EMSI I-O Model Documentation:

Establishing the NAICS 6-digit US Base Model

M. Henry Robison, Ph.D.
Chief Economist
EMSI
Moscow, Idaho

Timothy Nadreau
Research Faculty and Ph.D. Student
Washington State University
Pullman, Washington

Jonathan Crapuchettes
Senior Developer
EMSI
Moscow, Idaho

Dec 1^{st}, 2015

1. Introduction

This document provides formal documentation of the EMSI Input-Output (I-O) model construction process through formation of the national model we use as the basis for constructing regional I-O models. The intended audience is a person with considerable familiarity with I-O theory, looking for a general understanding of our specific method, perhaps for comparison with other non-survey models (IMPLAN, RIMSII, R/Econ, or REMI, for example). By "formal documentation" we mean expressed in equations where possible and otherwise free of the specifics a computer programmer might need to duplicate our approach. For those interested in more specific detail, we have provided the companion paper Crapuchettes et al., 2012.

Until now EMSI I-O model documentation has been limited to Crapuchettes et al., 2012, and available only after signing a non-disclosure agreement. This troubled some clients and we recognized it as self-defeating — outside review can only lead to improvement. Accordingly, we are making our methods freely accessible through online availability of both the present paper and Crapuchettes et al., 2012.

A comprehensive documentation of our process can be envisioned in three parts: 1) Establishing the US Base Model (the topic of the present paper), 2) A Gravity Approach to Regionalizing the National Model, and 3) The EMSI Approach to Social Accounts. Note that all three parts are touched on to some degree in Crapuchettes et al., 2012. The additional and more formal coverage of part 1 (Establishing the US Base Model) conveyed in the present paper is motivated in large degree by a novel feature of the EMSI I-O model, namely an expanded US National I-O technology matrix that covers each of the roughly

1,000 NAICS 6-digit industry classification codes.

Like virtually all generally available regional input-output models in the United States, the EMSI I-O model is based on the US Department of Commerce, Bureau of Economic Analysis (BEA) US National Input-Output Model. Thus the characterization "data reduction" applied to the family of modeling techniques that in effect reduce the national model to its sub-national (i.e., zip code, county, or multi-county) components. To serve as the basis for the EMSI regional I-O modeling process, it is necessary to reconfigure the "off-the-shelf" BEA Model in a number of ways. We refer to the reconfigured model as our "US Base Model."

We launch our documentation with the BEA Model expressed in formal notation. We follow this with three key reconfigurations. The first has the effect of ridding the national model's production functions of foreign imports. The second moves subsidies from an embedded location (embedded with "taxes on production and imports") to a location of its own. The third reconfiguration expands the sectors of the BEA Model to the NAICS 6-digit detail of the overall EMSI data set.

2. BEA Source Arrays[1]

Equations (1) through (5) show the fundamental source arrays of what we earlier termed the "off-the-shelf" BEA Model. That model conveys detail on m commodities, n industries, and f final demand sectors. Parenthetical terms beneath selected arrays indicate row-column dimensions. Where the dimensions are obvious, parenthetical expression is omitted.

$$\underset{(m\times n)}{U}\,1 + \underset{(m\times f)}{\bar{F}}\,1 + \underset{(m\times 1)}{\bar{E} - \bar{M}} = \underset{(m\times 1)}{q} \tag{1}$$

$$(\ 1\ )\,U + (\ 1\ )\underset{(3\times n)}{w} = \underset{(1\times n)}{X'} \tag{2}$$

$$(\ 1\ )\underset{(n\times m)}{V} = \underset{(1\times m)}{q'} \tag{3}$$

$$\underset{(n\times m)}{V}\,1 = \underset{(n\times 1)}{X} \tag{4}$$

$$\underset{(n\times m)}{D} = V\ \hat{q}^{-1} \tag{5}$$

Let us examine the model equation by equation. Equation (1) shows absorption of commodities in the national economy. Matrix U is the national "use matrix" showing the absorption of m commodities (shown on rows) by n industries shown on columns. Matrix $\bar{F}$ shows the absorption of m commodities by the f final demand sectors. $\bar{E}$ shows the nation's export of

[1] The formal presentation of the national model appearing in this section is well known. While freely adopting our own notation, wherever possible our presentation follows that of Miller and Blair (2009).

commodities while $\bar{M}$ shows its commodity imports. Matrices U and $\bar{F}$ are post-multiplied by sum vectors 1 and the entire system sums to the single vector q showing the economy-wide US national production of commodities.

