EMSI Multi-Regional Social Accounting Matrix (MR-SAM) Modeling System

Mr. Jonathan Crapuchettes
Dr. Hank Robison
Mr. Deacon James

EMSI

(Revision 03/30/17)

Table of Contents

1. Introduction

The technical document that follows describes Economic Modeling Specialists' (EMSI) **Multi-Regional Social Accounting Matrix** (MR-SAM) in detail. The SAM represents the flow of all economic transactions in a given region. It has replaced the traditional Input-Output (I-O) module in Analyst (EMSI's web-based labor market information and forecasting tool). EMSI's old I-O module operated with some 1,000 industries, four layers of government, a single household consumption sector, and an investment sector. It was used to simulate the ripple effects (multipliers) in the regional economy as a result of industries entering or exiting the region. This model served us well for several years. Over time, however, users have substantially increased in number and have become more sophisticated—hence the need to fully develop the SAM approach.

The SAM model performs the same tasks as the old I-O tool, but does much more. Along with the same 1,000 industries, government, household and investment sectors embedded in the old I-O tool, the SAM also exhibits much more functionality and a greater amount of data. Instead of simply showing the jobs, earnings, and sales multipliers, the SAM provides details on the demographic and occupational components of jobs (16 demographic cohorts and about 750 occupations are characterized).

The compilation of data for the U.S. national SAM model incorporates several sources: **Current Population Survey** (CPS), the **American Community Survey** (ACS), the **Bureau of Economic Analysis** (BEA) **National Income and Product Accounts** (NIPAs), and the BEA Input-Output **Make and Use Tables** (MUT). In addition, several EMSI in-house data sets are used. Adding data from Oak Ridge National Labs on the cost of transportation between counties allows for the creation of **multi-regional SAMs** (MR-SAM). MR-SAMs have the ability to model the flows of goods and services between regions (i.e., a collection of counties).

This document presents detail on the EMSI MR-SAM model from the ground up. It is, of necessity, a technical document not intended for wide readership. For additional information and clarification, readers are encouraged to contact the authors at EMSI or visit www.economicmodeling.com. (For a technical textbook on input-output economics, please refer to Ronald E. Miller and Peter D. Blair, *Input-Output Analysis*.)

2. The EMSI Model

EMSI's **multi-regional social accounting matrix** (MR-SAM) modeling system is a "comparative static" type model in the same general class as RIMS II (Bureau of Economic Analysis) and IMPLAN (Minnesota Implan Group).[1] It relies on a matrix representation of industry-to-industry purchasing patterns originally based on national data which are regionalized with the use of local data and mathematical manipulation (i.e., non-survey methods). Models of this type estimate the ripple effects of changes in jobs, earnings, or sales in one or more industries upon other industries in a region.

The EMSI model shows final equilibrium impacts—that is, the user enters a change that perturbs the economy and the model shows the changes required to establish a new equilibrium. As such, it is not a

1 The MR-SAM model is thus not an "econometric" model or an econometric-I-O hybrid model such as PolicyInsight (developed by Regional Economic Models Inc.).

“dynamic” type model that shows year-by-year changes over time (as REMI’s does).

2.1. Description of the National SAM

Following standard practice, our SAM model appears as a square matrix, with each row sum exactly equaling the corresponding column sum. Reflecting its kinship with the standard Leontief input-output framework, individual SAM elements show accounting flows between row and column sectors during a chosen base year. Read across rows, SAM entries show the flow of funds into column accounts (a.k.a., “receipts” or “the appropriation of funds” by those column accounts). Read down columns, SAM entries show the flow of funds into row accounts (a.k.a., “expenditures” or “the dispersal of funds” to those row accounts).

The SAM may be broken into three different aggregation layers: broad accounts, sub-accounts, and detailed accounts. The broad layer is the most aggregate and will be covered first. Broad accounts cover between one and four sub-accounts, which in turn cover many detailed accounts. This document will not discuss detailed accounts directly because of their number. For example, in the industry broad account, there are two sub-accounts and over 1,000 detailed accounts.

2.1.1. Broad Accounts: Most-Aggregated View

Figure 1 provides the most aggregated view of the national EMSI SAM. The SAM is partitioned according to eight broad accounts or accounting groups. Lower case elements within the body of the table denote trade between two different accounting groups. Upper-case elements on the far right and bottom of the table show respective row and column sums. Rows correspond to sales and columns to purchases. As indicated earlier, the row sum and the column sum in the SAM matrix will always be equal.

z_{zz}		z_{zd}	z_{za}	z_{zg}		z_{zc}	z_{ze}	X_z	Industry
z_{vz}				z_{vg}				X_v	Production Factors
	z_{dv}			z_{dg}			z_{de}	X_d	Demographic
	z_{av}	z_{ad}		z_{ag}			z_{ae}	X_a	Accumulation
	z_{gv}	z_{gd}		z_{gg}	z_{gb}		z_{ge}	X_g	Government
							z_{be}	X_b	Trade Balance
				z_{cg}				X_c	Subsidies
z_{ez}	z_{ev}	z_{ed}	z_{ea}	z_{eg}				X_e	External
X_z	X_v	X_d	X_a	X_g	X_b	X_c	X_e		

Figure 1: Basic SAM Framework, Broad Accounts

2.1.1.1. Industry Accounts

The first accounting group, denoted with subscript "z," refers to the activity of domestic industry. The row shows sales to other domestic industries (z_{zz}), to residents of the modeled region for the purposes of personal consumption (z_{zd}), sales for the purposes of domestic investment (z_{za}), sales to domestic government (z_{zg}), receipts from government subsidies (z_{zc}), and sales to non-residents, i.e., exports (z_{ze}). The first column shows purchases from other domestic industries (z_{zz}), purchases of primary factors (z_{vz}), and the purchase of imported goods (z_{ez}). The sum of the first row equals domestic industry total gross outputs or total receipts, while the sum of the column equals domestic industry total gross inputs or total payments. Total gross outputs exactly equal total gross inputs. Note: owner-occupied dwelling sales and expenditures are counted as a separate industry and are part of the industry broad group (see Appendix B: Owner-Occupied Dwellings).

2.1.1.2. Value Added Accounts

The second accounting group, shown on the second row and down the second column, and denoted by "v," captures the value added by factors of production (labor, capital, government depreciation, and taxes) located within the modeled region. The row shows value added by factors employed in domestic industry (z_{vz}), and by factors employed in domestic government (z_{vg}). The sum of the row equals the total value added in production, which is conceptually equal to the modeled country's **gross domestic product** (GDP). The second column shows the allocation of value added to residents of the modeled region (z_{dv}), depreciation of government fixed investment (z_{av}), taxes on production and imports to government (z_{gv}), and leakages outside the country (z_{ev}).

2.1.1.3. Demographic Accounts

The third accounting group, third row and column, denoted "d," captures the income and expenditure activities of the country's residents. The row shows residents' incomes stemming from the ownership of production factors located within the country (z_{dv}), government transfer payments (z_{dg}), and incomes obtained from sources outside the modeled region (z_{de}). The sum of this row equals the total of residents' income. The column shows the allocation of residents' incomes to purchase domestically produced consumer goods (z_{zd}), savings (z_{ad}), income taxes (z_{gd}), and monies spent to acquire imported goods (z_{ed}).

2.1.1.4. Investment Account

The fourth accounting group, fourth row and column, denoted "a," captures the source and spending of funds for current account investments in the modeled region. The "a" represents accumulation, as in "capital accumulation." The row shows the allocation of government fixed investment depreciation (z_{av}), the portion of residents' incomes saved and thereby invested (z_{ad}), government set-asides for investment (z_{ag}), and investment funding from sources outside the region (z_{ae}). The column shows investment goods purchases from domestic industries (z_{za}), and the purchase of investment goods produced outside the modeled region (z_{ea}).

2.1.1.5. Government Accounts

The fifth accounting group, fifth row and column, denoted "g," captures the income and expenditures of domestic government. The row shows the portion of government funds obtained from industries through taxes on production (z_{gv}), taxes on residents, or "income taxes" as above (z_{gd}), inter-governmental transfers (z_{gg}), trade balance (z_{gb}), and taxes paid by non-resident foreigners (z_{ge}). The column shows the allocation of government funds to purchase domestically produced goods (z_{zg}), to pay for primary factor inputs, chiefly labor (z_{vg}), to residents as government transfer payments (z_{dg}), to investment set-asides (z_{ag}), to itself as inter-governmental transfers (z_{gg}), to subsidies (z_{cg}), and to the purchase of goods produced outside the modeled region (z_{eg}).

2.1.1.6. Trade Balance Account

The sixth row and column, denoted by "b," captures an imaginary sector that helps to balance a national matrix with a trade deficit. It is assumed that the balance of trade is paid by the external sector (z_{be}) and that then goes to government (z_{gb}). This is essentially government borrowing from the international community to pay for the trade deficit.

2.1.1.7. Subsidies Account

The seventh row and column, denoted by "c", captures subsidies in the modeled region. The row shows the portion of government funds allocated to subsidies (z_{cg}). The column shows how those subsidies are given to specific industries (z_{zc}).

2.1.1.8. External Account

The final account, eighth row and column, denoted "e," captures import and export activity. The row shows the leakage of funds outside the modeled country. These occur through the import of goods by industry (z_{ez}), through payments to non-resident owners of production factors located in the modeled country[2] (z_{ev}), through the import of personal consumption goods purchased by residents of the modeled country (z_{ed}), through the import of investment goods (z_{ea}), and through imported-goods purchases by government (z_{eg}). The column shows the injection of funds into the modeled country's economy from outside. These occur through the export of domestically produced goods (z_{ze}), through payments to modeled country residents for the ownership of factors located outside the modeled country[3] (z_{de}), the injection of investment funds obtained from non-resident sources[4] (z_{ae}), from foreign sources of income (taxes) paid to domestic government (z_{ge}), and trade balance income (z_{be}).

2.1.2. Sub-Account View

So far, we have discussed the EMSI SAM in a broad account view in very general terms. This next section will show the SAM in its more complex, sub-account view. **Figure 2** provides the sub-account view with 17 rows and columns. Some sub-accounts have many detailed accounts, while others have none. Whenever there is a sub-account without any detailed accounts, the sub-account is named in the singular rather than the plural (e.g., "Capital *Account*" vs. "Labor *Accounts*"). Also note that broad accounts with sub-accounts add a superscript to denote each sub-account.

2 These might be thought of as "importing" claims for payment on ownership.
3 "Exporting" claims of factor payment.
4 "Exporting" bonds or promissory notes.

Z_{zz}	Z_{zo}					${}^{l}Z_{zd}$	${}^{k}Z_{zd}$	${}^{t}Z_{zd}$	Z_{za}	Z_{zs}	Z_{zl}	Z_{zf}	Z_{zm}		Z_{zc}	Z_{ze}	$\mathbf{X}_z$	Industry
Z_{oz}	Z_{oo}						${}^{k}Z_{od}$		Z_{oa}	Z_{os}	Z_{ol}	Z_{of}	Z_{om}			Z_{oe}	$\mathbf{X}_o$	
${}^{l}Z_{vz}$										${}^{l}Z_{vs}$	${}^{l}Z_{vl}$	${}^{l}Z_{vf}$	${}^{l}Z_{vm}$				${}^{l}\mathbf{X}_v$	Production Factors
${}^{k}Z_{vz}$	${}^{k}Z_{vo}$																${}^{k}\mathbf{X}_v$	
										${}^{g}Z_{vs}$	${}^{g}Z_{vl}$	${}^{g}Z_{vf}$	${}^{g}Z_{vm}$				${}^{g}\mathbf{X}_v$	
${}^{t}Z_{vz}$	${}^{t}Z_{vo}$																${}^{t}\mathbf{X}_v$	
		${}^{l}Z_{dv}$														${}^{l}Z_{de}$	${}^{l}\mathbf{X}_d$	Demographic
			${}^{k}Z_{dv}$													${}^{k}Z_{de}$	${}^{k}\mathbf{X}_d$	
										${}^{t}Z_{ds}$	${}^{t}Z_{dl}$	${}^{t}Z_{df}$					${}^{t}\mathbf{X}_d$	
				${}^{g}Z_{av}$		${}^{l}Z_{ad}$	${}^{k}Z_{ad}$	${}^{t}Z_{ad}$		Z_{as}	Z_{al}	Z_{af}	Z_{am}			Z_{ae}	$\mathbf{X}_a$	Accumulation
					${}^{t}Z_{sv}$	${}^{l}Z_{sd}$	${}^{k}Z_{sd}$	${}^{t}Z_{sd}$			Z_{sl}	Z_{sf}		Z_{sb}			$\mathbf{X}_s$	Government
					${}^{t}Z_{lv}$	${}^{l}Z_{ld}$	${}^{k}Z_{ld}$	${}^{t}Z_{ld}$		Z_{ls}		Z_{lf}		Z_{lb}			$\mathbf{X}_l$	
					${}^{t}Z_{fv}$	${}^{l}Z_{fd}$	${}^{k}Z_{fd}$	${}^{t}Z_{fd}$						Z_{fb}		Z_{fe}	$\mathbf{X}_f$	
												Z_{mf}					$\mathbf{X}_m$	
																Z_{be}	$\mathbf{X}_b$	Trade Balance
										Z_{cs}	Z_{cl}	Z_{cf}					$\mathbf{X}_c$	Subsidies
Z_{ez}	Z_{eo}		${}^{k}Z_{ev}$			${}^{l}Z_{ed}$	${}^{k}Z_{ed}$	${}^{t}Z_{ed}$	Z_{ea}	Z_{es}	Z_{el}	Z_{ef}	Z_{em}				$\mathbf{X}_e$	External

Figure 2: National SAM in Sub-Accounts

2.1.2.1. Industry Accounts

The first sub-account, denoted with subscript "z," captures the activity of domestic industry. The row shows sales to other domestic industries (z_{zz}) and to owner-occupied dwellings (z_{zo}); residents of the modeled region for the purposes of personal consumption and for which residents pay out of their wages (${}^{l}z_{zd}$), property income (${}^{k}z_{zd}$), and transfers from government entities (${}^{t}z_{zd}$); sales for the purposes of domestic investment (z_{za}); sales to state government (z_{zs}), to local government (z_{zl}), to federal government (z_{zf}), and to the military (z_{zm}); subsidies (z_{zc}); and sales to non-residents, (i.e., exports) (z_{ze}). The first column shows purchases from other domestic industries (z_{zz}) and owner-occupied dwellings (z_{oz}); purchases of regional labor (${}^{l}z_{vz}$), returns to capital (${}^{k}z_{vz}$), taxes on production and imports (${}^{t}z_{vz}$); and the purchase of imported goods (z_{ez}).

2.1.2.2. Owner-Occupied Dwellings Account

The second sub-account, denoted "o," imputes a value to and tracks various expenditures by people who own and occupy their own residences. The row shows sales to these sub-accounts: domestic industries (z_{oz}), to other owner-occupied dwellings (z_{oo}); to residents' property income for the items purchased by the sector (${}^{k}z_{od}$); sales for the purposes of domestic investment (z_{oa}); sales to state government (z_{os}), to local government (z_{ol}), to federal government (z_{of}), and to the military (z_{om}); and sales to non-residents, i.e., exports (z_{oe}). The second column shows the purchases from domestic industries (z_{zo}), to owner-occupied dwellings (z_{oo}); payments of capital (${}^{k}z_{vo}$), purchases of government services through taxes (${}^{t}z_{vo}$); and the purchase of imported goods (z_{eo}). For a more detailed explanation of owner-occupied dwellings (OOD), please refer to Appendix B: Owner-Occupied Dwellings.

2.1.2.3. Labor Accounts

The third row and column are the start of the second broad group of accounts (i.e., production factors) and capture the earnings and expenditures of occupations. The row shows the creation of labor income by occupation in domestic industries (${}^{l}z_{vz}$); state government (${}^{l}z_{vs}$), local government (${}^{l}z_{vl}$), federal government (${}^{l}z_{vf}$), and the military (${}^{l}z_{vm}$). The column shows the allocation of this income to demographic wages (${}^{l}z_{dv}$).

2.1.2.4. Capital Account

The fourth row and column capture capital income and expenditures. The row shows capital income creation in domestic industries (${}^{k}z_{vz}$) and owner-occupied dwellings (${}^{k}z_{vo}$). The column shows the allocation of that income to resident demographic property income (${}^{k}z_{dv}$) and to absentee-owning non-residents (${}^{k}z_{ev}$).

2.1.2.5. Government Capital Account

The fifth row and column show the depreciation of government capital. The row shows the allowance for depreciation of state government (${}^{g}z_{vs}$), local government (${}^{g}z_{vl}$), federal government (${}^{g}z_{vf}$), and the military ($^{g}z_{vm}$). The column shows the expenditure of those funds for capital replacement and maintenance ($^{g}z_{av}$).

2.1.2.6. Tax Accounts

The sixth row and column are the last sub-account in the production factors broad account and capture the purchases of government services from taxes on production and imports ($^{t}z_{vz}$) and owner-occupied dwellings ($^{t}z_{vo}$). Those taxes are claimed by state government ($^{t}z_{sv}$), local government ($^{t}z_{lv}$), and federal government ($^{t}z_{fv}$).

2.1.2.7. Demographic Wage Income Accounts

The seventh row and column are the start of the demographics broad account and capture the income and expenditures of demographic wage income. The row shows the income from labor ($^{l}z_{dv}$) and income obtained from sources outside the modeled region ($^{l}z_{de}$). The column shows the expenditures on industries for the purposes of personal consumption ($^{l}z_{zd}$); investment ($^{l}z_{ad}$); personal income taxes paid to state government ($^{l}z_{sd}$), local government ($^{l}z_{ld}$), and federal government ($^{l}z_{fd}$); and imports of goods and services for personal consumption ($^{l}z_{ed}$).

