Constructing US Interregional SAMs from IMPLAN Data: Issues and Methods

By

Randall W. Jackson

RESEARCH PAPER 2002-14

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

Paper presented at the Fourteenth International Conference on Input-Output Techniques, October 10-15, 2002, Montreal, Canada

Abstract: Many issues arise in the construction of interregional SAMs. In the US, a convenient point of departure is the foundation provided by commodity-by-industry subnational SAMS that can be generated from IMPLAN data. Unfortunately, because the SAMs are generated independently, there is no guarantee of consistency with known national totals. In particular, the sum of IMPLAN generated domestic regional exports across all regions will equal the sum of IMPLAN generated domestic regional imports only by chance. Additionally, while IMPLAN generates total domestic imports and total domestic exports for each region, the analyst must devise an acceptable method for distributing interregional domestic trade. These and other issues that arise in the organization and estimation of the interregional SAM are addressed in this paper.

1. INTRODUCTION

Social accounting matrices and methods are topics of increasing interest and use. A wide body of literature has developed around these topics. Most SAMs are constructed for nations or individual single regions. 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 a new approach to the construction of an interregional SAM for the US. It presents an export distribution estimation method, and describes the steps necessary to 1) generate the interregional trade flow portions of the ISAM, and 2) insuring the consistency of both the individual SAM accounts and the system as a whole.

The procedure for generating the ISAM can be presented in the following steps:

1. Defining the regions for the model
2. Defining the level of industry/commodity sector detail
3. Generating single-region social accounting matrices
4. Estimating interregional trade characteristics by commodity
5. Apportioning aggregate interregional commodity flow estimates
6. Adjusting foreign trade to insure the integrity of the intra-regional and system-wide accounts.

Steps 1 and 2 are self-explanatory and largely independent of the methodology itself. The regions in this paper were chosen for the purposes of constructing a model for a statewide travel demand model for Ohio. Industry sectors were defined in such a way as to correspond closely with the commodity codes used by the US Bureau of Transportation Statistics. There are eleven regions and 56 industry and commodity sectors, along with four sectors for factors of production and 18 institutional sectors. The following section describes the generation of single region SAMs.

2. SINGLE REGION SAMS

The construction of the single-region SAMs generated 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 SAM data were generated in this format to assist GAMS users in constructing single region CGE models from IMPLAN data.

The general structure of the *interregional* SAM is shown in Figure 2, which depicts a 3-region SAM, but which generalizes straightforwardly to our 11-region case. The challenge in constructing the interregional SAM lies in the estimation of values for the gray-shaded 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.

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.	

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				Total	
		Ind	Com	Factors	Inst	Ind	Com	Factors	Inst	Ind	Com	Factors	Inst	Foreign Trade	Domestic Trade
Region 1	Industry		r011x2				>				>			r011x7	interregional replaces r011x8
	Commodity	r012x1			r012x4	!			!	!			!		
	Factors	r013x1													
	Institutions		r014x2	r014x3	r014x4		>				>			r014x7	interregional replaces r014x8
Region 2	Industry		>				r021x2				>			r021x7	interregional replaces r021x8
	Commodity	!			!	r022x1			r022x4	!			!		
	Factors					r023x1									
	Institutions		>				r024x2	r024x3	r024x4		>			r024x7	interregional replaces r024x8
Region 3	Industry		>				>				r031x2			r031x7	interregional replaces r031x8
	Commodity	!			!	!			!	r032x1			r032x4		
	Factors									r033x1					
	Institutions		>				>				r034x2	r034x3	r034x4	r034x7	interregional replaces r034x8
	F Trade	r017x1			r017x4	r012x1			r0217x4	r037x1			r037x4		
	F Trade No Detail			r015x3				r025x3				r035x3			
	Dom Trade No Detail			r016x3				r026x3				r036x3			
	Domestic Trade	interregional use replaces r018x1			interregional use replaces r018x4	use replaces r028x1			interregional use replaces r028x4	interregional use replaces r038x1			interregional use replaces r038x4		
		TIO	TCO	Total Factor Payments	Total Institutional Expenditures										

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)

3. EXPORT DISTRIBUTIONS

The US Bureau of Transportation Statistics publishes data collected through its commodity flow survey (CFS data). 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.[1] 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 were 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* are 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

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

[1] Also, because the Ohio model region includes counties in bordering states, there is no strict comparability between model region flows and state-to-state BTS estimates. This would not be as great a problem for other models were state boundaries strictly adhered to in regional definition.