Equation (2) shows the purchase of commodities and primary inputs by n industries. Matrix w shows value added by industry with three rows of detail: 1) employee compensation, 2) gross operating surplus, and 3) taxes on production and imports less subsidies. Pre-multiplying the use and value added matrices by sum vectors 1 provides a row vector of total industry inputs X'.[2]

Equations (3) and (4) exhibit the national "make matrix" V, showing commodity production by industries on columns (equation (3)) and industry production of commodities on rows (equation (4)). Normalizing the make matrix on columns yields equation (5), the "market shares matrix" D, showing the share of a given commodity's production accounted for by each of the n industries.

[2]For clarity, we note row sum vectors by parentheses, e.g., (1). The single accent on the total gross input vector X' denotes the transpose of the column vector X.

3. Ridding Use and Final Demand Arrays of Foreign Imports

Inspection of equation (1) shows that while q conveys US commodity production, use matrix U and final demand matrix $\bar{F}$ include both foreign-produced and domestically-produced commodities. Foreign-produced commodities are given by $\bar{M}$ (foreign imports), which is subtracted, leaving domestically-produced commodities q.

The productions functions (fixed coefficient) we seek for regional modeling indicate the technical requirement for inputs to production. Our regionalizing process reduces this technical requirement to the portion obtained from local sources. Input requirements not obtained locally are imported, either from other US sources (i.e., other counties) or from foreign suppliers. Consistency with the national model requires that in the aggregate (i.e., summing all intra and inter-county trade) leaves a residual of unmet requirement just equal to foreign imports.

One way to achieve this result is to rid the national model of foreign imports — assuming, in effect, that all counties obtain the same proportion of total requirements from foreign sources. This expedient approach to the problem of foreign imports is the one adopted by EMSI and described in this section. Note that anything more sophisticated would effectively require a two-tier data reduction process, whereby foreign imports by county are estimated first, followed by a traditional intra-national regionalizing process. We are not aware of a literature pointing the way to the first tier process just described, i.e., to the foreign imports component. Our approach therefore is to adopt the expedient and develop our US Base Model rid of foreign imports.

Using Hadamard division, the foreign imports ridding process starts with formation of the scaling vector Φ as follows:[3]

$$\underset{m\times 1}{\Phi} = \left\{ \frac{q_i - e_i}{q_i - e_i + m_i} \right\} \tag{6}$$

Premultiplying both sides of equation (1) by scaling vector Φ (diagonalized[4]) provides equation (7), an approximation of the US production and absorption of commodities account free of foreign imports.[5]

$$\hat{\Phi}U1 + \hat{\Phi}\bar{F}1 + \bar{E} = q \tag{7}$$

Meanwhile foreign imports appear in equation (8) as a partitioned matrix, showing the foreign import of m commodities by n industries and f final demand sectors.

$$\underbrace{\left\{I - \hat{\Phi}\right\}}_{(m\times m)} \left\{ \underbrace{U}_{(m\times n)} \mid \underbrace{\bar{F}}_{(m\times f)} \right\} = \left\{ \underbrace{M}_{(m\times n)} \mid \underbrace{M_f}_{(m\times f)} \right\} \tag{8}$$

[3]The approach to ridding the national model of foreign imports presented in this section follows closely the presentation offered by Lahr (2001).

[4]The hat, ˆ,over an array denotes a vector that has been "diagonalized," i.e., a square matrix showing vector elements on the principal diagonal and zeros elsewhere.

[5]The procedure assumes all industries and final demand sectors import the same proportion of a given commodity and thereby only provides an approximation.

4. The Industry-by-Industry Sales Account

To this point our presentation conveys a commodity-by-industry structure: commodity absorption appears across rows, industry purchases appear down columns. Our purpose in this section is to reconfigure the accounts from a commodity-by-industry to an industry-by-industry structure.[6]

Premultiplying the absorption of US-made commodities equation (7) by the market shares matrix (equation (5)) converts the former from a commodity-by-industry to an industry-by-industry structure. The conversion appears as follows, yielding a set of *industry sales accounts*:

$$\underset{(n\times n)}{Z_n} 1 + \underset{(n\times f)}{F} 1 + \underset{(n\times 1)}{E} = \underset{(n\times 1)}{X} \tag{9}$$