2.1.2.8. Demographic Property Income Accounts

The eighth row and column capture the income and expenditures of demographic property income or **dividends**, **interest**, and **rent** (DIR). The row shows the income from capital ($^{k}z_{dv}$) and income obtained from sources outside the modeled region ($^{k}z_{de}$). The column shows expenditures for the purposes of personal consumption on industries ($^{k}z_{zd}$) and owner-occupied dwellings ($^{k}z_{od}$); investment ($^{k}z_{ad}$); corporate income taxes paid to state government ($^{k}z_{sd}$), local government ($^{k}z_{ld}$), and federal government ($^{k}z_{fd}$); and imports of goods and services for personal consumption ($^{k}z_{ed}$).

2.1.2.9. Demographic Transfer Income Accounts

The ninth row and column capture the income and expenditures of demographic transfer income from government entities. The row shows the income from state government ($^{t}z_{ds}$), local government ($^{t}z_{dl}$), and the federal government ($^{t}z_{df}$). The column shows the expenditures on industries for the purposes of personal consumption ($^{t}z_{zd}$); investment ($^{t}z_{ad}$); personal income taxes paid to state government ($^{t}z_{sd}$), local government ($^{t}z_{ld}$), and federal government ($^{t}z_{fd}$); and imports of goods and services for personal consumption ($^{t}z_{ed}$).

2.1.2.10. Investment Account

As with the broad investment account, the tenth row and column denoted “a,” captures the source and spending of funds for current account investments in the modeled region. The “a” represents accumulation, as in “capital accumulation.” The row shows income from government capital depreciation ($^{g}z_{av}$); the portion of residents’ incomes saved and thereby invested in terms of wages ($^{l}z_{ad}$), property income ($^{k}z_{ad}$), and transfers ($^{t}z_{ad}$); government set-asides for investment from state government

(z_{as}), local government (z_{al}), federal government (z_{af}), and the military (z_{am}); and investment funding from sources outside the region (z_{ae}). The column shows investment goods purchases from domestic industries (z_{za}), from owner-occupied dwellings (z_{oa}); and the purchase of investment goods produced outside the modeled region (z_{ea}).

2.1.2.11. State Government Accounts

The 11th row and column captures the income and expenditures of all state governments. The row shows the income from taxes on production and imports (${}^{t}z_{sv}$); demographic income tax from personal wages (${}^{l}z_{sd}$), corporate income (${}^{k}z_{sd}$), and personal transfers (${}^{t}z_{sd}$); income from local government (z_{sl}) and from federal government (z_{sf}); and income from the trade balance sector (z_{sb}). The column shows expenditures to industries (z_{zs}), owner-occupied dwellings (z_{os}); labor (${}^{l}z_{vs}$), depreciation (${}^{g}z_{vs}$); demographic transfers (${}^{t}z_{ds}$); investment (z_{as}); local government (z_{ls}); subsidies (z_{cs}); and imports (z_{es}).

2.1.2.12. Local Government Accounts

The 12th row and column capture the income and expenditures of all local governments. The row shows the income from taxes on production and imports (${}^{t}z_{lv}$); demographic income tax from personal wages (${}^{l}z_{ld}$), corporate income (${}^{k}z_{ld}$), and personal transfers (${}^{t}z_{ld}$); income from state government (z_{ls}) and from federal government (z_{lf}); and income from the trade balance sector (z_{lb}). The column shows expenditures to industries (z_{zl}), owner-occupied dwellings (z_{ol}); labor (${}^{l}z_{vl}$), depreciation (${}^{g}z_{vl}$); demographic transfers (${}^{t}z_{dl}$); investment (z_{al}); state government (z_{sl}); subsidies (z_{cl}); and imports (z_{el}).

2.1.2.13. Federal Government Account

The 13th row and column capture the income and expenditures of the federal government. The row shows the income from taxes on production and imports (${}^{t}z_{fv}$); demographic income tax from personal wages (${}^{l}z_{fd}$), corporate income (${}^{k}z_{fd}$), and personal transfers (${}^{t}z_{fd}$); income from the trade balance sector (z_{fb}); and income from foreign taxes (z_{fe}). The column shows expenditures to industries (z_{zf}), owner-occupied dwellings (z_{of}); labor (${}^{l}z_{vf}$), depreciation (${}^{g}z_{vf}$); demographic transfers (${}^{t}z_{df}$); investment (z_{af}); state government (z_{sf}), local government (z_{lf}), and military (z_{mf}); subsidies (z_{cf}); and imports (z_{ef}).

2.1.2.14. Military Account

The 14th row and column capture the income and expenditures of the military. Even though this sector is part of the government broad account, it is very different from the other sectors in the group. Military does not get income from taxes directly; instead it receives all its funding from the federal government. The only element in the row is the military's income from the federal government (z_{mf}). The column shows expenditures to industries (z_{zm}), owner-occupied dwellings (z_{om}), labor (${}^{l}z_{vm}$), depreciation (${}^{g}z_{vm}$), investment (z_{am}), and imports (z_{em}).

2.1.2.15. Trade Balance Account

The 15th row and column is the imaginary account added to the matrix to handle the international trade imbalance or difference between imports and exports. The row shows the income for the sector coming from the external account (z_{be}) allowing the sum of the external account row to match that of the column. The column shows the expenditure of that income to state government (z_{sb}), local government (z_{lb}), and federal government (z_{fb}).

2.1.2.16. Subsidies Account

The 16th row and column capture government subsidies. The row shows the sector's income from state government (z_{cs}), local government (z_{cl}), and the federal government (z_{cf}). The column shows the expenditure of those subsidies to industries (z_{zc}).

2.1.2.17. External Account

The 17th row and column capture the imports and exports of all other sectors, respectively. The row shows imports of industries (z_{ez}), owner-occupied dwellings (z_{eo}); capital (${}^{k}z_{ev}$); consumption paid for by wages (${}^{l}z_{ed}$), property income (${}^{k}z_{ed}$), and transfers (${}^{t}z_{ed}$); investments (z_{ea}); and purchases from state government (z_{es}), local government (z_{el}), federal government (z_{ef}), and the military (z_{em}). The column shows exports of industries (z_{ze}) and owner-occupied dwellings (z_{oe}); wages (${}^{l}z_{de}$) and property income (${}^{k}z_{de}$) from outside of the region; outside investment (z_{ae}); taxes paid by foreign entities (z_{fe}); and trade balance income (z_{be}).

2.2. Description of the Multi-Regional Aspect

Multi-regional (MR), in this context, describes a non-survey model that has the ability to analyze the transactions and ripple effects (multipliers) of not just a single region, but multiple regions interacting with each other. Regions in this case are made up of a collection of counties.

EMSI's multi-regional model is built off of gravitational flows, assuming that the larger a county's economy, the more influence it will have on the surrounding counties' purchases and sales. The equation behind this model is essentially the same that Isaac Newton used to calculate the gravitational pull between planets and stars. In Newton's equation, the masses of both objects are multiplied, then divided by the distance separating them and multiplied by a constant. In EMSI's model, the masses are replaced with the supply of a sector for one county and the demand for that same sector from another county. The distance is replaced with an impedance value that takes into account the distance, type of roads, rail lines, and other modes of transportation. Once this is calculated for every county-to-county pair, a set of mathematical operations is performed to make sure all counties absorb the correct amount of supply from every county and the correct amount of demand from every county. These operations produce more than 200 million data points. More detail on how this model is built may be found later in this document, in the section describing the **Gravitational Flows Model** (p. 31).

With the flows finalized, EMSI is able to use industry standard equations to adjust the national SAM and bring it into focus for the given region or regions. If the model being created is multi-regional, the amount and kind of transactions that occur between those regions is also calculated.

3. Data Used by the EMSI Model

To produce regional data, the EMSI model relies on a number of internal and external data sources, mostly compiled by the federal government. What follows is a listing and short explanation of our sources. The use of these data will be covered in more detail later in this document.

EMSI Data are produced from many data sources to produce detailed industry, occupation, and demographic jobs and earnings data at the local level. This information (especially sales-to-jobs ratios derived from jobs and earnings-to-sales ratios) is used to help regionalize the national matrices as well as to disaggregate them into more detailed industries than are normally available.

BEA Make and Use Tables (MUT) are the basis for input-output models in the US. The make table is a matrix that describes the amount of each commodity made by each industry in a given year. Industries are placed in the rows and commodities in the columns. The use table is a matrix that describes the amount of each commodity *used* by each industry in a given year. In the use table, commodities are placed in the rows and industries in the columns. The BEA produces two different sets of MUTs, the benchmark and the summary. The benchmark set contains about 500 sectors and is released every five years, with a five-year lag time (e.g. 2002 benchmark MUTs were released in 2007). The summary set contains about 80 sectors and is released every year, with a two-year lag (e.g. 2010 summary MUTs were released in late 2011/early 2012). The MUTs are used in the EMSI model to produce an industry-by-industry matrix describing all industry purchases from all industries.

BEA Gross Domestic Product by State (GSP) describes gross domestic product from the value added perspective. Value added is equal to employee compensation, gross operating surplus, and taxes on production and imports, less subsidies. Each of these components is reported for each state and an aggregate group of industries. This dataset is updated once per year, with a one-year lag. The EMSI model makes use of this data as a control and pegs certain pieces of the model to values from this dataset.

BEA National Income and Product Accounts (NIPA) cover a wide variety of economic measures for the nation. This dataset is updated periodically throughout the year and can be between a month and several years old depending on the specific account. NIPA data are used in many of the EMSI MR-SAM processes as both controls and seeds.

BEA Local Area Income (LPI) encapsulates multiple tables with geographies down to the county level, including CA05 (Personal income and earnings by industry). CA05 is used in several processes to help with place-of-work and place-of-residence differences, as well as to calculate personal income, transfers, dividends, interest, and rent.

BLS Consumer Expenditure Survey (CEX) reports on the buying habits of consumers along with some information as to their income, consumer unit, and demographics. EMSI utilizes this data heavily in the creation of the national demographic by income type consumption on industries.

Census of Government's (CoG) state and local government finance dataset is used specifically to aid breaking out state and local data that is reported in the MUTs. This allows EMSI to have unique production functions for each of its state and local government sectors.

Census' OnTheMap (OTM) is a collection of three datasets for the census block level for multiple years. **Origin-Destination** (OD) offers job totals associated with both home census blocks and a work census block. **Residence Area Characteristics** (RAC) offers jobs totaled by home census block. **Workplace Area Characteristics** (WAC) offers jobs totaled by work census block. All three of these are used in the commuting submodel to gain better estimates of earnings by industry that may be counted as commuting. This dataset has holes for specific years and regions. These holes are filled with Census' Journey-to-Work described later.

Census' Current Population Survey (CPS) is used as the basis for the demographic breakout data of the EMSI MR-SAM model. This set is used to estimate the ratios of demographic cohorts and their income for the three different income categories (i.e. wages, property income, and transfers).

Census' Journey-to-Work (JtW) is part of the 2000 census and describes the amount of commuting jobs between counties. This set is used to fill in the areas where OTM does not have data.

Census' American Community Survey (ACS) **Public Use Microdata Sample** (PUMS) is the replacement for Census' long form and is used by EMSI to fill the holes in the CPS data.

Oak Ridge National Lab (ORNL) **County-to-County Distance Matrix** (Skim Tree) "contains a matrix of distances and network impedances between each pair of county centroids via highway, railroad, water, and combined highway-rail paths." Also included in this set are minimum impedances utilizing the best combination of paths. This is used in EMSI's gravitational flows model that estimates the amount of trade between counties in the country.

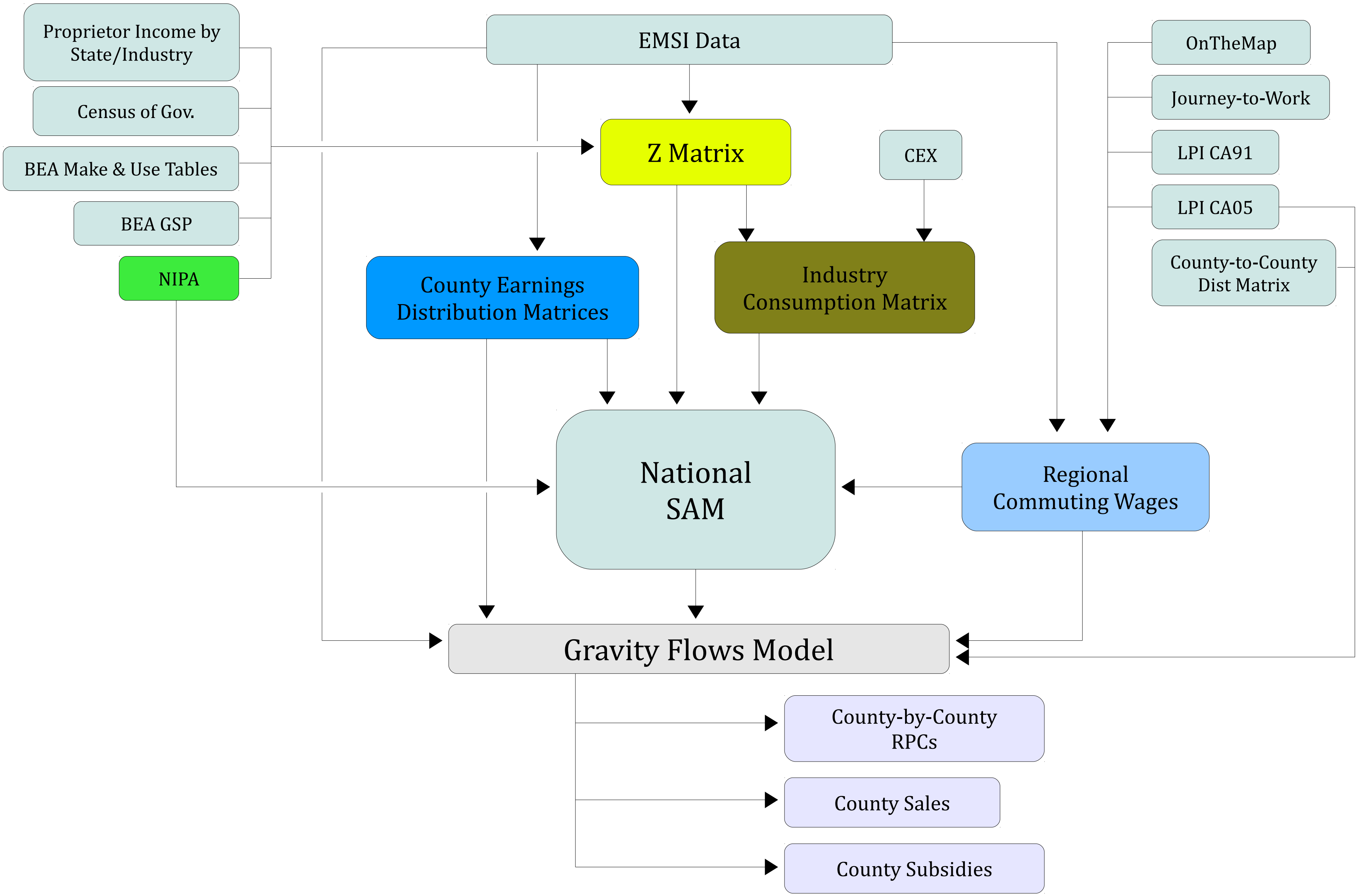


Figure 3: Dataset Map

4. Components of the EMSI Model

The EMSI MR-SAM is built from a number of different components that are gathered together to display information whenever a user selects a region. What follows is a description of each of these components and how each is created. EMSI's internally created data are used to a great extent throughout the processes described below, but its creation is not described in this document.

4.1. County Earnings Distribution Matrix

The county earnings distribution matrices describe the earnings spent by every industry on every occupation for a year. The matrices are built utilizing EMSI's industry earnings, occupational average earnings, and staffing patterns.

Each matrix starts with a region's staffing pattern matrix which is multiplied by the industry jobs vector. This produces the number of occupational jobs in each industry for the region. Next, the occupational average hourly earnings per job is multiplied by 2,080 hours, which converts the average hourly earnings into a yearly estimate. Then the matrix of occupational jobs is multiplied by the occupational annual earnings per job, converting it into earnings values. Last, all earnings are adjusted to match the known industry totals. This is a fairly simple process, but one that is very important. These matrices describe the place-of-work earnings used by the MR-SAM. **Figure 4** shows a simple example for a single industry and its occupations.

