_i^{mn} - y_i^{mn}\right|$. Because of the gaps in the BTS CFS data, we do not use observed interregional flows, *per se*. However, we do make use of the BTS commodity-specific summary data to synthetically generate an observed flow estimate. Each 2-digit SCTG[1] code commodity has associated aggregate BTS data on distances shipped in the US[2]. These data report commodity value shipped by distance range (0-50 miles, 50-99 miles, 100-249 miles, etc.). These values were then used to parameterize equation (2.1) via optimization.

---

[1] SCTG – Standard Classification of Transported Goods codes are used by the BTS.
[2] Where the SCTG sectors do not match the model sectors precisely, data for industries with similar output characteristics is used.

Figure 3 serves to further illustrate the method employed. The locus of points corresponds to the cumulative shipment at any distance. The points x' and x" in the diagram correspond to cumulative flows at distance d, with a buffer width of s, seen as $d_{mn} + s$ and $d_{mn} - s$ respectively. Thus calculating the difference between the two flows gives the total flow over the range $[d_{mn} + s, d_{mn} - s]$. Such a distance range is an ideal representation for the purposes of this study, since the shipment from one state to another cannot really be defined by a single distance value. The buffer, s, also provides a convenient mechanism by which we can adjust the regression generated flows to better match the IMPLAN generated regional import totals, which is the goal of the first stage optimization.

Figure 3. Cumulative Distance Function



We use a double log regression specification (natural logs of flows and distance) to parameterize the distance decay function for each commodity.

(2.2) $$\ln(f_i) = \beta_1 + \beta_r \ln(d_r)$$

Where f_i are the cumulative flows for commodity *i*, d_r are the distance range upper limits, and the β_r are commodity-specific distance coefficients. The result of this parameterization is a generalized distance decay function for each commodity, founded on commodity-specific BTS Commodity Flow Survey data.

These functions are then used to generate synthetic "observed" flows corresponding to state centroid interregional distances:

(2.3) $$F_g^{mn} = [\exp(\hat{\beta}_1 + \hat{\beta}_2 \ln(d_{mn} + s)) - \exp(\hat{\beta}_1 + \hat{\beta}_2 \ln(d_{mn} - s))] * X_r$$

where F_g^{mn} is the regression-generated (synthetically observed) commodity flow from region *m* to region *n*, d_{mn} represents interregional distance, s is the size of buffer around interregional "point-to-point" distances, and X_r represents domestic export shares derived from the IMPLAN CGE file corresponding to 1X8 in Figure1. As has been noted previously, this step requires the specification of distance buffers around the interregional "point-to-point" distances. The width of these buffers was determined by minimizing the sum of the absolute differences between the sums of the synthetically observed (regression generated) region-specific imports and the known totals of region-specific domestic demand for imports (from the single-region SAMs), while accounting for each region's share of total system exports of the commodity[3]. Using the following

(2.4) $$\underset{s}{Min} \left| \sum_m F_g^{mn} - \mathrm{IM} \right|$$

where $\sum_m F_g^{mn}$ is the Regression-generated total import demand for region *n*, and *IM* is the corresponding IMPLAN import demand.

With the first step complete λ_i and β_i can be calibrated by minimizing the squared percentage error between logit-predicted and regression-generated flows :

(2.5) $$\underset{\lambda,\beta}{Min} \sum \sum \left(\frac{\hat{Y}_i^{mn} - F_g^{mn}}{F_g^{mn}} \right)^2$$

where $\hat{Y}_i^{mn}$ is the predicted flow of commodity i from region *m* to region *n*, and F_g^{mn} is the regression-generated commodity flow from region *m* to region *n*.

Given commodity-specific values for λ_i and β_i, the aggregate commodity trade flow distributions in the interregional SAM can be derived by applying the generalized

[3] For non-goods commodity sectors (such as higher-level services), averages of the regression parameters from the goods sectors were used. This reflects the assumption that interregional trade in these sectors is related to information flows, which should be reflected by patterns of overall trade. Note that the IMPLAN provides the estimate of total exports of these commodities, while this procedure estimates only the interregional distributions of the exports.

function to domestic export estimates from the single-region SAMs. The procedure described generates considerable variation in interaction parameters across commodities. Depending on the commodity, both population and distance can be very important flow determinants or have virtually no effect on flow determination.

Sector Specific Interregional Commodity Flows

The export distributions for each commodity are first used to apportion the IMPLAN generated domestic export matrices to destination regions. This apportionment is applied equally to commodities exported by institutions and by industries. The export distributions are then unstandardized by IMPLAN export estimates, and normalized by column sum. The result is a set of commodity specific import distributions by region. That is, entries in the new table correspond to the proportion of regional domestic imports that originate in each other region. This new table is then used to apportion aggregate commodities imported by industries and institutions to regions of origin. Because it was derived from the actual export distributions, its use assures consistency between exports from region r to region s and imports by region s from region r (which appear in two separate partitions in the interregional SAM).

Since it is unlikely that an ISAM resulting from this procedure would result in a balanced system, an additional step is implemented *prior* to the import and export apportionment to insure the integrity of both the individual SAMs and of the system as a whole.