Where:

$$Z_n = \underset{(n\times m)}{D} \underset{(m\times m)}{\hat{\Phi}} \underset{(m\times n)}{U} \tag{10}$$

$$F = \underset{(n\times m)}{D} \underset{(m\times m)}{\hat{\Phi}} \underset{(m\times f)}{\bar{F}} \tag{11}$$

$$E = \underset{(n\times m)}{D} \underset{(m\times 1)}{\bar{E}} \tag{12}$$

[6] At the regional level, data with specifically commodity detail are not generally available.

$$X = \underset{(n\times m)}{D}\underset{(m\times 1)}{q} \tag{13}$$

Equation (2) above shows the purchase of commodities (domestically produced and imported) and primary input purchases by each of the n industries. The imports ridden, industry-by-industry counterpart shows purchases from other domestic industries (interindustry purchases), plus purchases of primary inputs (as before), plus foreign import purchases.

$$(\ 1\ )\,Z_n + (\ 1\ )\,w + (\ 1\ )\,\{M_n\} = X' \tag{14}$$

By way of proof, equation (14) is easily derived from equation (2) in three steps. First note that:

$$(\ 1\ )\,Z_n = (\ 1\ )\,D\left\{\hat{\Phi}\right\}U = (\ \Phi\ )\,U \tag{15}$$

Next note that:

$$(\ 1\ )\,\{M_n\} = (\ 1\ )\left\{I - \hat{\Phi}\right\}U = (\ 1\ )\,U - (\ \Phi\ )\,U \tag{16}$$

Finally, substitute (15) and (16) into (14) and simplify to obtain equation (2).

5. Relocating Subsidies

As described above with reference to BEA source arrays and equation (2), matrix w conveys value added by industry and exhibits three rows of detail: 1) employee compensation, 2) gross operating surplus, and 3) taxes on production and imports less subsidies. Note that this last item combines a positive and a negative: taxes on production and imports (positive) and subsidies (subtracted off or otherwise entered as a negative). We have data on subsidies (see: Crapuchettes et al., 2012) and with an eye to the extended I-O accounts (Social Accounting Matrix) described in Crapuchettes et al., 2012, we find it useful to breakout subsidies and carry it in its own account.

Equation (17) introduces w^*, a gross income measure that includes employee compensation, gross operating surplus, and taxes on production and imports. Later, in our extended (or SAM) accounts, and following standard practice, we subtract subsidies from w^* when measuring regional value added (called gross regional product, or GRP).

$$\underset{(1\times 3)}{(\;1\;)}\,\underset{(3\times n)}{w^*} = \underset{(1\times 3)}{(\;1\;)}\,\underset{(3\times n)}{w} + \underset{(1\times n)}{(\;s\;)} \tag{17}$$

Equation (18) introduces X_n, a revised measure of industry total gross output. The revision is straightforward. X_n is simply our original industry total gross output, plus subsidies. X_n thereby indicates the total monies available to industries to discharge their expenses, including monies obtained through subsidies.

$$\underset{(n\times 1)}{X_n} = \underset{(n\times 1)}{X} + \underset{(n\times 1)}{s'} \tag{18}$$

With subsidies included in total gross output X_n, equation (19) presents the revised interindustry purchases, primary inputs, and foreign imports account:

$$(\; 1 \;) \, Z_n + (\; 1 \;) \, w^* + (\; 1 \;) \, \{M_n\} = \underset{(1\times n)}{X_n}{}' \qquad (19)$$

Likewise, the revised industry sales account (i.e., revision of equation (9)) appears as equation (20).

$$Z_n 1 + F1 + E + S = \underset{(n\times 1)}{X_n} \qquad (20)$$

6. Expanding the Accounts to NAICS 6-digit

6.1 Overview

The BEA US I-O Model conveys approximately 400 industries of detail that we have designated by n. Meanwhile EMSI data conveys detail on regional (i.e., county-level) jobs and earnings at roughly the NAICS 6-digit level (approximately 1,000 industries), which we designate immediately below by z. Our final reconfiguration of the BEA model expands its accounts and other components from n sectors to z sectors.

We will show that our process amounts to repeating a given n-industry-based production function for each of the NAICS 6-digit, z-based industries contained within the more aggregated n-based industry groups. For example, the n-based industry "Lime and Gypsum Product Manufacturing (BEA 32740)" contains two NAICS 6-digit, z-based sectors: "Lime Manufacturing (NAICS 327410)" and "Gypsum Product Manufacturing (NAICS 327420)." Our expansion of the national model described formally below provides separate NAICS 6-digit sectors, thus one for lime and another for gypsum. Both however employ the same n-level BEA 32740 production function. We present the simple mathematics of the expansion process in this section (section 6). In the following section (section 7) we form the expanded z-level national I-O coefficients matrix, (i.e., the matrix of national model production functions), demonstrate the manner in which our process amounts to a repeat of n-level production functions, and discuss the implication of our process in I-O applications.