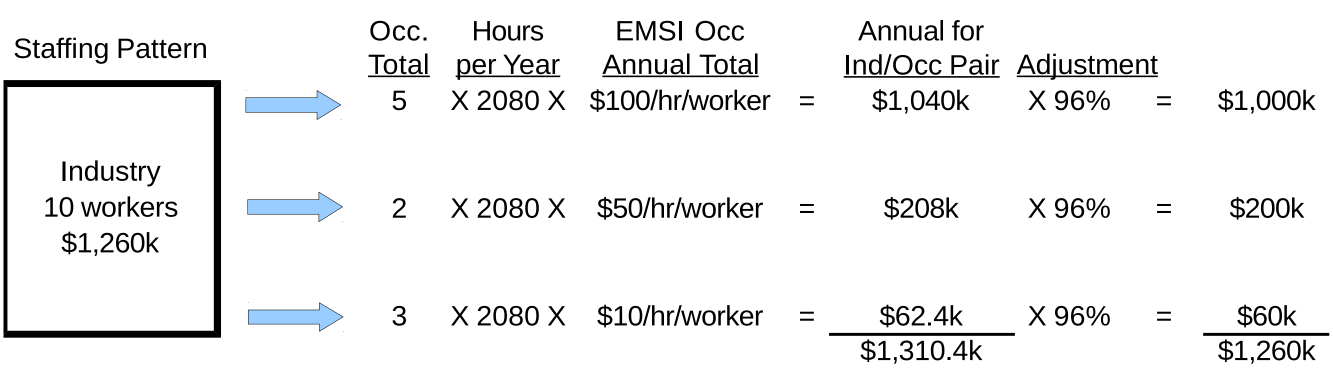


Figure 4: Occupation-By-Industry Earnings Process

4.2. Commuting Model

The commuting sub-model is an integral part of EMSI's MR-SAM model. It allows the regional and multi-regional models to know what amount of the earnings can be attributed to place-of-residence vs. place-of-work. The commuting data describe the flow of earnings from any county to any other county (including within the counties themselves). For this situation, the commuted earnings are not just a single value describing total earnings flows over a complete year, but are broken out by occupation and demographic. Breaking out the earnings allows for analysis of **place-of-residence** (PoR) and **place-of-**

work (PoW) earnings. These data are created using BLS's OnTheMap dataset, Census' Journey-to-Work, BEA's LPI CA05 tables, and some of EMSI's data. The process incorporates the cleanup and disaggregation of the OnTheMap data and the creation of finalized commuting data.

4.2.1. Origin-Destination and Journey-to-Work

The two main datasets that are used to inform the spatial component of this dataset are **Origin-Destination** (OD) and **Journey-to-Work** (JtW). Both of these were discussed previously in the data section above. Together these datasets describe flows of jobs from every county to every other county. OD is the main dataset and describes the number of private and total jobs that commute, using either three earnings ranges, three age ranges or three industry groupings for every census-block-to-census-block relationship for a year. Unfortunately, not all states have contributed to OD and therefore it is lacking some census blocks. JtW is an older dataset from the 2000 census that describes the total number of jobs commuting from one county to another, but it covers all counties and can be used to fill in the areas that are not covered by OD. The gaps in OD are filled in the following two steps:

1. Aggregating the census blocks to counties, expanding the industry groupings and converting the 2000 census data to current geographies.
2. Filling in the states' missing data for given years when the data exist for other years.

Once the OD data have been modified, they are turned into three-dimensional cubes of job data for each county-to-county pair. This is done by utilizing EMSI's detailed industry data, which reports industry-by-demographic jobs for a place of work; and by applying EMSI's demographic population data to help with the place-of-residence aspect. Once the cubes are created, a set of consistent earnings flows are calculated from the BEA's data. These steps are explained in more depth in the next sections.

4.2.1.1. OD Cleaning

The aggregation of jobs to the county level is a simple task because the first five digits of a census block code are the same as the state/county commodity flow survey code. But before the aggregation occurs, additional information must be introduced to the OD dataset.

OD comes from a larger set called OnTheMap. OnTheMap data all come raw at the census block and has three subsets: OD, **Residence Area Characteristics** (RAC), and **Workplace Area Characteristics** (WAC). RAC and WAC describe jobs by place of residence and place of work, respectively. They also describe the jobs in greater detail than OD. This detail is used to break out the three OD industry groups into 20 groups that correspond to NAICS two-digit industry codes.

Once the industry data have been broken out and the census blocks aggregated to counties, we adjust the data so that they are consistent with the current county geographies used by EMSI.

To fill in the missing data from a particular year, we copy available data from the closest year in the future. Here we assume that the commuting patterns will not change drastically between years. This allows for a complete dataset with all the necessary data.

4.2.1.2. Cubes

Now that OD has been aggregated to the county level, we combine age, earnings, and industry groups to estimate the jobs from county to county for a specific age, earnings, and industry group triplet. This estimate is accomplished through the use of a tri-proportional adjustment. The data from OD are used as the margins of the proportional, assuming that the difference between the sets of "total jobs" and "private jobs" can be ascribed to government. This gives the margins of a 3 x 3 x 21 cube, which is seeded with data from the EMSI detailed industry jobs and earnings dataset.

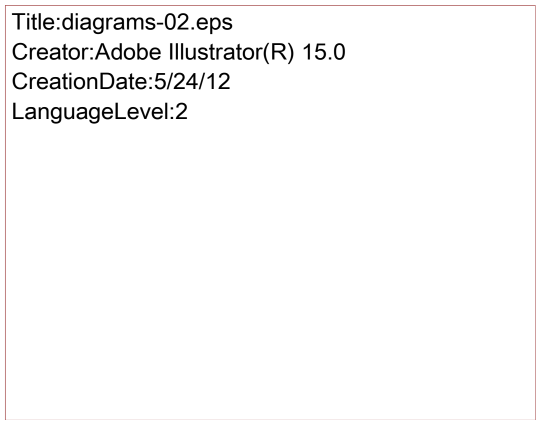


Figure 5: Jobs/Earnings Data Cube

Since the detailed data don't group jobs into earnings brackets, the average wages are calculated and the jobs that fall into the OD groups are summed, then aggregated to the OD industries. The age and gender dimensions are also summed together into the OD groups.

Since the commuting process discussed in the next section requires a complete dataset with all states accounted for, JtW cubes are created for the locations not captured by OD. The seeds for JtW cubes are created in the same way as the seeds for the OD cubes, except they are scaled to match the JtW value, once the JtW seeds are created.

4.2.2. Creating the Commuting Model

The commuting model process takes the cubes, detailed industry data, and county occupation-by-industry data and merges them into a county-by-county occupation-by-demographic earnings dataset. The process that merges these data is broken into three sections: flow creation, flow expansion, and flow conversion. (Note: This methodology is based on a reverse engineering of the residence adjustment documentation published by the BEA.[5])

4.2.2.1. Flow Creation

The creation of these flows relies on cubes and detailed jobs and earnings mentioned previously. To

[5] See part VII of Local Area Personal Income and Employment Methodology, accessed at http://www.bea.gov/regional/pdf/lapi2010.pdf

begin, the process collects all of the cubes together, along with the average PoW earnings from the detailed industry data. The aggregation and grouping of the detailed industry data is done in the same way discussed in the **Cubes** section above in order to match the dimensionality of the cubes dataset.

With these data available, the amount of earnings moved from county to county are estimated, as well as the earnings retained in each county. The first step in this estimation is calculating the ratio of PoR to PoW jobs from OD cubes, which is accomplished with a simple devision of each data point by the national total for that PoW county. We then multiply that ratio by the previously grouped detailed industry data to break it out by PoR.

We correct for any discrepancies between cubes and detailed by reinserting data points where cubes did not have any data, but detailed did. Because we have no information about the county-county flow for these data, we assume no commuting and label all earnings as retained in region (PoW is PoR).

By combining cubes and detailed industry data, we have produced a balanced data set of inflow, outflow, and retained earnings for U.S. counties. At this point each of the flows are a three age group-by-21 industry matrix. The next step is to expand these data to full demographic and industry detail.

4.2.2.2. Flow Expansion

For every flow matrix that has been created, the three OD age groups must be expanded to 16 age/gender groups and the 21 industries expanded to the full EMSI NAICS sectors. The detailed industry data (industry-by-demographic) have the full detail, so they are used for the expansion. We run into two problems when we use these data, however. First, the detailed industry data only exist in terms of PoW and the flows matrices exist in terms of both PoW and PoR. Second, the age groups don't match up. The OD age groups cover age 15 to 29, age 30 to 54, and age 55 and older. The final output age groups for the commuting model cover 15-18, 19-21, 22-24, 25-34, 35-44, 45-54, 55-64, and 65-99. Unfortunately, the OD groups cannot be split up nicely between the final groups. To handle this problem, the detailed industry data for the 25-34 group are split using percentages from the EMSI demographic population data. To align the expansion data more closely with PoR, the population percentages are pulled for the place-of-residence county for the specific flow matrix. The 25-34 group is then split into 25-29 and 30-34 year olds.

<table>
<tr><th>OD</th><th></th><th>EMSI</th></tr>
<tr><td rowspan="4">15-29</td><td>15-18</td><td>15-18</td></tr>
<tr><td>19-21</td><td>19-21</td></tr>
<tr><td>22-24</td><td>22-24</td></tr>
<tr><td>25-29</td><td rowspan="2">25-34</td></tr>
<tr><td rowspan="3">30-54</td><td>30-34</td></tr>
<tr><td>35-44</td><td>35-44</td></tr>
<tr><td>45-54</td><td>45-54</td></tr>
<tr><td rowspan="2">55+</td><td>55-64</td><td>55-64</td></tr>
<tr><td>65-99</td><td>65-99</td></tr>
</table>

Figure 6: Age Mappings

With the expansion data formatted correctly, the expansion data are proportionally adjusted to values in the flows matrix. Since these data are also supposed to represent earnings from PoR and the only data that we have for PoR are demographic population, we use the demographic population percentages (from EMSI data) and adjust the expanded data to match. Also, if a demographic that existed in the PoW doesn't exist in a PoR, the adjustment is done to make sure the earnings for that demographic are adjusted to a demographic that does exist. Remember, the expanded flows matrices are supposed to represent both the PoW and the PoR earnings.

4.2.2.3. Flow Conversion

In comparison to the previous two sections, the conversion of the industry-by-demographic data to occupation-by-demographic data is simple. Each PoW's industry-by-demographic earnings flow is multiplied by a normalized occupation-by-industry for that PoW county.

Each county's totals are scaled to match the detailed occupation totals for each county. Scaling the totals removes any inconsistencies that have existed up to this point.

After all these processes are finished, we have a finalized set of commuting earnings sufficient for use later in the regional SAMs and in total for the national SAM.

4.3. National SAM

The national SAM as described above is made up of several different components. Many of the elements already discussed are filled in with values from the national Z or transactions matrix. This matrix is built from BEA data that describe which industries make and use what commodities at the national level. These data are manipulated with some industry standard equations to produce the national Z matrix. The data in the Z matrix act as the basis for the majority of the data in the national SAM. The rest of the values are filled in with data from the county earnings distribution matrices, the commuting data, and the BEA's **National Income and Product Accounts** (NIPA).

One of the major issues that affects any SAM project is the combination of data from multiple sources that may not be consistent with one another. Matrix balancing is the broad name for the techniques used to correct this problem. EMSI uses a modification of the "diagonal similarity scaling" algorithm to balance the national SAM.[6]

4.3.1. Creating the National Z Matrix

The BEA **make and use tables** (MUTs) show which industries make or use commodities, what commodity types they make or use, and the quantities of each. EMSI combines these two tables to replace the industry-commodity-industry relationships with industry-industry relationships in dollar terms. We call this the national "Z" matrix, which shows the total amount ($) each industry purchases from others. Industry purchases run down the columns, while industry sales run across the rows.

6 *Operations Research* May/June 1990, *Michael Schneider and Stavros Zenios* "A comparative study of algorithms for matrix balancing"

Figure 7: Sample "Z" Matrix (values in $ millions)				
	Industry 1	*Industry 2*	. . .	*Industry N*
Industry 1	3.3	1,532.5	. . .	232.1
Industry 2	9.2	23.0	. . .	1,982.7
. . .	. . .	. . .	. . .	. . .
Industry N	819.3	2,395.6	. . .	0

The value 1,532.5 in this table means that Industry 2 purchases $1,532,500,000 worth of commodities and/or services from Industry 1. Most of the table is an economic double-entry accounting system, configured so that all money inflows have corresponding outflows elsewhere.

In addition to regular industries (such as "oil and gas extraction," "machinery manufacturing," "food and beverage stores," "hospitals," and so on), there are three additional rows representing labor earnings, profits, and business taxes, which together represent industry "value added" and account for the fact that industries do not spend all of their income on inputs from other industries. There are also several columns representing federal, military, and state plus local government (we later separate state plus local government into their separate sectors using Census of Government expenditures).

We create two separate Z matrices since there are two sets of MUTs—annual and benchmark. The benchmark data are produced every five years with a five-year lag and specify up to 500 industry sectors; annual data have a two-year lag but specify only about 80 industrial sectors.

The basic equation for both Z matrices is:

$$Z = V\hat{Q}^{-1}U \tag{1}$$

where *V* is the industry "make" table, Q is a vector of total gross commodity output, and U is the industry "use" table.[7]

In reality, this equation is more complex because we also need to "domesticate" the Z matrix by drawing out all imports and putting them in their own rows.[8] It is necessary to domesticate the matrix because the national model must be a closed system.

There are a number of additional modifications that need to be made to the BEA data. Most of these are related to the conversion of certain data in BEA categories to new categories that are more compatible with the datasets we use later in the process. Describing all of these modifications in detail is beyond the scope of this document, but the majority will be covered.

7 See Miller and Blair, *Input-Output Analysis*, 2nd Edition, section 5.3.6.

8 *Economic Systems Research* Vol. 13 #2 June 2001, Michael L. Lahr "Reconciling Domestication Techniques, The Nothing of Re-exports and Some Comments on Regional Accounting."

4.3.1.1. Initial Z Matrix Creation

Equation (1) is correct for a general model, but because we are creating a national model that will be regionalized, we split the data into domestic absorption and foreign imports, thereby domesticating the model.[8] The equation we use for creating a domesticated matrix (Z^D) is:

$$Z^D = V\hat{Q}^{-1}\Phi U \quad (2)$$

where Φ is a hatted vector from a Hadamard product[9] that calculates imports as a percentage of total commodity output for each industry. Formally:

$$\Phi_i = \left\{ \frac{q_i - e_i}{q_i - e_i + m_i} \right\} \quad (3)$$

where q_i is an element of the total gross commodity output vector, e_i is an element of the exports column (from the Use table), and m_i is an element of the imports column (from the Use table).

To add to the domesticated matrix, an imports matrix is created with the following equation:

$$Z^M = V\hat{Q}^{-1}\{I - \Phi\} U \quad (4)$$

The imports matrix is used at a later time in this process to help populate an imports row.

4.3.1.2. Splitting the State and Local Government Expenditures Vector

Now that we have created Z^D (2) and Z^M (4), we need to split the combined state and local government column in each matrix into four separate columns: state government, state government education, local government, and local government education. Splitting the columns gives us additional production functions for later analysis. In the Use table, the BEA gives several column vectors for state and local government. To inform the splits, we use **Census of Government** (CoG) finance expenditure data. CoG offers data about the expenditures of state and local governments for a given fiscal year. These expenditures are not at any NAICS industry level, but can be summed into the following four pairs of values from which percentages may be derived: state vs. local education consumption, state vs. local education investment, state vs. local consumption, and state vs. local investment. Each pair of percentages describes how its respective state/local vector may be split by multiplying a percentage times a vector. Once all of the vectors have been split, the respective state and local vectors are aggregated to make up four vectors: state, state education, local, and local education.

---

9 A Hadamard product is also known as an entrywise product. When multiplying vectors or matrices, dot products are not used. Instead element by element multiplication is used. A larger description may be found at http://en.wikipedia.org/wiki/Hadamard_product_(matrices).

4.3.1.3. Exports

An exports column needs to be added to the Z^D matrix. The equation used to create the exports vector is:

$$E = V \hat{Q}^{-1} [U_e] \quad (5)$$

where $[U_e]$ is the exports column from the Use table. The exports vector is then appended to the Z matrix columns.

4.3.1.4. Separating Subsidies

Recall from section 4.3.1. Creating the National Z Matrix that there are two sets of MUTs from the BEA, annual and benchmark. We separate subsidies differently for each of these data sets when creating the Z matrices. We begin with a description of the methodology for breaking out subsidies in the annual matrix.

Subsidies in the BEA data from the annual matrix are a combined part of the **Taxes on Production and Imports, less Subsidies** (TPIS) value added row in the Use table. Subsidies sometimes account for more than the total taxes of a specific industry (e.g., agriculture) and thus give TPIS a negative value. EMSI's final Z matrix does not allow for negative values. This problem can be fixed by zeroing out the negative TPIS data, but by doing so, information can be lost and other issues can arise. The solution is to split TPIS into two separate vectors—a **Taxes on Production and Imports** (TPI) vector and a separate Subsidies vector that is composed entirely of negative values. The split is based on percentages drawn from the BEA's GSP subsidies component dataset, which describe the amount of subsidies given to aggregated industries for each year. Since the subsidies are reported in the negative, we subtract rather than add them to TPIS in order to derive TPI. Without this subtraction, some of the industries that have negative TPIS would appear to have zero taxes, but after making the subtraction, we find industries that have both taxes and subsidies. The subsidies are then turned into a column vector and appended, with sign change from negative to positive, to the working Z^D matrix.

The benchmark matrix is more disaggregated than the annual matrix, so we modify TPIS differently. Negatives are removed from the current TPIS numbers and set aside. Then, the ratio of the subsidies over employee compensation is used to calculate a provisional subsidy value. If the provisional subsidy for an industry is less than the absolute value of the TPIS value for that industry, the provisional value is replaced with the original and all of the industries in the group are adjusted according to the known GSP subsidy value.

4.3.1.5. Removal of General Government Sectors and Addition of Imports

Now that we have created the two Z matrices, we need to make sure the matrices are compliant with NAICS codes. Several industries have special codes that need to be aggregated into federal, state, local and military government. These codes are **general federal government** services (GFG), **general military government** services (GMG), and **general state and local government** services (GSLG). The two matrices are aggregated at two different levels, but undergo the same changes with a small additional change for the update matrix.