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*,

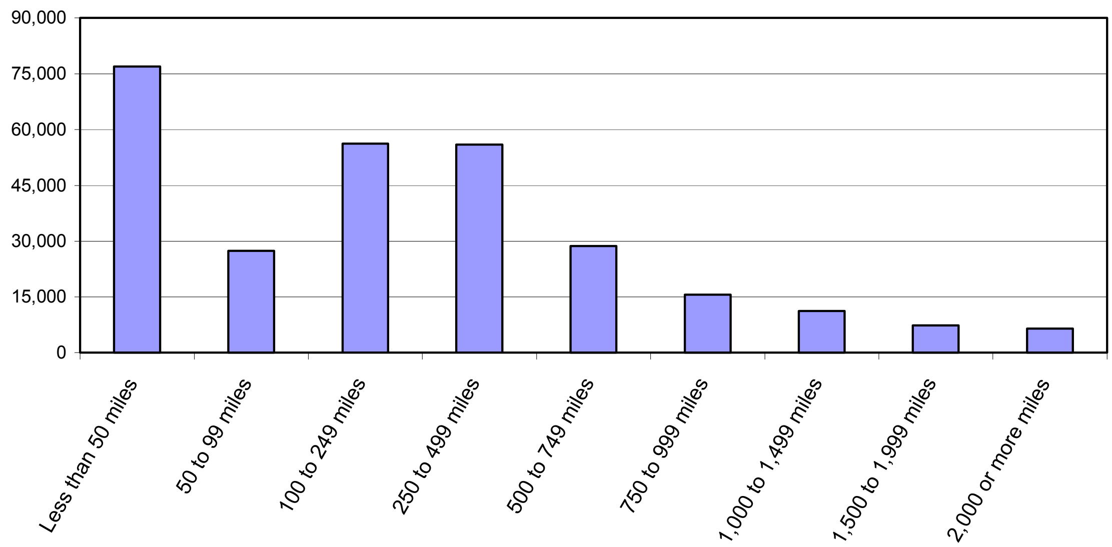
$y_i^{m\bullet} = \sum_{n \neq m} y_i^{mn}$ is total domestic commodity *i* exports from region m, where the y_i^{mn}, ideally, are actual shipments derived from observed values published in the 1997 BTS Commodity Flow Survey (CFS).

and λ_i and β_i are elasticities on distance and population, 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[2] code commodity has associated aggregate BTS data on distances shipped in the US.[3] These data report commodity value shipped by distance range (0-50 miles, 50-99 miles, 100-249 miles, etc.). Figure 3 is typical of the CFS data by 2-digit SCTG.

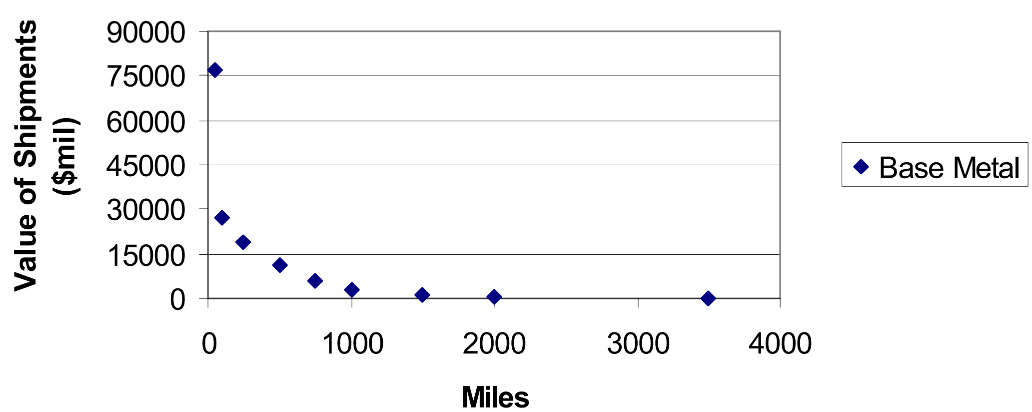
[2] SCTG – Standard Classification of Transported Goods codes are used by the BTS.
[3] Where the SCTG sectors do not match the model sectors precisely, data for industries with similar output characteristics is used.