The Adjustment Procedure

Adjustment is performed using a RAS type bi-proportional adjustment procedure to generate a balanced matrix. For each commodity, the sum of IMPLAN generated foreign exports by region should equal the corresponding foreign exports from the national SAM. When this is not the case, total regional exports are increased or decreased in equal proportion, with an offsetting adjustment to regional domestic imports. The matrix to be adjusted, in this case, is a 51 x 2 matrix with regions as rows and regional domestic and regional foreign exports as columns. The first column margin is set equal to the original estimate of total regional exports less known total national exports, which is the second column margin. Row margins are set equal to total regional exports by region. The bi-proportional adjustment is then implemented until convergence is obtained. The results of this adjustment procedure insures individual SAM and overall ISAM consistency.

Summary and Discussion

This paper has described an approach to the construction of an interregional SAM for the US, using IMPLAN data as a foundation and incorporating commodity flow data from the US Bureau of Transportation Statistics. The export distribution method provides a generalized function for each commodity, and in so doing, overcomes major obstacles in the use of the CFS data while still taking advantage of the information that is available. The method generates an interregional SAM that is consistent from an accounting perspective, both within each regional SAM and for the interregional modeling system as a whole.

Two areas warrant additional attention within this context. First, generalized export functions for non-commodity (e.g., service) sectors were estimated as a composite function of all flows. Although there is some theoretical justification for this approach, additional research is needed in this area to assess the viability of the embedded assumptions. Second, no attempt has been made in the approach described to estimate interregional factor flows or inter-institutional transfers. Theory and methods underlying the estimation of these flows await further development.

References Cited

Bell, C.L.G., P. Hazell, and R. Slade. (1982) *Project Evaluation in Regional Perspective*. Baltimore, MD; Johns Hopkins Press.

Cole, Sam (1993) "Cultural Accounting in Small Economies," *Regional Studies*, 27: 121-136.

[Jack] Faucett Associates, Inc. (1983) The Multiregional Input-Output Accounts, 1977: Introduction and Summary, Vol.1 (Final Report), prepared for the US Department of Health and Human Services, Washington, D.C. (Jack Faucett Associates Inc., Chevy Chase, MD).

Hewings, Geoffrey J.D., Michael Sonis, Jiemin Guo, Philip R. Israilevich, and Graham R. Schindler. (1998) "The Hollowing-out Process in the Chicago Economy, 1975-2015." *Geographical Analysis* 30 (3): 217-33.

Hitomi, Kazumi., Yasuhide Okuyama, and Geoffrey J.D. Hewings. (2000) "The Role of Interregional Trade in Generating Change in the Interregional Leontief Inverse in Japan, 1980-1990," *Economic Systems Research*, 12 (4): 515-537.

Horiba, Yutaka. (2000) "U.S. Interregional Migration and Trade," paper presented in the Murphy Institute Conference on the Political Economy of Migration, Murphy Institute of Political Economy, Tulane University: March 24-25, 2000.

Isard, Walter. (1951) "Interregional and Regional Input-Output Analysis: A Model of a Space Economy," *Review of Economics and Statistics*, 33 (4): 318-328.

Jackson, R. (2002) "Constructing US Interregional SAMs from IMPLAN Data: Issues and Methods," *Regional Research Institute Working Paper* 2002-13.

Kilkenny, M. (1995) "Operationalising a Rural-Urban General Equilibrium Model Using a Bi-Regional SAM," in G.J.D. Hewings and M. Madden eds. *Social and Demographic Accounting*, Cambridge, UK; Cambridge University Press: pp. 164-179.

Kilkenny, M., and A. Rose (1995) " Interregional SAMs and Capital Accounts," in G.J.D. Hewings and M. Madden eds. *Social and Demographic Accounting*, Cambridge, UK; Cambridge University Press: pp. 41-59.

Miller, R.E., and G. Shao (1990) "Spatial and Sectoral Aggregation in the Commodity-Industry Multiregional Input-Output Model," *Environment and Planning A*, 22: 1637-1656.

Polenske, Karan R. (1972) “The implementation of a Multiregional Input-Output Model for the United States,” in A.P. Carter and A. Brody eds. *Contributions to Input-Output Analysis*, Amsterdam; Elsevier-North Holland: 171-189.

Polenske, Karen R. (1980) *The US Multiregional Input-Output Accounts and Model*. Lexington, MA: Lexington Books.

Pyatt, G., and J.I. Round. (1983) “Improving the Macroeconomic Database: A SAM for Malaysia, 1970. *World Bank Staff Working Paper 646*.

Pyatt, G., and E. Thorbecke. (1976) *Planning Techniques for a Better Future*. Geneva: International Labor Office.

Round, J.I. (1985): Decomposing Multipliers for Economic Systems Involving Regional and World Trade,” *Economic Journal*, 95: 383-399.

Round, J.I. (1995) “A SAM for Europe: Social Accounts at the Regional Level Revisited,” in G.J.D. Hewings and M. Madden eds. *Social and Demographic Accounting*, Cambridge, UK; Cambridge University Press: pp. 15-40.

Stone, Richard. (1961) “Social Accounts at the Regional Level,” in Walter Isard and J.H. Cumberland eds. *Regional Economic Planning*, Paris; Organization for European Economic Cooperation: pp. 263-269.

US Army Corps of Engineers. (1994) *Regional Economic Development: Red River Waterway Project Reevaluation Study In-Progress Review*, May 1994.