GFG, GMG and GSLG are moved to their respective government columns proportionally based on the values at $Z_{GFG,Fed}$ for GFG, $Z_{GMG,Mil}$ for GMG and $Z_{GSLG,State}$ and $Z_{GSLG,Local}$ for GSLG. The equation below shows the structure (7) before and after the change for just the GSLG sector. In this set of equations, T_{GSLG} is the sum of the GSLG column and W is the value added rows of the Z matrices:

$$f_1 = \frac{Z_{\text{GSLG,State}}}{T_{\text{GSLG}}} \qquad f_2 = \frac{Z_{\text{GSLG,Local}}}{T_{\text{GSLG}}} \tag{6}$$

$$\frac{\begin{Bmatrix} Z & Z_{GSLG} & Z_{\text{State}} & Z_{\text{Local}} \\ (0) & 0 & Z_{GSLG,\text{ State}} & Z_{GSLG,\text{ Local}} \\ W & W_{GSLG} & [0] & [0] \end{Bmatrix}}{T_{GSLG}} \rightarrow \begin{Bmatrix} Z & \left[Z_{\text{State}} + \left(Z_{GSLG} * f_1\right)\right] & \left[Z_{\text{Local}} + \left(Z_{GSLG} * f_2\right)\right] \\ (W) & W_{GSLG} * f_1 & W_{GSLG} * f_2 \end{Bmatrix} \tag{7}$$

After the portions of the GSLG column have been moved to the state and local columns, the GSLG row and column are removed from the matrix. Then the same procedure is performed for GFG and GMG.

The postal service sector in the annual Z matrix is included in a sector called **general federal enterprises** (GFE) and needs to be broken out. The postal service sector (491) doesn't suppress its earnings, which means that we can create a proportion to separate the postal service's data from GFE, as seen in the following equation (Y = earnings):

$$Z_{GFE}\left(\frac{Y_{491}}{Y_{GFE}}\right) = Z_{491} \tag{8}$$

Now that we have handled the extra government sectors, we need to calculate and add an imports row to the Z matrices. For most sectors, this is done by summing across all rows of the Z matrix (including exports) and subtracting the column sums:

$$Z_R^D[1] - (1) Z_R^D = \boldsymbol{m_R} \tag{9}$$

The final demand sectors (e.g., government) are calculated by summing down the columns of the final demand sectors from the Z^M matrix:

$$(1) Z_F^M = \boldsymbol{m_F} \tag{10}$$

The reason for adding imports is twofold: 1) to balance the regular industries in the Z matrix and 2) to prepare for the block bi-proportional adjustment, since the margins in the block bi-proportional only make sense when total sales includes imports.

4.3.1.6. Value Added Changes

The BEA provides value added data using the following three components: 1) **Employee compensation** (EC), 2) **Taxes on Production and Imports, less Subsidies** (TPIS), and 3) **Gross operating surplus** (GOS) which is equal to profits plus proprietors' income. At this point in the process, subsidies have already been removed from TPIS as discussed in the **4.3.1.4. Separating Subsidies** section, so TPIS is actually TPI. The value added (VA) stack for the BEA appears as the following:

$$\begin{bmatrix} EC \\ TPI \\ GOS \end{bmatrix} \tag{11}$$

The EMSI model, on the other hand, breaks value added into the following three components: 1) **Earnings** (ER), including proprietors' income, 2) TPI, and 3) **Profits** (PRT). Thus, the VA stack for EMSI appears as follows:

$$\begin{bmatrix} ER \\ TPI \\ PRT \end{bmatrix} \tag{12}$$

To create earnings coefficients that are consistent with other EMSI data, the BEA EC is replaced by EMSI earnings (ER). BEA employee compensation data do not include proprietors' income and are not smoothed—meaning they include excessive gains and (more importantly) excessive losses. EMSI earnings, on the other hand, include proprietor earnings and are smoothed. The main reason behind this replacement is that the earnings coefficients will be multiplied by EMSI earnings to produce regional sales, which keeps the “units” of data the same.

Since GOS is the sum of profits and proprietors' income, we can easily estimate industry profits by subtracting EMSI's modified BEA **State Personal Income** (SPI) proprietors' income from GOS. GOS is not aggregated in the same way as SPI data, but we create aggregation groups that map the GOS industries to the SPI industries. The total earnings for the aggregation group are then subtracted from GOS, which provides the total profits for the aggregation group. The total profits are distributed proportionally as the value of the specific sector over the total sum for the aggregation group.

Now that GOS has been changed to include only industry profits, we need to smooth the profits to remove excessive gains and losses. The profits are smoothed to make the sales multipliers created by the EMSI model more reasonable for any given year. In order to gauge excessive profits (large or small), we start by forming a ratio of profits to value added. In effect, we form a histogram and eliminate both high and low end values, but we deliberately skew the results in the direction of high values. Values greater than 2.5 standard deviations to the right of the mean are eliminated and replaced with new values, as are values less than 1 standard deviation to the left of the mean. In algebraic terms, our adjustments are expressed as follows:

$$\bar{x}=\frac{\sum_{i}^{I}\frac{\pi_i}{VA_i}}{I} \qquad \sigma=\sqrt{\frac{\left(\sum_{i}^{I}\frac{\pi_i}{VA_i}\right)^2}{I}-\bar{x}^2} \tag{13}$$

$$\forall i\in I:\ \pi_i=\begin{cases}(\bar{x}-\sigma)*VA_i,\ \ if\ \frac{\pi_i}{VA_i}<\bar{x}-\sigma;\\(\bar{x}+2.5*\sigma)*VA_i,\ \ if\ \frac{\pi_i}{VA_i}>\bar{x}+2.5*\sigma;\\\pi_i,\ otherwise\end{cases} \tag{14}$$

Since the above equations to do not preclude negative profits, a floor must be established. We choose to limit profits to being no less than 5% of the sum of profits plus earnings:

$$\frac{\pi_i}{\pi_i+Y_i}\geq .05 \quad \text{Condition}$$

Applying some algebra yields: (15)

$$\pi_i\geq\frac{.05*Y_i}{1-.05}$$

Now that the value added rows have changed, the matrix rows are no longer balanced with the columns. In order to re-balance the matrix, we simply assume that the final sums will be equal to the column sums. Under this assumption, we know the row and column sums for the regular industries. We also know the column sums for the final demand sectors since these sectors haven't changed. We do not know the sums for the imports and exports, but we can get around this problem with a floating bi-proportional adjustment. A floating bi-proportional automatically adjusts its margins to fit the data, allowing us to work without defining the sums. Finally, we remove any negatives in the matrix and bi-proportional margins by setting the values to one-tenth of a cent. This ensures that the values remain above zero.

After the matrices have been balanced, they are both used in a process that disaggregates the data to full EMSI NAICS industries, which will be described below.

4.3.2. Disaggregation of the National Z Matrix

The previous step resulted in two national Z matrices—one based on the benchmark BEA data and the other based on the annual BEA data. These two national Z matrices are then combined and disaggregated to roughly 1,000 industry sectors. Since the annual data are more recent but the benchmark data are more detailed, combining the matrices allows us to capitalize on the strengths of both. Each initial Z matrix is disaggregated by using probability matrices to estimate industry transactions for the more detailed sectors, based on the known transactions of their parent sectors. The probability matrices are created from detailed EMSI industry earnings data, which are available for all sectors and are created using a separate process.

For years that are beyond the most recently published BEA MUTs, we project the most recent disaggregated matrix forward using data from NIPA and EMSI industry earnings.

4.3.2.1. Seeded Base Matrix and Sales Vector

Before continuing on, it must be noted that the NAICS industry levels in the benchmark data are not the same across all industries. The levels range between two-digit and six-digit depending on the industry.

The probabilities matrix is the first step used in the expansion of the two matrices. It describes an estimation (based on earnings) of the expansion of every value in the benchmark matrix down to full EMSI detailed NAICS. It is created by aggregating earnings associated with 6-digit NAICS codes up to the corresponding level of the benchmark matrix and then calculating the percentage of each earnings value over the respective aggregated total. This produces a vector of percentages, which we cross with itself to produce a matrix.

The seed base matrix is created by multiplying the probabilities matrix with values from the benchmark Z matrix. This operation results in a full-detail disaggregated Z matrix and acts as the seed base for the block bi-proportional adjustment, described in the next section.

4.3.2.2. Block Bi-proportional

The block bi-proportional that is used here is very similar to a standard bi-proportional (See **Appendix C: Proportional Adjustments Explained**), but it includes an additional factor. An aggregate matrix (i.e., the annual matrix) is added to the mix, each cell of which controls a block of values in the full detail matrix. This gives each cell of the full detail matrix three specific controls: row margin, column margin, and aggregate matrix value.

Occasionally, zeros appear in a seed base matrix block that is mapped to an annual matrix cell containing a non-zero. In this situation, the block bi-proportional will not converge. However, by blanketing all the zeros in the seed base with a very small number, we create extra connections between sectors. Not only does this solution solve the problem of non-convergence with a minimal amount of extra values, it also creates necessary connections that would not otherwise exist in the data. The very small number used to blanket the seed base is one-tenth of a cent.

Before we use the block bi-proportional, we perform one final step. We create the full detail sales vectors that will be used as the margins of the proportional. These vectors are created using EMSI's earnings for a given year and the sales-to-earnings ratios from the benchmark data. In the following equation, X is sales, Y is earnings, y is the year of the annual matrix, and b is the benchmark year.

$$X^y = Y^y \left(\frac{X^b}{Y^b} \right) \tag{16}$$

After this, the full detail sales vectors are adjusted to make sure that they are consistent with the annual matrix.

The bi-proportional also accounts for the fact that the given full detail margins may not be correct when they allocate sales to different industries in a benchmark matrix industry. As the adjustments occur, the convergence is monitored. If it halts or begins to diverge, an algorithm is introduced that reorders the sales margins to gain a better estimate, without losing the integrity of the original ratios.

4.3.2.3. Projections

Sometimes a Z matrix is created for a year that has not yet been analyzed by the BEA (or the data has not yet been produced). This matrix is called a projected Z matrix, and it is based on the most recent full-sized Z matrix that is supported by BEA Make and Use tables. Industry sales are projected forward using earnings from different years to create ratios, while non-industries use BEA NIPA data for their projection ratios. Imports are not projected directly through the use of outside values, but are calculated as the residual of the new projected column margin less the new row margin.

We create the final projected Z matrix by applying a standard bi-proportional that holds the earnings row constant. This is done to ensure that the earnings-to-sales ratios do not change.

4.3.3. Consumption Matrix

The industry consumption matrix is a national dataset used in the SAM that describes the final demand expenditures of demographics on industries. The data are based on the industry consumption vector from the national Z matrices and the **Community Expenditures Survey** (CEX). The CEX website states that it "provides information on the buying habits of American consumers, including data on their expenditures, income, and consumer unit (families and single consumers) characteristics. The survey data are collected for the Bureau of Labor Statistics by the U.S. Census Bureau."[10] The data describe different expenditures and income groups by age, income level before taxes, Hispanic or Latino origin, and race, among other indexes. These are all combined to produce an industry-by-demographic-by-income type dataset.

The consumption vector can be disaggregated into a matrix using three different sets of percentages, each slightly more detailed than the last. The most broad division is the non-industry-specific set, which separates total national expenditures into percentages by demographic. The non-industry-specific percentages are used to disaggregate the consumption vector's industries in the absence of static or CEX percentages. These are calculated as the ratio of the three income type expenditures (wages, property income, and transfers) over total expenditures.

The next set of percentages is the static set, which recognizes the expenditures of specific income types on industries. Because the static set only considers data based on demographic and income type, and not which industries are receiving expenditures, its percentages are only applied to one income type

[10] "Consumer Expenditure Survey," accessed May 30, 2012, http://www.bls.gov/cex/home.htm.

expenditure at a time. For example, “236220 - Commercial and Institutional Building Construction” is allocated only to property income expenditures, on the assumption that such an industry will only receive business from other businesses, indirectly through capital (profits) and personal income.

The static and non-industry-specific sets use CEX's Average Annual Expenditures to create percentages. The most detailed set, the industry-specific, uses CEX data, but the set is derived through a different method. Industry-specific percentages take into account the income type, demographic, and the specific industry that is receiving income. Since CEX does not specify industries in its data but only describes expenditures on products, the industry-specific percentages are created by mapping products to their respective industry (or industries) and calculating the ratio of a particular set of income type expenditures over that total product's expenditures.

Once the three different kinds of percentages are calculated, each is applied to their respective industries, creating a matrix of values.

4.3.4. Building the National SAM

All of the data populating the national SAM are derived from either EMSI or NIPA data. Since many of the elements are filled directly by values from the Z matrix, **Figure 8** shows where the data originate;

In **Figure 8** the color coding is as follows:

- Yellow represents elements filled directly, without modification, from the national Z matrix.
- Dark yellow represents elements filled by the consumption matrix.
- Blue represents elements filled by a sum of the county earnings distribution matrices.
- Light blue represents elements filled by a sum of the earnings in the commuting model.
- Dark green represents elements filled by a modification of the TPI data from the Z matrix.
- Green represents elements filled by NIPA data or a demographic disaggregation of those data.
- Gray represents elements that are filled from the sum of other elements or are calculated as a residual.
- Dark Gray represents elements that are logically zero or otherwise non-applicable.

It is important to note that the tax vector that comes out of the Z matrix process accounts for total federal, state, and local TPI. Because the SAM is a double accounting matrix, all columns must have a corresponding row. The final demand columns for government in the Z matrix are fully split out to the separate federal, military, state, and local government sectors, each with its corresponding row. Since the military does not receive taxes directly from individuals or businesses, the military row contains zeros. The splitting of the TPI vector is a simple task that applies ratios from the NIPA government tables, producing three separate vectors of taxes. Since **Owner-occupied Dwellings** (OOD) is considered an industry, its taxes are automatically split along with the rest of TPI. OOD pays property taxes, however, so the OOD taxes are resplit according to which government entity receives the property taxes. This split is identical to the way the TPI vector was broken out, except that we use personal current taxes instead of TPI.

Readers should also note that the demographic disaggregation of NIPA data is accomplished through the use of national demographic income data created by EMSI, which utilize ACS PUMS, CPS personal income (PINC) tables 1 and 9. These data are disaggregated and then used to create percentages that expand the NIPA data.

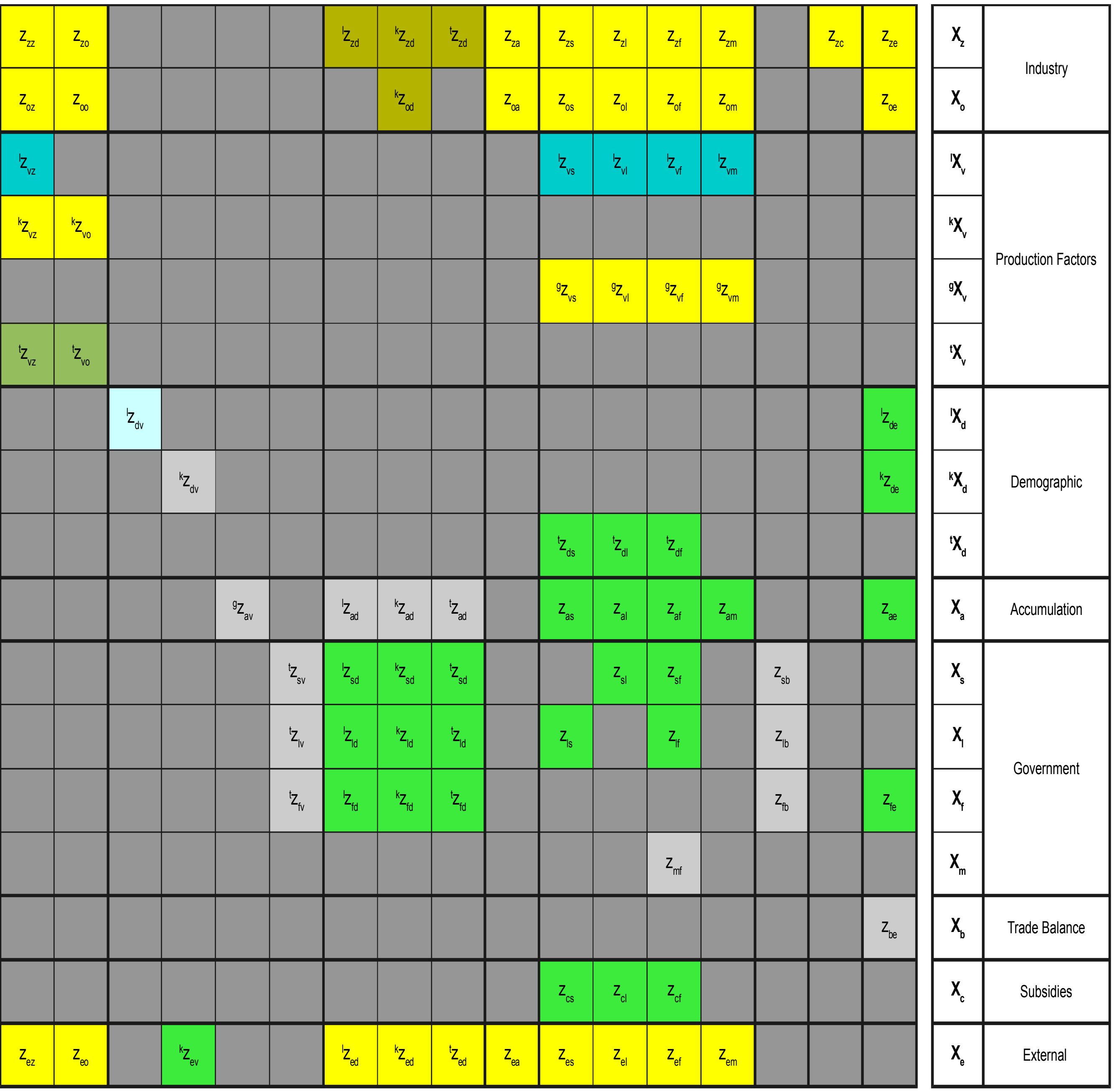


Figure 8: Origin of National SAM Data

4.4. Gravitational Flows Model

The most important piece of the EMSI MR-SAM model is the gravitational flows model that produces county sales, subsidies, taxes, and profits, and county-by-county **regional purchasing coefficients** (RPCs). County sales are the vector of total output for every sector in the SAM applied to a given county. County subsidies are an estimation of the governmental subsidies given to specific industries in a given county. County taxes and profits are an estimation of taxes on production and imports and profits by industry for a given county. RPCs estimate how much an industry purchases from other industries inside and outside of the defined region. This information is critical for calculating regional economic SAM and I-O models. As discussed earlier, the national SAM incorporates data from the national Z matrix, so from this point on, the national SAM will be referred to as the national Z SAM.