Figure 3. Value of Shipments by Distance Ranges



Although drawn from CFS data, the graph is somewhat visually deceptive because of the variable distance ranges corresponding to each bar. Figure 4 presents the same data values appropriately distributed.

Figure 4. Normalized Flow Estimates



We use a double log regression specification (natural logs of flows and distance) to parameterize the distance decay function. 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 the ODOT model interregional distances. 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. [4]

These computed interregional commodity flows were then used to parameterize equation (0.1) via optimization. Then, 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 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.

4. 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.

[4] 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.

The Adjustment Procedure

A first inclination might be to implement some form of bi-proportional adjustment procedure to solve the balancing problem. The structure of available data, however, provides us with a more direct method of insuring the integrity and consistency of the interregional SAM and its components. Because domestic exports and imports data by region are generated by IMPLAN (partitions 1x8, 4x8, 8x1, and 8x4 from Figure 1), these can be collected, compiled, and compared. For each commodity, the discrepancy between total domestic imports and exports is identified. Obviously, one of three possibilities can occur: the two estimates can be equal, domestic imports can exceed domestic exports, or domestic exports can exceed domestic imports.

If there is no discrepancy, neither domestic nor foreign trade need be adjusted. Should domestic imports exceed domestic exports by, say 5%, then aggregate domestic imports of that commodity for each region are decreased by 5%, while foreign imports of that commodity for each region are increased by the associated dollar values. This is preferred to increasing domestic exports estimates, since the necessary exports increases could result in production levels that exceed total commodity output in a given region. Were domestic exports to be increased, foreign exports would, of course, be reduced by corresponding amounts, although there would be no guarantee that foreign exports values would be larger than the necessary reductions, which could lead to negative foreign exports values.

Should domestic exports exceed domestic imports, all domestic exports for that commodity for all regions are decreased by the appropriate percentage, with corresponding dollar values being added to the appropriate foreign exports commodity, again for all regions. This is preferred here to adjusting imports for reasons parallel to the reason for adjusting imports rather than exports in the previous paragraph.

The results of these adjustments insure individual SAM and overall ISAM consistency. The ISAM that results will have the structure shown in Figure 5 below.

Figure 5. Final Structure of the US Interregional Social Accounting Matrix

		Region 1				Region 2				Region 3				
		Industry	Commodity	Factors	Institutions	Industry	Commodity	Factors	Institutions	Industry	Commodity	Factors	Institutions	Foreign Trade
Region 1	Industry		r011x2				r01021x8				r01031x8			r011x7
	Commodity	r012x1			r012x4	r01028x1			r01028x4	r01038x1			r01038x4	
	Factors	r013x1												
	Institutions		r014x2	r014x3	r014x4		r01024x8				r01034x8			r014x7
Region 2	Industry		r02011x8				r021x2				r02031x8			r021x7
	Commodity	r02018x1			I	r022x1			r022x4	r02038x1			r02038x4	
	Factors					r023x1								
	Institutions		r02014x8				r024x2	r024x3	r024x4		r02034x8			r024x7
Region 3	Industry		r03011x8				r03021x8				r031x2			r031x7
	Commodity	r03018x1			I	r03028x1			r03028x4	r032x1			r032x4	
	Factors									r033x1				
	Institutions		r03014x8				r03024x8				r034x2	r034x3	r034x4	r034x7
	F Trade	r017x1			r017x4	r012x1			r0217x4	r037x1			r037x4	
	F Trade No Detail			r015x3				r025x3				r035x3		
	Dom Trade No Detail			r016x3				r026x3				r036x3		
		TIO	TCO	Total Factor Payments	Total Institutional Expenditures									

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)

r0#5x3 and r0#8x3 represent factor payments outside the region for which there is no detail.

Interregional trade partitions replace aggregrate domestic trade blocks 1x8, 8x1, 4x8 and 8x4.

5. 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 provided 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 non-commodity interregional transfers. Theory and methods underlying the estimation of these flows await further development.