Before we explain how EMSI creates RPCs, one more concept must be introduced, namely the **A matrix**. An A matrix is mathematically derived from a Z matrix and shows the production function for each sector (i.e., what a sector requires from all other sectors in order to maintain its output). The matrix is calculated by normalizing the columns of a Z matrix with respect to the sales for that column. In other words, each column is scaled so that it sums to 1.

Figure 9: Sample "A" Matrix

	Industry 1	*Industry 2*	. . .	*Industry n*
Industry 1	.001	.112	. . .	.035
Industry 2	.097	0	. . .	.065
. . .	. . .	. . .	. . .	. . .
Industry n	.002	.076	. . .	0

Each cell value represents the percentage of a column industry's output that goes toward purchasing inputs from each row industry. So the cell containing *.112* above shows that Industry 2 spends 11.2% of its total output to obtain inputs from Industry 1. A matrices are an integral part of I-O economics and will appear many times throughout this document.

When calculating RPCs, EMSI uses two methods:

Supply/demand pool method: This method uses regional industry presence and the national A matrix to estimate the regional industry demand that remains unmet by regional industry supply. The difference is assumed to be imported or exported, which defines the basis for all RPC calculation methods.

Gravitational flows method: This is a far more complex method for estimating RPCs, but it yields multi-regional data. Gravity modeling starts with the creation of an impedance matrix that values the difficulty of moving a product from county to county. Next, the impedance matrix is converted into a base matrix that contains seeds of multi-regional flows between counties in a given sector. This base matrix is then fed to a bi-proportional with supply and demand as the row and column constraints, respectively. The result is an estimate of multi-regional flows from every county to every county. These flows are divided by each respective county's demand to produce multi-regional RPCs.

4.4.1. County Data

It is necessary to calculate sales, TPI, profits, subsidies, supply, and demand for every county and for every sector that the SAM covers. This is done using a combination of national and state data which are then regionalized to the county level based on county earnings. During the creation of these data, EMSI matches certain components of its GSP estimate – TPI, subsidies, and profits – to published BEA GSP components. In this section we describe how we combine BEA GSP data with our national SAM and state level earnings to produce state level coefficients for earnings, TPI, profits, and subsidies. These coefficients are then used in turn to calculate county data. A few things to note while reading the descriptions of the state and county calculations in the following sub-sections:

- The equations are presented in a necessary order
- **Bold** letters denote vectors
- Upper case letters denote matrices
- Lower case letter "*a*" denotes values from the national A matrix
- $\mathbf{x}^r$ is the sales vector for area *r*
- $\mathbf{y}^r$ is the earnings vector for area *r*
- Superscript *c* denotes the county and *s* denotes the state
- Terms without superscripts indicate national
- All other superscripts and subscripts denote the sub-accounts as described in **2.1.2. Sub-Account View**.

4.4.1.1. State Calculations

State level calculations are a vital step in the process of calculating county data. The standard method for estimating taxes, profits, and subsidies at the state and county level is to apply national coefficients to state and county earnings. Because the BEA publishes GSP component totals for each state, however, EMSI can use these to control the state values created through the application of national coefficients. This produces more accurate state level data which is used in turn to derive more accurate county data, while keeping the overall totals consistent with the national model.

Equations (17) to (19) show the national coefficients for taxes, profits, and subsidies. These coefficients are then applied to state earnings in equations (20) to (22) to get an initial estimate of the six-digit breakout for TPI, profits, and subsidies at the state level. Note that, if the BEA hasn't reported the working year's data, data are pulled from the closest year and then projected forward using quarterly NIPA data for the working year.

National coefficients:

$$\underset{t\,x\,n}{[A_T]} = \underset{t\,x\,n}{[Z_T]}\underset{n\,x\,n}{[\hat{\mathbf{y}}]}^{-1} \tag{17}$$

$$\underset{\pi x n}{\left[A_{\pi}\right]}=\underset{\pi x n}{\left[Z_{\pi}\right]}\underset{n x n}{\left[\hat{\boldsymbol{y}}\right]}^{-1} \quad (18)$$

$$\underset{n x 1}{\left[\boldsymbol{a}_{c}\right]}=\underset{n x n}{\left[\hat{\boldsymbol{y}}\right]}^{-1}\underset{n x 1}{\left[\boldsymbol{z}_{c}\right]} \quad (19)$$

State estimates:

$$\underset{t x n}{\left[\bar{T}^{s}\right]}=\underset{t x n}{\left[A_{T}\right]}\underset{n x n}{\left[\hat{\boldsymbol{y}}^{s}\right]} \quad (20)$$

$$\underset{\pi x n}{\left[\bar{\Pi}^{s}\right]}=\underset{\pi x n}{\left[A_{\pi}\right]}\underset{n x n}{\left[\hat{\boldsymbol{y}}^{s}\right]} \quad (21)$$

$$\underset{n x 1}{\left[\bar{\boldsymbol{c}}^{s}\right]}=\underset{n x n}{\left[\hat{\boldsymbol{y}}^{s}\right]}\underset{n x 1}{\left[\boldsymbol{a}_{c}\right]} \quad (22)$$

Using the initial estimates from equations (20) to (22) as seeds, we apply proportionals to control the state level data to both the published BEA GSP data and the national totals. As a reminder, the EMSI SAM has three tax sectors, two profits sectors[11], and one subsidy sector. This means that when we use a proportional on the tax and profits sectors, we don't use a standard bi-proportional but use a tri-proportional instead.

With the outputs of the proportionals we recalculate state value added. As a reminder, value added is equal to the sum of earnings, TPI, and profits, less subsidies. However, because subsidies are contained within their own column,[12] the value added rows in the model comprise just earnings, TPI, and profits, as shown in equation (23). Note that *v* represents value added and *i* represents a single sector.

$$v_i^s=y_i^s+t_i^s+\pi_i^s \quad (23)$$

The recalculation of value added by means of proportionals makes it necessary to also calculate sales at the state level in order to ensure that the two data sets remain compatible. Here too the standard method for calculating state sales is to apply national coefficients to state earnings. Now that we have state level value added data, however, we can approach the calculation of state level sales differently, based on the understanding that sales are equal to the sum of the total requirements of an industry on other industries, the industry's value added, and its portion of non-competitive international imports (m^s).

11 Profits in the EMSI SAM comprise profits for industries and depreciation of fixed investments for the government sectors. BEA's GSP accounts don't report profits the same way as EMSI, since the BEA reports proprietors' income together with profits (see **4.3.1.6. Value Added Changes** above).

12 Subsidies are a single row and column in the SAM. However, because subsidies are part of GRP and are reported as negative values, we use the column values (which account for expenditures) to derive the industry detail for each county. Recall that in all other cases we use the row values to derive the industry detail for our GRP calculations. As such, we calculate a vector of subsidies for each county, which we then use in our creation of regional models to replace the initially calculated column values of the regional Z matrix. For additional insight on the mechanical treatment of subsidies in the EMSI SAM, see Robison et al, 2015.

This appears in equation (24) for sector *i*. Note that industry requirements are defined as the sum of the industry's inter-industry national technical coefficients (α) times the industry's total state sales.

$$\alpha_i x_i^s + v_i^s + m_i^s = x_i^s \tag{24}$$

Equation (25) provides more detail on the calculation of m^s, which is estimated by distributing international imports to industries in each state according to their portion of national value added.

$$m_i^s = \frac{m_i}{v_i} v_i^s \tag{25}$$

Equation (26) substitutes equation (25) into equation (24).

$$\alpha_i x_i^s + v_i^s + \frac{m_i}{v_i} v_i^s = x_i^s \tag{26}$$

Finally, equation (27) simplifies and solves for x^s.

$$x_i^s = \left(\frac{v_i + m_i}{v_i (1 - \alpha_i)} \right) v_i^s \tag{27}$$

The calculation of state sales in equation (27) allows us to derive sales-to-earnings ratios for each state. These ratios are then applied to county earnings to estimate county sales. State TPI, profits, and subsidies are also divided by state sales to derive ratios that are used in estimating the corresponding county values (see equations (28) to (30)).

$$\underset{t\,x\,n}{\left[A_T^s \right]} = \underset{t\,x\,n}{\left[T^s \right]} \underset{n\,x\,n}{\left[\hat{\boldsymbol{x}}^s \right]}^{-1} \tag{28}$$

$$\underset{\pi\,x\,n}{\left[A_\pi^s \right]} = \underset{\pi\,x\,n}{\left[\Pi^s \right]} \underset{n\,x\,n}{\left[\hat{\boldsymbol{x}}^s \right]}^{-1} \tag{29}$$

$$\underset{n\,x\,1}{\left[\boldsymbol{a}_c^s \right]} = \underset{n\,x\,n}{\left[\hat{\boldsymbol{x}}^s \right]}^{-1} \underset{n\,x\,1}{\left[\boldsymbol{c}^s \right]} \tag{30}$$

4.4.1.2. County Calculations

This section presents the equations used to calculate the county level data. Equations are presented in a necessary order for each sector in the SAM.

Industry & Government

$$\left[\boldsymbol{x}_{z,g}^{c}\right]=\left[\frac{\boldsymbol{x}^{s}}{\boldsymbol{y}^{s}}\right]\otimes\left[\boldsymbol{y}_{z,g}^{c}\right] \tag{31}$$

The industry (z) and government (g) sales are the basis for much of the sales vector. Sales for these sectors are calculated using state sales-to-earnings ratios ($\left[\boldsymbol{x}^{s}/\boldsymbol{y}^{s}\right]$), times the county's earnings ($\boldsymbol{y}_{z,g}^{c}$). $\boldsymbol{x}_{z,g}^{c}$ is a vector.

Owner-occupied Dwellings

$$\lambda^{c}=\frac{PI^{c}}{PI} \tag{32}$$

$$\boldsymbol{x}_{o}^{c}=\boldsymbol{x}_{o}*\lambda^{c} \tag{33}$$

County-based OOD sales are simply national OOD sales ($\boldsymbol{x}_{o}$) scaled down by the county's LPI property income ratio (λ^{c}). $\boldsymbol{x}_{o}^{c}$ is a scalar.

Labor Income

$${}^{l}\boldsymbol{x}_{v}^{c}=\left\{{}^{l}Z_{v}^{c}\right\}[1] \tag{34}$$

Labor income (${}^{l}\boldsymbol{x}_{v}^{c}$) is a vector of occupational earnings from the county earnings distribution matrices. It is calculated as the row sums of the specific county earnings distribution matrix ($\left\{{}^{l}Z_{v}^{c}\right\}$).

Capital and Government Depreciation

$${}^{k,g}\boldsymbol{x}_{v}^{c}=A_{\pi}^{s}[\boldsymbol{x}_{z,o,g}^{c}] \tag{35}$$

Capital and government depreciation output is calculated using state ratios from equation (28) and industry, OOD, and government sales for the county ($[\boldsymbol{x}^{c}_{z,o,g}]$).

Taxes on Production and Imports

$$^{t}\boldsymbol{x}^{c}_{v} = A^{s}_{T}[\boldsymbol{x}^{c}_{z,o,g}] \tag{36}$$

Tax output is calculated in the same way as capital and government depreciation. $^{t}\boldsymbol{x}^{c}_{v}$ is a vector.

Subsidies

$$\boldsymbol{c}^{c} = [\boldsymbol{x}^{c}_{z}]\{\hat{\boldsymbol{a}}^{s}_{c}\} \tag{37}$$

$$\boldsymbol{x}^{c}_{c} = (1)[\boldsymbol{c}^{c}] \tag{38}$$

Like the rest of the GRP components, subsidies are calculated using state ratios. The subsidy output scalar ($\boldsymbol{x}^{c}_{c}$) in the county sales vector is equal to the sum of the final subsidy vector as shown in equation (38).

Investment/Accumulation

$$\boldsymbol{x}^{c}_{a} = \left(\frac{\sum L^{c}}{\sum L}\right)\boldsymbol{x}_{a} \tag{39}$$

Investment output is calculated as the national accumulation total output ($\boldsymbol{x}_{a}$) times the regional-to-national ratio of total industry employment.

Demographic Wages

$$^{w}\boldsymbol{x}^{c}_{d} = \boldsymbol{W}^{c}\left(1 + \frac{Z_{de}}{\boldsymbol{W}}\right) \qquad \boldsymbol{W}^{c} = \{^{w}Z^{c}_{dv}\}[1] \tag{40}$$

Calculating demographic wages begins with the row sum of the earnings from the county's commuting data, denoted $\boldsymbol{W}^{c}$. This factor is not sufficient because there is a portion of the nation's total demographic wages that comes from outside the country. To allow all equations to add correctly to the nation, a portion of the nation's wages must be allocated to each county. The $\frac{Z_{de}}{\boldsymbol{W}}$ part of the equation adds that portion of external wages to the county's wage output. $\boldsymbol{W}$ denotes national demographic wages.

Demographic Transfers

$$ {}^{t}\boldsymbol{x}_{d}^{c} = {}^{t}\boldsymbol{x}_{d}\left(\frac{T^{c}}{T}\right) \tag{41}$$

Demographic transfer output is an allocation of the national transfers vector (${}^{t}\boldsymbol{x}_{d}$) proportional to the amount of regional transfers reported by the BEA in the LPI data.

Demographic Property Income

$$ {}^{k}\boldsymbol{x}_{d}^{c} = {}^{k}\boldsymbol{x}_{d}\left(\frac{DIR^{c}}{DIR}\right) \tag{42}$$

Demographic property income output is an allocation of the national property income vector (${}^{k}\boldsymbol{x}_{d}$) proportional to the amount of regional DIR reported by the BEA in the LPI data.

Trade Balance

$$\boldsymbol{x}_{b}^{c} = \boldsymbol{x}_{b} * \lambda^{c} \tag{43}$$

A portion of the national trade balance is allocated to each county proportionally according to the county's PI ratio (λ , see equation (32)) from the LPI.

External

$$\boldsymbol{x}_{e}^{c} = 0 \tag{44}$$

Currently, EMSI does not allocate foreign exports directly to each county. Instead, a portion of foreign exports are aggregated with domestic exports when the regional and multi-regional models are created.

Supply and Demand

The last two equations deal with the calculation of supply and demand for the county. Supply is different from sales in that it subtracts an estimate of foreign exports from the sales vector. This is done to estimate the amount of county sales that are supplied to domestic demand.

$$\boldsymbol{s}^c = \boldsymbol{x}^c - \{\hat{\boldsymbol{x}}^c\}\{\hat{\boldsymbol{x}}\}^{-1}[Z_e] \tag{45}$$

Demand is calculated as the national A matrix times county output (i.e., sales). In this situation, the external row and column of the A matrix is filled with zeros so that the demand calculated is specifically domestic demand. The subsides row and column is also filled with zeros. After multiplying $A\boldsymbol{x}^c$, we add our estimate of subsidies, from equation (37), to complete our demand calculation.

$$\boldsymbol{d}^c = A\boldsymbol{x}^c + [\boldsymbol{c}^c] \tag{46}$$

The special handling of subsidies is necessary to allow subsidies to match the BEA's GSP component totals.

4.4.2. Allocation of Supply and Demand

Now that we have calculated supply and demand for every county, we need to allocate the absorption of supply for each county-to-county flow so that we can derive RPCs. Here we switch from operating by county to operating by sector. For each sector, we calculate RPCs with one of the following two different methods: supply-demand pool and gravity. Otherwise we assume that RPCs are 0. (A list of sub-accounts and which RPC calculation method we use for each appears in **Appendix D: Sub-account RPC methods**.) Of the two RPC calculation methods, only gravity has the ability to produce multi-regional RPCs. Supply-demand pool only produces regional RPCs. In the case where we assume RPCs to be 0, no calculations are done.

4.4.2.1. Supply-Demand Pool

The most basic method of calculating RPCs is supply-demand pool, which assumes that regional demand will be satisfied as much as possible by regional supply. If supply is greater than demand, the excess supply will be exported. If supply is less than demand, the excess demand will be imported.

$$SDP_i = \begin{cases} \dfrac{S_i}{D_i}, & if\ S_i < D_i \\ 1, & otherwise \end{cases} \tag{47}$$

4.4.2.2. Gravity Modeling

Gravity modeling, a much more complicated method, allows for the calculation of multi-regional RPCs and the estimation of trades between counties. Gravity modeling begins with the creation of an impedance matrix, which we describe more fully in the next section. Next, the impedance matrix is converted into a base matrix that contains seeds of multi-regional flows between counties for the given sector. This base matrix is fed to a bi-proportional with supply and demand as the row and column constraints, respectively. The result is an estimate of multi-regional flows from every county to every other county (including the source county). To aid the gravity model, we use data from the **Commodity Flow Survey** (CFS). In CFS, data on commodity and industry trade may be found for the nation and state-to-state levels. We use these data to estimate impedance functions, derive proportions to inform the use of ORNL's county-to-county distance matrix, and measure average distances for the trade of products. These topics will be developed later in the document.

4.4.2.3. Impedance Matrix

For each sector, an impedance matrix (ω) is created based on a set of distance impedance methods for that sector. A distance impedance method is one of the measurements reported in the Oak Ridge National Laboratory's County-to-County Distance Matrix or Skim Tree. In this matrix, every county-to-county relationship is accounted for in seven different measures: **great-circle distance** (GCD), highway impedance, rail miles, rail impedance, water impedance, highway-rail-highway impedance (HRH), and “all”. “All” is a minimum impedance that uses the lowest impedance path regardless of mode. According to the Oak Ridge web site, HRH is the only explicitly intermodal path in the set and looks at paths that “were forced to start on a highway, pass through a highway/rail terminal, then through another rail/highway terminal, and complete the trips by truck.” The site goes on to say,

> “Impedance units on each modal link were chosen to be relative to the best facilities, which start out with about 1 impedance unit per mile (and 0.9 for rural Interstates). These 'native' impedances are multiplied by intermodal adjustment factors to bring their approximate 'costs' into common units. The intermodal factors used were: 1/1 for highway 1/3.3 for rail 1/5.0 for inland barge 1/5.8 for Great Lakes 1/6.5 for marine shipping. Also, since some paths are impossible, the mileage for that path will be -1 and the impedance will be 99999.9.”[13]

When an impedance matrix is created and the inter-county impedances are loaded from these measures for a sector, two possible situations arise:

(1) CFS may report information in its NAICS-by-mode-by-distance set on the value of products transported via a mode of transportation. Weights may be calculated for each measure from these values.

(2) For all other sectors that utilize the gravity model but lack CFS information, we set primary, secondary, and tertiary modes of transportation, weighted 7/12, 4/12, and 1/12, respectively.

In both situations, an impedance measure is multiplied by its respective weight and summed to produce the inter-county impedances for the sector. We use weights so that these values account for the multiple modes of transportation that may be used by any given sector.

For intra-county impedances, the area of the county is considered as a circle and two-thirds its radius[14] is used for the impedance value.

---

13 “County Skim Tree,” revised April 2011, http://cta.ornl.gov/transnet/SkimTree.htm.

14 $A^c = \pi r^2 \rightarrow \; r = \sqrt{\frac{A^c}{\pi}}$

$$\omega^{c,c} = \frac{2}{3}\sqrt{\frac{A^c}{\pi}} \tag{48}$$

4.4.2.4. Base Matrix

Before we run the gravity model through the bi-proportional, we seed it with data from the base matrix. In gravity and spatial interaction models, there are two functional forms that we can apply to values of the impedance matrix: the exponential function (e^{-bd}) and the power function (d^{-b}), where "d" is distance and "b" is a constant.[15] While "d" is equal to $\boldsymbol{\omega}$ from the previous section, "b" is unknown. Using CFS NAICS by distance shipped and graphing distance by tonnage shipped, we find which of the functional forms best matches the industry by graphing the data and then finding the best regression form. For sectors where CFS NAICS data is unavailable, we assume that they best conform to the power function. Estimating "b" is far more difficult. As yet, there are no sources on which to base an educated guess. We solve this problem by using an iterative method, which will be discussed later. For now, let us assume b = 1.

The seeds are created based on the supply of a product from one county, the demand of the same product from another county, and the impedance (with function applied) between the two counties.

$$\forall\, r, s: \widetilde{N}^{rs} = \frac{\left(\boldsymbol{s}^r \boldsymbol{d}^s\right)}{\boldsymbol{\omega}^{rs}} \tag{49}$$

This equation assumes that the amount of absorption ($\widetilde{N}$) allocated between two counties r and s is greatest when one is pushing (selling) a product with a large amount of supply, the other is pulling (buying) the same product with a large amount of demand, and the distance between the counties is small.

4.4.2.5. Allocation/Solving

Having created the seeds, we accomplish the true allocation of supply and demand through the use of a bi-proportional adjustment. Remember that "b" is still unknown. During the allocation, we solve for "b" indirectly. When bi-proportional adjustments are run, very small values often appear. Since the county-to-county trade is an estimate, it is likely that a very small amount of trade is merely a discrepancy. In fact, the smaller the trade, the less likely it is to be real. (For example, if the trade for the fast food industry between two counties is only $7 for an entire year, it appears that only one person traveled between the counties that year and only bought one value meal. If the counties are close together, this is highly unlikely.) To minimize these potential inaccuracies, we setup the bi-proportional to zero out trades below a certain limit, but without losing the integrity of the original matrix. This can sometimes be problematic since it reduces the chances for the bi-proportional to converge. To make sure this does not happen while solving "b," we run the bi-proportional with b=1, which gives it the best chance of converging. If this operation succeeds, we know that the bi-proportional will work with

[15] Walter Isard et al., *Methods of Interregional and Regional Analysis* (Brookfield, VT: Ashgate, 1998).

“b” values greater than 1 and so we can move on to solving “b.”

The values of “b” can range between 1 and ∞ , but realistically, they exist between 1 and 50 (although each industry can vary greatly). The method for solving for “b” is dependent on an average trade distance defined by CFS or by in-house estimates when CFS data are not available. Using average trade distance allows for a smaller range (1-3000 miles vs. 1- ∞) and a better understood metric. It also allows methods developed for industries with CFS data to be reused. The optimal value for “b” is considered to be whatever value best constrains national average trade distance to within 10% of the average trade distance defined by CFS or by our in-house numbers. We solve the “b” values through a binary search iteration algorithm, which is designed to find a value by continuously halving a defined range until the specified conditions are met. After we test a “b” value through the convergence of the bi-proportional adjustment, we calculate the average trade distance in the following way:

$$\frac{\sum_i (v_i d_i)}{\sum_i v_i} \tag{50}$$

Here v_i is the value of a transaction between two counties and d_i is the great circle distance between those two counties. Comparing the averages informs the algorithm of whether “b” should be increased or decreased, since “b” is inversely related to the differences in defined and calculated average distances. If the optimal value is not found, the process repeats with a different “b” value, starting with the creation of the base seed matrix.

Once we find the optimal “b” value, we calculate the inter-county RPCs (γ). We do this by taking the county-by-county final matrix and multiplying it by the inverse of the diagonalized county demand vector, as in the following equation:

$$\{\gamma\}=\{N\}\{\hat{\boldsymbol{d}}_i\}^{-1} \tag{51}$$

This produces a matrix of county-by-county RPCs for one specific sector. We then repeat this operation for all other sectors. Once this step is finished, we are left with county sales, county subsidies, and county-by-county RPCs, adding up to approximately 220 million values.

5. Model Creation and Usages

The previous section described the components of the EMSI model and the data used to create regional and multi-regional models. This section describes how we use the data to create the models, beginning with an in-depth discussion of regional models and moving on to a less comprehensive overview of multi-regional models (multi-regional models are essentially the same as regional models but with additional information). We close the section with a discussion of some of the most common equations used to extract data from the models.

5.1. Regional Models

Regional models are simply county or ZIP code models that we aggregate together. Because the aggregated data would fill approximately 3,000 terabytes, we keep the models to a manageable size by constructing them using only the national SAM, county-by-county RPCs, county sales, county subsidies, county earnings distribution matrices, and the commuting data. For ZIP code models, we use county models as a basis and then scale them to the correct size. This will be discussed later in the document.

5.1.1. County Models

County models rely on the following six components: the national A matrix, county-by-county RPCs, county sales, county subsidies, the county earnings distribution matrix, and county-by-county wage flows from the commuting model. To produce the final regional model, we individually aggregate the county earnings distribution matrices, wage flows, county sales, and county subsidies, and then we place their respective aggregations into a regional Z matrix. We create the regional Z matrix from the national A matrix, county-by-county RPCs, and county sales. The national A matrix, also known as a technical coefficients matrix, is created from the national Z SAM by dividing all columns by their respective total sales figure. (The creation of an A matrix was discussed in detail in section 4.4.)

$$A = Z_{SAM} \{ \hat{\mathbf{x}} \}^{-1} \tag{52}$$

To regionalize the A matrix, we scale the rows by pre-multiplying the matrix by the RPCs for a specific county. Then we convert the new matrix into a regional Z SAM by multiplying it by the county's sales. We repeat this process for each county in the selected region and we aggregate all of the matrices together. The multi-regional RPCs used here are not to be confused with regional RPCs, which examine multi-county regions as a single unit without specifying or aggregating inter-county relationships. When working with multi-regional RPCs, there are RPCs for intra-county and inter-county purchases, both of which are necessary when building a regional model. A two region model with counties **a** and **b** makes use of the intra-county RPCs ($\boldsymbol{\gamma}^{aa}$, $\boldsymbol{\gamma}^{bb}$) and the inter-county RPCs ($\boldsymbol{\gamma}^{ba}$, $\boldsymbol{\gamma}^{ab}$).

$$Z_{SAM}^{r} = \left(\{ \widehat{\boldsymbol{\gamma}^{aa} + \boldsymbol{\gamma}^{ba}} \} A \{ \widehat{\mathbf{x}^{a}} \} \right) + \left(\{ \widehat{\boldsymbol{\gamma}^{ab} + \boldsymbol{\gamma}^{bb}} \} A \{ \widehat{\mathbf{x}^{b}} \} \right) \tag{53}$$

This calculation allows us to count every transaction in each county and between each county.

Before considering the regional Z SAM to be final, we need to perform the following three aggregations and replace the corresponding data in the matrix:

(1) County earnings distribution matrices are aggregated for all counties in the region and then inserted into the matrix in their correct location (${}^{1}z_{v}$).

(2) Demographic-by-occupation data for earnings generated and received in the region are

aggregated and added in their correct location (${}^{l}z_{dv}$).

(3) Subsidies for the region are aggregated and then placed in their matrix column (z_{zc}).

(4) TPI for the region are aggregated and then placed in their matrix rows (${}^{t}z_{vz}$).

(5) Profits for the region are aggregated and then placed in their matrix rows (${}^{k}z_{vz}$).

All this together gives a regional county SAM.

5.1.2. ZIP Code Models

ZIP code models are more complex and require additional calculations in their creation. First, we find all of the counties that encompass the requested ZIP codes and then create a model for those counties. Next, ZIP code sales are calculated in almost the same way as the county sales described in section **4.4.1.2**. One of the major differences is the use of ZIP code earnings and a regional ZIP-to-county-population ratio (μ).

$$\mu = \frac{Pop^{z}}{Pop^{c}} \tag{54}$$

The ratio is calculated as the ratio of total population in the ZIP code region over the total population of the encompassing counties. We use this ratio to scale owner-occupied dwellings (OOD) and demographic sales. Other differences between the creation of ZIP code and county models are outlined in the following equations. Please refer to the above section for any comparisons.

Owner-occupied Dwellings

$$\boldsymbol{x}_{o}^{z} = \boldsymbol{x}_{o}^{c}\mu \tag{55}$$

OOD output for the ZIP code region is derived by scaling OOD output for the corresponding county region by the regional population ratio.

Labor Income

$${}^{l}\boldsymbol{X}_{v}^{z} = \left[{}^{l}Z_{v}^{c}\right]\left\{\left[(1)^{l}Z_{v}^{c}\right]\right\}^{-1}\left[\boldsymbol{Y}^{z}\right] \tag{56}$$

Labor income for a ZIP region is calculated in much the same way as it is for a county region, but in this case we sum the earnings distribution matrices for the counties and then scale them to match the earnings for the ZIP region. Here we assume that the occupational breakdown of the ZIP code region is similar to the counties that encompass it.

GRP Components, not Subsidies

$$ {}^{kgt}\boldsymbol{x}_v^z = \left[{}^{kgt}A_v\,\boldsymbol{x}^z\right]\left[\left\{{}^{kgt}A_v\,\hat{\boldsymbol{x}}^c\right\}^{-1}\,{}^{kgt}\boldsymbol{x}_v^c\right] \tag{57} $$

Capital, government depreciation, and taxes for ZIP regions are also calculated in much the same way as they are for county regions, with an important difference. When creating ZIP regions, LPI and GSP data are not accessible. To compensate, we use the A matrix and actual county output to derive a ratio of actual to estimated output for each sector in the county region. We then apply this ratio to the estimated output for the ZIP code region. Scaling to the ratios in this way allows us to control the ZIP code region output in a manner that is consistent with published GSP data.

Demographics

$$ \boldsymbol{x}_d^r = \boldsymbol{x}_d^c\,\mu \tag{58} $$

Demographic income (wages, personal income, and transfers) output for the county region is scaled based on the regional population ratio to derive the ZIP code region output.

Trade Balance

$$ X_b^z = X_b^c\left(\frac{\sum ER^z}{\sum ER^c}\right) \tag{59} $$

Trade balance output for the county is scaled based on the total ZIP code region earnings over total county region earnings.

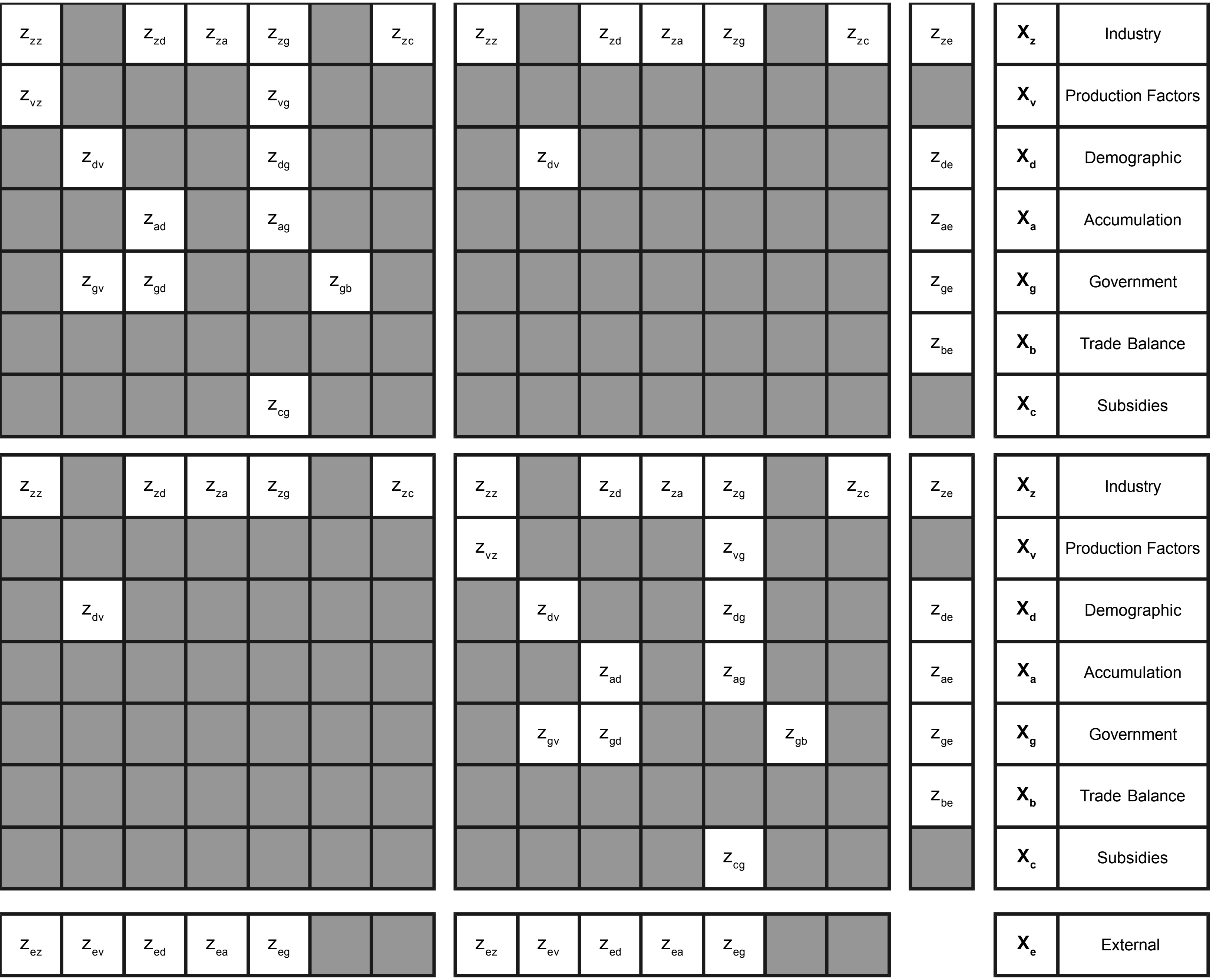
Using the new sales vector as margins, the regional county model is scaled down using a bi-proportional adjustment. All this together gives a regional ZIP code SAM.

5.2. Multi-Regional Models

A multi-regional model is able to look at trade between several different county regions. It works by creating a very large matrix with each region's model in the diagonal and inter-region trade matrices in the off-diagonals. These off-diagonal matrices are created in a similar way to the regional county matrices. The major differences are the number of zeros in the matrix and which RPCs are used. Flows between regions are only accounted for within industries (calculated with RPCs) and residence adjustment earnings (from the commuting model). The following is an example calculation for the inter-region industry trade from region r1, with counties **a** and **b**, to region r2, with counties **c** and **d**:

$$Z_{SAM}^{r1r2} = \left(\{ \hat{\boldsymbol{\gamma}^{ca} + \boldsymbol{\gamma}^{da}} \} A \{ \hat{\boldsymbol{x}^{a}} \} \right) + \left(\{ \hat{\boldsymbol{\gamma}^{cb} + \boldsymbol{\gamma}^{db}} \} A \{ \hat{\boldsymbol{x}^{b}} \} \right) \tag{60}$$

Once this off-diagonal matrix has been calculated, the residents' adjustment earnings data are placed in the matrix (${}^{l}z_{dv}$) and then the matrix is placed in its respective location in the large matrix.



Z_{zz} Z_{zd} Z_{za} Z_{zg} Z_{zc}
Z_{vz} Z_{vg}
Z_{dv} Z_{dg}
Z_{ad} Z_{ag}
Z_{gv} Z_{gd} Z_{gb}
Z_{cg}
Z_{zz} Z_{zd} Z_{za} Z_{zg} Z_{zc}
Z_{dv}
Z_{ze} Z_{de} Z_{ae} Z_{ge} Z_{be}
X_z Industry
X_v Production Factors
X_d Demographic
X_a Accumulation
X_g Government
X_b Trade Balance
X_c Subsidies
Z_{zz} Z_{zd} Z_{za} Z_{zg} Z_{zc}
Z_{dv}
Z_{zz} Z_{zd} Z_{za} Z_{zg} Z_{zc}
Z_{vz} Z_{vg}
Z_{dv} Z_{dg}
Z_{ad} Z_{ag}
Z_{gv} Z_{gd} Z_{gb}
Z_{cg}
Z_{ze} Z_{de} Z_{ae} Z_{ge} Z_{be}
X_z Industry
X_v Production Factors
X_d Demographic
X_a Accumulation
X_g Government
X_b Trade Balance
X_c Subsidies
Z_{ez} Z_{ev} Z_{ed} Z_{ea} Z_{eg}
Z_{ez} Z_{ev} Z_{ed} Z_{ea} Z_{eg}
X_e External

5.3. Usage of the Models

There are a large number of uses for regional and multi-regional SAM models, many of which have not yet been implemented. Before we discuss some of the uses and how they are calculated, we need to understand some terms that appear in the following sections. This section of the document moves away from focusing on national data and brings regional data into the foreground. *This means that variables in the following equations should be assumed to apply to regional data instead of national, unless otherwise denoted.*

We have already discussed the creation of regional Z SAMs and the national A matrix. Regional A matrices are created in the same way as the national A matrix, i.e., by normalizing the columns of the Z matrix with respect to the sales for each column. This matrix is used heavily in input-output analysis.

Using a regional A matrix, we create a new matrix (B) called the "multiplier" matrix. It describes the effect on every industry given an amount of dollars of sales to any combination of industries. One way to create these data is to add the identity matrix to the regional A matrix, then multiply the regional A matrix by itself and add it to the result. This process is continued recursively, as seen in the following equation.

$$B = I + A + A^2 + A^3 + \ldots \tag{61}$$

Wassily Leontief proved that another equation may be used to solve this problem without adding every value out to infinity. It is called the Leontief inverse method.

$$B = (I - A)^{-1} \tag{62}$$

The A matrix used in this equation is not the raw A SAM created for the region. The model can be closed in one of three ways that each produce different multipliers by zeroing out the exogenous sectors of the A matrix. The three different multipliers available in the EMSI model are Type I, Type II and Type EMSI.

- Type I multipliers are produced by closing the model only in relation to industries.
- Type II multipliers are produced by closing the model in relation to the industries, production factors, and demographics (consumption).
- Type EMSI multipliers are produced by closing the model with respect to industries, production factors, demographics (consumption), investment and portions of government based on the RPCs.

5.3.1. Regional Requirements

Regional requirements include both requirements satisfied inside the region and those satisfied outside the region (imports). They are reported for a given region with the following equations:

$$\boldsymbol{r}_{Total} = A^N \boldsymbol{x} \tag{63}$$

$$\boldsymbol{r}_{InReg} = A \boldsymbol{x} \tag{64}$$

$$\boldsymbol{m} = \boldsymbol{r}_{Total} - \boldsymbol{r}_{InReg} \tag{65}$$

The vector **m** in equation (65) is imports of particular industries. Total regional requirements are also equal to total demand for the region.

5.3.2. Regional Multipliers and Scenarios

The B matrix is used as the basis for multiplier analysis and running regional scenarios. If the B matrix represents sales, we get jobs (l) and earnings (y) multipliers using jobs-to-sales and jobs-to-earnings ratios, respectively, as shown in the following equations:

$$\boldsymbol{b}^x = (1) B \tag{66}$$

$$\boldsymbol{b}^l = \left\{ \frac{\boldsymbol{l}}{\boldsymbol{x}} \right\} B \left\{ \frac{\hat{\boldsymbol{x}}}{\boldsymbol{l}} \right\} \tag{67}$$

$$\boldsymbol{b}^y = \left\{ \frac{\boldsymbol{y}}{\boldsymbol{x}} \right\} B \left\{ \frac{\hat{\boldsymbol{x}}}{\boldsymbol{y}} \right\} \tag{68}$$

Here we can see the equations used for total sales (66), jobs (67), and earnings multipliers (68).

One of the strengths of the EMSI model is the fact that it splits multipliers into the following four different measured multiplier effects: **initial**, **direct**, **indirect**, and **induced**.[16] The initial represents the set of changes applied to the model that begin the effects. Direct multipliers are the effects caused by the initially changed sectors and describe the effects on those sectors' supply chain. Indirect multipliers extend the concept of the direct multipliers to the supply chain's supply chain. Induced multipliers describe the effects of the spending on production factors and those on demographics, investment, and government.

The following equations show the decomposition of the B matrix to give initial, direct, indirect, and induced sales multipliers. In these equations, the subscripts reference broad accounts (p. 4), B_{zz} represents a multiplier matrix closed to Type I, and F_{zz} represents a sub-matrix derived from a Type EMSI multiplier matrix.

$$Initial = (1) \tag{69}$$

[16] This methodology is consistent with Miller and Blair, *Input-Output Analysis*, 2nd Edition.

$$Direct = (1)\, A_{zz} \tag{70}$$

$$Indirect = (1)\{B_{zz} - I - A_{zz}\} \tag{71}$$

$$Induced = (1)\{F_{zz} - B_{zz}\} \tag{72}$$

Using these equations and applying the jobs and earnings ratios produces initial, direct, indirect, and induced jobs and earnings multipliers.

A major use of these multipliers is in impact analysis scenarios where a single sector or multiple sectors are shocked with an initial change and the output describes specific impacted industries. EMSI makes this output possible by translating a user's inputs into sales and then running the sales through sales multiplier matrices. In the following, Δ is the initial change vector:

$$Initial = I\, \mathbf{\Delta}_z \tag{73}$$

$$Direct = A_{zz}\, \mathbf{\Delta}_z \tag{74}$$

$$Indirect = \{B_{zz} - I - A_{zz}\}\, \mathbf{\Delta}_z \tag{75}$$

$$Induced = \{F_{zz} - B_{zz}\}\, \mathbf{\Delta}_z \tag{76}$$

Changes to government require different equations because government is not part of the industry broad account as seen in the SAM descriptions:

$$Initial = \begin{Bmatrix} 0 \\ I \end{Bmatrix} \mathbf{\Delta}_g \tag{77}$$

$$Direct = \begin{Bmatrix} A_{zg} \\ 0 \end{Bmatrix} \mathbf{\Delta}_g \tag{78}$$

$$Indirect = \begin{Bmatrix} B_{zz} A_{zg} - A_{zg} \\ 0 \end{Bmatrix} \mathbf{\Delta}_g \tag{79}$$

$$Induced = \begin{Bmatrix} F_{zg} - B_{zz} A_{zg} \\ F_{gg} - I \end{Bmatrix} \mathbf{\Delta}_g \tag{80}$$

Equation (80) is a little different from the previous equations. It is here to capture the induced feedback of government spending. The stacked equation represents a matrix with z and g rows and g columns.

The EMSI model allows users to run scenarios even when the affected industry does not exist in the analyzed region. This greatly complicates the equations since there isn't any local information about that industry. To compensate for this, national ratios are used and regionalized through the use of calculated RPCs:

$$\hat{\boldsymbol{\gamma}} = \left(\widehat{A\boldsymbol{x}}\right)\left(\widehat{A^N\boldsymbol{x}}\right)^{-1} \tag{81}$$

The RPCs that will be used to estimate the effects of the change are themselves estimated here:

$$Initial = \begin{Bmatrix} I \\ 0 \end{Bmatrix} \boldsymbol{\Delta}_g \tag{82}$$

$$Direct = \begin{Bmatrix} \hat{\boldsymbol{\gamma}}\, A_{zz}^N \\ 0 \end{Bmatrix} \boldsymbol{\Delta}_z \tag{83}$$

$$Indirect = \begin{Bmatrix} B_{zz}\hat{\boldsymbol{\gamma}}\, A_{zz}^N - \hat{\boldsymbol{\gamma}}\, A_{zz}^N \\ 0 \end{Bmatrix} \boldsymbol{\Delta}_z \tag{84}$$

$$Induced = \begin{Bmatrix} (F_{zz} - B_{zz})\hat{\boldsymbol{\gamma}}\, A_{zz}^N + F_{zv}\, A_{vz}^N \\ F_{gz}\hat{\boldsymbol{\gamma}}\, A_{zz}^N + F_{gv}\, A_{vz}^N \end{Bmatrix} \boldsymbol{\Delta}_z \tag{85}$$

These equations are possible based on a derivation that starts with equation (61), post-multiplying by $\hat{\gamma} A^N$,

$$\begin{aligned} B\hat{\gamma} A^N &= (I + A + A^2 + A^3 + \ldots)\hat{\gamma} A^N \\ &= \hat{\gamma} A^N + A\hat{\gamma} A^N + A^2\hat{\gamma} A^N + A^3\hat{\gamma} A^N + \ldots \end{aligned} \tag{86}$$

and then adding the identity matrix.

$$I + B\hat{\gamma} A^N = I + \hat{\gamma} A^N + A\hat{\gamma} A^N + A^2\hat{\gamma} A^N + A^3\hat{\gamma} A^N + \ldots \tag{87}$$

This gives us an equation that is in the form of the Leontief inverse (62) and from this, equations (82), (83), (84), and (85) may be derived.

For situations that require an earnings distribution matrix, national data are used. Since we are calculating for an industry that doesn't exist in the region, we assume the industry will have the same distribution data as the nation.

With industry results calculated, EMSI uses its regional industry-by-demographic data to translate the industry results into demographics. Occupational results are calculated by multiplying the industry results by the region's earnings distribution matrix.

5.3.3. Regional Exports

Regional exports are calculated as the difference between regional sales and regional absorption:

$$\boldsymbol{e} = \boldsymbol{x} - A\boldsymbol{x} \tag{88}$$

5.3.4. Gross Regional Product (GRP)

GRP can be calculated as total value added (earnings), taxes on production and imports less subsidies, and profits. Since all of these components are in a regional matrix, the EMSI model allows for these to be extracted:

$$\boldsymbol{y} = (1)^{l} Z_{v} \tag{89}$$

$$\boldsymbol{\pi} = (1)^{k} Z_{v} \tag{90}$$

$$\boldsymbol{t} = (1)^{t} Z_{v} \tag{91}$$

$$\boldsymbol{s} = [Z_{c}]' \tag{92}$$

$$\boldsymbol{grp} = \boldsymbol{y} + \boldsymbol{\pi} + \boldsymbol{t} - \boldsymbol{s} \tag{93}$$

5.3.5. Multi-Regional Data

Multi-regional data are currently not available but will be added in the near future.

Appendix A: Glossary

A.1. Terms

ACS	**American Community Survey**: the replacement for Census' long form; used by EMSI to fill holes in the Current Population Survey data. http://www.census.gov/acs/www/
BEA	**Bureau of Economic Analysis**: a branch of the U.S. Department of Commerce that produces economic accounts statistics by collecting source data and conducting research and analysis. www.bea.gov
BLS	**Bureau of Labor Statistics**: a branch of the U.S. Department of Labor that is the principal federal agency responsible for measuring labor market activity, working conditions, and price changes in the economy. www.bls.gov
CA05	Personal income and earnings by industry, from the Bureau of Economic Analysis. www.bea.gov/itable
CEX	**Consumer Expenditure Survey**: a survey conducted by the Bureau of Labor Statistics that reports on the buying habits of consumers along with some information as to their income, consumer unit, and demographics. EMSI utilizes these data heavily in the creation of the national demographic by income type consumption on industries. http://www.bls.gov/cex/
CFS	**Commodity Flow Survey**: provides data on commodity and industry trade for the nation and state-to-state levels. EMSI uses these data to estimate impedance functions, among other things. http://www.bts.gov/publications/commodity_flow_survey/
CoG	**Census of Government**: the Census' state & local government finance dataset is used specifically to split state and local data reported in the Make and Use Tables. This allows EMSI to have unique production functions for each of its state and local government sectors. http://www.census.gov/govs/
CPS	**Current Population Survey**: conducted by the Census of Government, the results of this survey are used for the demographic breakout data of the EMSI MR-SAM model. This set is used to estimate the ratios of demographic cohorts and their income from the three different income categories. http://www.census.gov/cps/
DIR	**Dividends, interest, and rent**
EC	**Employee compensation**
EMSI	**Economic Modeling Specialists, International**: www.economicmodeling.com
FIPS	**Federal Information Processing Standard**: a public standardization developed by the U.S. government for use in computer systems. http://www.itl.nist.gov/fipspubs/

GCD **Great-circle Distance**: the shortest distance between two points on the surface of a sphere, as measured along the surface of the sphere ("as the crow flies").

GDP **Gross Domestic Product**: the market value of all officially recognized final goods and services produced within a country in a given period. From www.wikipedia.org.

GFE **General Federal Enterprises**: for definition please refer to page 9-3 in http://www.bea.gov/national/pdf/ch9 govt ce&GI for posting.pdf

GFG **General Federal Government services**: for definition please refer to page 9-3 in http://www.bea.gov/national/pdf/ch9 govt ce&GI for posting.pdf

GMG **General Military Government services**: for definition please refer to page 9-3 in http://www.bea.gov/national/pdf/ch9 govt ce&GI for posting.pdf

GOS **Gross Operating Surplus**: a balancing item in the generation of income account representing the excess amount of money generated by incorporated enterprises' operating activities after paying labor input costs. More information available here: http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Glossary:Gross_operating_surplus_%28GOS%29_-_NA

GRP **Gross Regional Product**: the market value of all final goods and services produced within a metropolitan area in a given period of time. From www.wikipedia.org.

GSLG **General State and Local Government services**: for definition please refer to page 9-3 in http://www.bea.gov/national/pdf/ch9 govt ce&GI for posting.pdf

GSP **Gross Domestic Product by state (Gross State Product)**: the market value of all final goods and services produced within a state in a given period of time.

HRH **Highway-rail-highway impedance**: the only explicitly intermodal path in the Oak Ridge National Laboratory's County-to-county Distance Matrix. It looks at paths that start on a highway, pass through a highway/rail terminal, then through another rail/highway terminal, and complete the trips by truck. http://cta.ornl.gov/transnet/SkimTree.htm

I-O **Input-Output** analysis is the name given to an analytical framework developed by Professor Wassily Leontief in the late 1930s, in recognition of which he received the Nobel Prize in Economic Science in 1973. Taken from Miller and Blair 2nd ed.

IR **Inter-regional**: A type of analysis that uses survey based data to model the trade and dependencies of a set of regions with each other.

JtW **Journey-to-Work**: part of the 2000 government census describing the amount of commuting jobs between counties. EMSI uses this set to fill in the areas where the Bureau of Labor Statistics' OnTheMap does not have data. http://www.census.gov/hhes/commuting/

LPI **Local area Personal Income**: a dataset created by the Bureau of Economic Analysis that encapsulates multiple tables with geographies down to the county level and some with industry detail, including "Personal income and earnings by industry" (CA05). http://www.bea.gov/regional/

MUT — **Make and Use Tables**: data collected by the Bureau of Economic Analysis that show which industries make or use commodities, what commodity types they make or use, and the quantities of each. http://www.bea.gov/industry/io_benchmark.htm

MR-SAM — **Multi-Regional Social Accounting Matrix**: a "comparative static" type model that relies on a matrix representation of industry-to-industry purchasing patterns. These patterns are based on national data that are regionalized with local data and mathematical manipulation (i.e., non-survey methods).

NAICS — **North American Industry Classification System**: the standard used by Federal statistical agencies in classifying business establishments for the purpose of collecting, analyzing, and publishing statistical data related to the U.S. business economy. http://www.census.gov/eos/www/naics/

NIPA — **National Income and Product Accounts**: a set of economic accounts that provide detailed measures of the value and composition of national output and the incomes generated in the production of that output. PDF accessed at http://www.bea.gov/national

OD — **Origin-Destination database**: part of the Census Bureau's OnTheMap dataset that describes the number of private and total jobs commuting, using either three earnings ranges, three age ranges or three industry groupings for every census-block-to-census-block relationship for a year.

OOD — **Owner-occupied Dwellings**: a sub-account that captures the activity of people who own and occupy their residences. See Appendix B: Owner-Occupied Dwellings.

ORNL — **Oak Ridge National Laboratory**: www.ornl.gov

OTM — **OnTheMap**: a collection of three datasets compiled by the Census Bureau at the census block level for multiple years: the Origin-Destination set, the Residence Area Characteristics set, and the Workplace Area Characteristics set. http://onthemap.ces.census.gov/

PINC — **Personal Income tables**: from the Census's Current Population Survey. http://www.census.gov/cps/

PoR — **Place-of-residence**

PoW — **Place-of-work**

PRT — **Profits**

RAC — **Residence Area Characteristics**: part of the Census Bureau's OnTheMap dataset that describes the number of private and total jobs by place of residence, using either three earnings ranges, three age ranges or 20 industry groupings by census-block for a year.

RPC — **Regional Purchasing Coefficient**: the proportion of regional demand for a sector's output that is fulfilled from regional production. (See Ronald E. Miller and Peter D. Blair, *Input-Output Analysis* (New York: Cambridge University Press, 2009), 357.)

SAM — **Social Accounting Matrix**: an economic model that represents the flow of all economic transactions in a given region.

SPI — **State Personal Income**: personal income for individuals in a given state. EMSI use SPI data provided by the Bureau of Economic Analysis. http://www.bea.gov/regional/spi/

TPI — **Taxes on Production and Imports**

TPIS — **Taxes on Production and Imports, less Subsidies**

VA — **Value Added**

WAC — **Workplace Area Characteristics**: part of the Census Bureau's OnTheMap dataset that describes the number of private and total jobs by place of work, using either three earnings ranges, three age ranges or 20 industry groupings by census-block for a year.

A.2. Equation Variables

Z = Z matrix (1)(2)(4)(6)(8)(9)(10)(34)(40)(45)(52)(53)(56)(60)(89)(90)(91)(92)

V = Make table (1)(2)(4)(5)(50)

Q = vector of total gross community output (1)(2)(4)(5)

U = Use table (1)(2)(4)(5)

X = sales (16)(31)(33)(34)(35)(36)(37)(39)(40)(41)(42)(43)(44)(45)(46)(52)(53)(55)(56)(57)(58)(59)(60)(63)(64)(67)(68)(81)(88)

Y = earnings (8)(15)(16)(31)(56)(68)(93)

A = A matrix or technical coefficients matrix (31)(35)(36)(37)(46)(48)(52)(53)(57)(60)(61)(62)(63)(64)(70)(71)(74)(75)(78)(79)(80)(81)(83)(84)(85)(86)(87)(88)

VA = value added (13)(14)

λ = LPI property income ratio (32)(33)(43)

β = state scalars for taxes, profits, and subsidies (35)(36)(37)

L = Jobs (39)(67)

S = Supply (45)(47)(49)(93)

D = Demand (46)(47)(49)(50)(51)

ω = Impedance (48)(49)

N = Absorption (49)(51)

γ = regional purchasing coefficient (51)(53)(60)(81)(83)(84)(85)(86)(87)

B = multiplier matrix or Type I multiplier matrix (61)(62)(66)(67)(68)(71)(72)(75)(76)(79)(80)

(84)(85)(86)(87)

F = selectively typed multiplier matrix (72)(76)(80)(85)

Δ = change vector (73)(74)(75)(76)(77)(78)(79)(80)(82)(83)(84)(85)

Appendix B: Owner-Occupied Dwellings

Owner-Occupied Dwellings (OOD) is a fictitious industry that nonetheless captures a genuine income-generating activity within the economy. Recall that an important function of the SAM is to show the relationship between income creation and expenditures in the regional (and national) economy. Income (labor and property) is generated in industry and government, received by individuals and assorted institutions, and spent on the various outputs of industry and government. This is the well-known "circular flow of income and product" conveyed in introductory economics textbooks, and it is captured and conveyed well through the double-entry accounting system of the SAM model.

The existence of privately owned and occupied dwellings presents a notably missing element in the circular flow framework. For the purposes of presentation, imagine two people, A and B, with identical jobs and salary incomes and living in identical houses. These two individuals are identical in every way except one: A owns the house he lives in, while B rents. In terms of consumption, both consume the services of identical housing, and spend the rest of their disposable incomes on the usual collection of non-housing consumer goods, taxes and such. But oddly, A has more to spend on non-housing items. If they both have the same salary income, and are alike in almost every respect, how can this be the case?

The difference is that while they both have the same salary income, A actually receives an additional income (a property income) in the form of a flow of capital services from the home that he owns. A has more income, and therefore consumes a larger portion of the economy's goods and services. Without the OOD sector, this additional income and consumption of goods is not captured in the SAM.

	A	B
Income-side account		
Salary Income	100	100
TOTAL	**100**	**100**
Consumption-side account		
Consumer Goods	100	50
Rental Housing		50
TOTAL	**100**	**100**

Table 1: Without OOD Sector

	Person A	Person B
Income-side account		
Salary Income	100	100
Imputed Income	50	
TOTAL	**150**	**100**
Consumption-side account		
Consumer Goods	100	50
Rental Housing		50
Value OOD	50	
TOTAL	**150**	**100**

Table 2: With OOD Sector

Tables 1 and 2 summarize the issue and its solution. Table 1 shows an accounting of A and B's income and spending without the OOD sector. As stipulated, both receive the same salary income, shown here on the "income-side account" as $100. Both spend their incomes, but the difference is that A enjoys $100 in consumer goods while B can only afford $50, since the other $50 is paid as rent. Table 1 totals show both with salary income of $100, and both with $100 in consumption-side spending. But therein lies the problem: A is actually better off. A enjoys the same housing services as B, and the value of this can quite reasonably be viewed as an item of income, albeit income in-kind.

Table 2 shows the OOD solution. On the consumption-side account, the stream of housing services provided to A by the home he owns is given an appropriate value (in this hypothetical case, \$50) and entered as the "commodity" OOD. In the meantime, an imputed value is entered on the income-side account, which recognizes that the OOD stream of housing services is in fact an addition to A's income. Note that alternatively, the imputed value of OOD recognizes an opportunity cost for homeowners' investments in their occupied dwellings.

The above set-up details the place of the OOD sector within the SAM framework. Let us turn now to the SAM framework itself and incorporate the account. Let us start with a closer look at B's traditional rental situation which we can use as a model for constructing the OOD sector. On the consumption side of the rental transaction, B's rental payment appears in the SAM as a consumption item (i.e., consumption of rental housing, in Figure 2's vector ${}^{l}\mathbf{Z}_{zd}$.) At the same time, on the income side of the transaction, the rent payment goes to cover the landlord's expenses for repairs and maintenance (regular business inputs tracked in the rental property sector, part of Figure 2's vector $\mathbf{Z}_{zz}$). In addition, there is a property income entry, to compensate the landlord for his investment in the rental property (tracked in Figure 2's vector ${}^{k}\mathbf{Z}_{vz}$), and a business tax entry, to pay an assortment of property taxes (tracked in Figure 2's vector ${}^{t}\mathbf{Z}_{vz}$). That is how the rental of a residential dwelling appears in the accounts of the SAM.

Let us turn now to our home-owning person A, and the construction of the OOD income and consumption accounts. The first step is to impute the equivalent rental value of A's home. This imputation is carried out in the aggregate by the BEA as part of the construction of the U.S. Input-Output model. EMSI then allocates this aggregated value to counties in proportion to BEA-reported personal incomes. The imputed rental value of owner-occupied dwellings is entered into the EMSI SAM as a consumption spending item from the OOD sector (element ${}^{k}\mathbf{Z}_{od}$ in the Figure 2 SAM). So B consumes the services of rental property equal in value to her rent, ${}^{l}\mathbf{Z}_{zd}$, while A consumes an identically-valued stream of services from his own home which is entered as a consumption good purchase from the OOD sector.

The entry just described accounts for the consumption side of the OOD transaction. The next step is to fill in the associated income side. As discussed above, B's rent includes a charge to cover the landlord's expense of maintaining the rental property (which appears as a regular input expense for the rental property sector in Figure 2's vector $\mathbf{Z}_{zz}$). Homeowner A faces a similar (and in this example, identical) collection of expenses which would appear as regular consumption items in the consumption vectors (Figure 2 vectors $\mathbf{Z}_{zd}$), if it were not for the OOD sector. But these expenditures are simply moved to the OOD sector and become OOD regular input purchases (appearing in Figure 2 vector $\mathbf{Z}_{zo}$). Note that this process is actually performed by the BEA and the national I-O data is delivered with the OOD sector in place.

Two additional entries complete the construction of the OOD account. Property taxes paid by homeowner A would, in the absence of the OOD sector, appear in Figure 2's vectors showing tax payments out of disposable personal income (Figure 2 vectors $\mathbf{Z}_{sd}$, $\mathbf{Z}_{ld}$). These taxes are moved to the OOD sector and become a business tax (Figure 2 item ${}^{t}\mathbf{Z}_{vo}$). Finally, an income item equal in value to the net stream of services of OOD is entered as an item of the economy's value added (Figure 2, item ${}^{k}\mathbf{Z}_{vo}$).

The above outline provides all the essential detail on the purpose and placement of the OOD sector within our SAM framework, but one last issue remains. The modern input-output framework underlying the U.S. National Input-Output Model is based on a commodity-industry, or "Make and Use Table" format. The distinction between commodities and industries stems mainly from the old primary-secondary output problem that, prior to the Make and Use format, troubled I-O model builders. For example, consider two commodities: lodging and restaurants. Provided these are delivered in separate and entirely distinct businesses, no problem. But how does the I-O model builder handle a case where a large hotel includes its own restaurant? The answer is to construct a Make Table, showing all the commodities produced by a given industry. Accordingly, the “Lodging” industry will mainly produce the commodity “Lodging,” its “primary output,” but will also produce a small amount of the commodity “Restaurants,” a “secondary output.”

In the EMSI SAM, with its industry-by-industry structure, Make and Use Tables are combined.[17] One result of this combination is that a given “industry” will exhibit an input structure that reflects the input requirements of its component secondary outputs. In the case of the Lodging sector example above, the inputs of Lodging will include inputs needed for Restaurants (e.g., food items), because the Lodging sector includes Restaurants as a secondary output.

So how does this impact our OOD sector? In constructing the U.S. I-O Model, the BEA shows as secondary outputs of the OOD industry two non-OOD commodities: “Residential permanent site single- and multi-family structures,” and “Residential maintenance and repair.” These two sectors respectively account for 1.61% and 1.55% of the output of the OOD industry. The other 98.84% is the “commodity” OOD. The complicating up-shot of their inclusion is that there will be small entries for the “OOD industry” in parts of the SAM other than those discussed above.

[17] The mathematics for combining Make and Use tables is well known, and presented for example in Miller and Blair, *Input-Output Analysis*, 2nd Edition, section 5.3.6.

Appendix C: Proportional Adjustments Explained

Throughout this document, proportionals of various types are used in many different ways. They are used to aggregate, project, and adjust vectors, matrices, and cubes. The types of proportionals are discussed in this section. First, we will cover the concept of proportionally adjusting a vector of data—the foundation on which all proportionals are built. Then we will cover bi-proportionals and their variants and lastly, tri-proportionals and quad-proportionals.

C.1. Proportionals

A proportion is a share or part of a whole. This definition is fundamental to the concept of proportional adjustments. To proportionally adjust a vector, all of the elements in the vector must be adjusted so that their total matches a specific value, but without losing the proportions that already exist between their separate values. The best way to understand how to proportionally adjust a vector is with an example:

2
3
5

Figure 11: Initial Vector

If we want that vector to sum to 50, but the current sum of the vector is 10, we calculate a ratio of the target sum over its current total: 50 / 10 = 5. Multiplying each element in the vector by this ratio produces the vector in figure 12,

10
15
25

Figure 12: Final Vector

and gives us the new total of 50. Now the vector has been proportionally adjusted or “scaled.” This is an example of a one dimensional problem.

C.2. Bi-Proportionals

A bi-proportional uses the same concepts that were used to adjust the vector, but takes those concepts to a two dimensional mathematical problem. Bi-proportionals are also known as the RAS scaling method. In general, bi-proportionals solve problems where a matrix exists with some initial values and two known vectors of margins. Just as we projected a sum of 50 for the vector above, which allowed us to proportionally adjust it, the margins of a bi-proportional define what the final row and column sums of the matrix should be.

				Row Sums	Row Margin
	5	1	10	16	6
	7	2	5	14	12
	10	1	6	17	13
Column Sums	22	4	21		
Column Margin	12	8	11		

Figure 13: Pre Bi-Proportional

Here is the technical description of the problem:

Given an m x n non-negative matrix A and positive vector (margins) **u** and **v**, we need to find a new matrix B that is "similar" to A such that

$$\forall i \in m: \sum_{j}^{n} b_{ij} = \boldsymbol{u}_i$$
$$\forall j \in n: \sum_{i}^{m} b_{ij} = \boldsymbol{v}_j \quad (94)$$
$$\text{and } b_{ij} > 0 \equiv a_{ij} > 0$$

One of the things that usually is not stated in the descriptions of bi-proportionals is the fact that **u** and **v** must both sum to the same value.

$$\sum_{i}^{m} \boldsymbol{u}_i = \sum_{j}^{n} \boldsymbol{v}_j \quad (95)$$

To solve this problem, the bi-proportional alternates between proportionally adjusting the rows and the columns of the matrix. Another way to describe this is that each row is multiplied by a positive scalar so that the row has the desired sum (**u**). Then the same operation is performed on the columns until they reach the desired column sums or column margins (**v**). Of course, scaling the columns changes the row sums and similarly, scaling the rows changes the columns sums. The algorithm alternates between scaling the rows and columns until the problem is satisfied and the matrix has converged. During the iterations of the algorithm, checks are done to make sure that the error between the row sums and their margins or column sums and their margins is shrinking. If that error grows, we say that the matrix is diverging.

				Row Sums	Row Margin
	1.4984	1.1289	3.3727	6	6
	4.1663	4.4844	3.3493	12	12
	6.3353	2.3866	4.278	13	13
Column Sums	12	8	11		
Column Margin	12	8	11		

Figure 14: Solved Bi-Proportional

To full converge the example above, the bi-proportional made five row and column adjustment passes. As can be seen, proportional adjustments often result in non-integer values. Depending on the size of the problem at hand and how close the initial matrix row and column sums are to the margins, the number of required passes may be significantly higher.

C.2.1. Other Kinds of Bi-Proportionals

In the creation of the data for EMSI's model, two modifications of the bi-proportional are used specifically: the floating bi-proportional and the block bi-proportional.

The floating bi-proportional is essentially a standard bi-proportional, but with one small change. In the event that an element in each of the margins is unknown, a standard bi-proportional cannot adjust that row and column. A floating bi-proportional extends a standard bi-proportional by allowing a specific row and column to be adjusted indirectly. As an example of this, we can take the previous matrix and margins shown in figure 13 and ignore the final element of the row and column margin (the values 13 and 11, respectively). When the floating bi-proportional adjusts columns 1 and 2, the values in row 3 are adjusted. When rows 1 and 2 are adjusted, the values in column 3 are adjusted. This is how the last row and column are adjusted indirectly. (Note: in this example, the initial value of 6, int the matrix, will not change.)

Another modification of the standard bi-proportional is the block bi-proportional. In the event of unknown margins, the block bi-proportional goes in the opposite direction of the floating bi-proportional. Instead of relaxing constraints, the block bi-proportional adds a new one: the aggregate matrix. This additional matrix is more aggregated than the base matrix and acts like a margin, but in two dimensions. Each of the elements in the aggregate matrix is associated with some elements in the base matrix. After a row and column pass, a block adjustment is done to force blocks of elements in the base matrix to match the respective element out of the aggregate matrix.

Figure 15: Aggregate matrix being projected on the base matrix

There are two restrictions that the aggregate matrix and margins must adhere to in order to ensure that a block bi-proportional will converge: (1) the mass of the aggregate matrix must match the sums of the row and column margins and (2) the sum of the blocks in the margins of the block bi-proportional must match the row and column sums of the aggregate matrix.

C.3. Other Proportionals

Tri-proportionals take the concept of the bi-proportional into the third dimension. The only difference is an additional adjustment: depth. Tri-proportionals operate on matrices with three factors (cubes), and then do row, column, and depth adjustments. This also means that there are three sets of margins instead of two.

The quad-proportional used in this document doesn't operate in four dimensions. It operates on a matrix, but does have four sets of margins. Two control the entire row or column and the other two control rows or columns, but not the diagonal values.

Appendix D: Sub-account RPC methods

Sub-account	RPC Calculation Method
Industries (z)	Gravity
General Government Industries	Assumed 0
Owner-Occupied Dwellings (o)	Supply-Demand Pool
Labor Income (vl)	Known from Occupation-By-Industry data
Capital (vk)	Supply-Demand Pool
Government Capital (vg)	Supply-Demand Pool
Taxes (vt)	Supply-Demand Pool
Demographic Wages (dw)	Known from commuting model
Demographic Property Income (dk)	Supply-Demand Pool
Demographic Transfers (dt)	Assumed 0
Accumulation (a)	Supply-Demand Pool
Local Government (l)	Supply-Demand Pool
State Government (s)	Gravity only in state with b = 1
Federal Government (f)	Assumed 0
Military (m)	Assumed 0
Trade Balance (b)	Assumed 0
Subsidies (c)	Assumed 0
External (e)	Assumed 0