6.2 The Sector Expansion Process

Our expansion process starts with formation of a z-row, n-column matrix P, which we will refer to as the "expansion matrix." Columns of P refer to n-based industries, rows to z-based industries. Elements of a given column, designating a particular n-based industry, exhibit fractions in the rows, designating z-level industries included in the n-level industry, zeros elsewhere. The individual fractions indicate the portion of an n-level industry's transactions that are distributed to its z-level member industries. By its nature, matrix P columns sum to unity:

$$\underset{(1\times z)}{(1)}\underset{(z\times n)}{P} = (1) \tag{21}$$

There appears to be no one optimal way to form the fractional elements of P, and a simple approach based on NAICS 6-digit earnings appears as good as any. Accordingly, the elements of P are formed based on shares of overall industry earnings. Suppose 25% of a given n-based industry's is total earnings appear in a particular z-based industry. Then the element of P for that particular n-z industry combination will indicate 25%, and our procedure will allocate that portion of n-based industry sales to that z-based industry. The approach thus assumes a constant earnings/output ratio among the z-based industries contained in a particular n-based industry.

Let us proceed to the expansion mathematics. We start by premultiplying the industry sales account of equation (20) by expansion matrix P:

$$PZ_n1 + PF1 + PE + PS = PX_n \tag{22}$$

We next define the expanded z x z matrix of interindustry sales thus:

$$PZ_nP' = \underset{(z\times z)}{Z_{zz}} \tag{23}$$

Finally, matrix (23) is substituted into equation (22) and other terms redefined as follows:

$$Z_{zz}1 + F_z1 + E_z + S_z = \underset{(z\times 1)}{X_z} \tag{24}$$

Equation (24) provides our z-based *industry sales account.*

Let us turn next to industry purchases. Equation (17) introduced our n-based matrix of industry value added w^*. Postmultiplying this term by expansion matrix P' converts it from n industries to z industries thus:

$$\underset{(3\times n)}{w^*}\underset{(n\times z)}{P'} = \underset{(3\times z)}{w_z^*} \tag{25}$$

A similar mapping converts the n-based vector of foreign imports to a z-based counterpart:

$$\underset{(1\times n)}{(1)}\, M_nP' = \underset{(1xz)}{M_z} \tag{26}$$

Finally, the expanded matrix of interindustry sales (equation (23)) is combined with the expanded array of value added (equation (25)) and foreign imports (equation (26)) to provide our expanded *interindustry purchases, primary inputs, and foreign imports account*:

$$(1)\, Z_{zz} + (1)\, w^* + M_z = X_z' \tag{27}$$

Together, the expanded sales account (equation (24)) and the expanded interindustry purchases, primary inputs, and foreign imports account (equation (27)) provide the "dog-leg" form of the z-based industry-by-industry national base model we use for building county and multi-county regional input-output models.

7. Use of the Expanded Model in I-O Applications

7.1 N-level and Z-Level Production Functions

Formation of an expanded national base model is a step EMSI must take if its z-based regional data set (in EMSI Analyst, for example) is to sync with its regional input-output model. However, users take note: while the additional sectoral detail might seem an immediate advantage, the expansion process imposes some rather strict conditions on applied modeling outcomes. Understanding these conditions can be important to a correct interpretation of EMSI model results.

The essential character and assumptions underlying the EMSI expanded model are best expressed in terms of associated I-O model fixed-coefficient production functions. Let us start by expressing the n unique production functions available from the BEA US input-output model:

$$A_{nn} = Z_{nn} \left\{ \hat{X}_n \right\}^{-1} \tag{28}$$

The corresponding z production functions conveyed by our expanded model are expressed by:

$$A_{zz} = Z_{zz} \left\{ \hat{X}_z \right\}^{-1} \tag{29}$$

The assumptions underlying the expanded model are disclosed by expressing z-based production functions A_{zz} in terms of the BEA's original n-based production functions A_{nn}. Our definition of A_{zz} in terms of A_{nn} starts with the re-expression of equation (23) thus:

$$A_{zz} = PA_{nn}\hat{X}_nP' \left\{\hat{X}_z\right\}^{-1} \tag{30}$$

The next step is pivotal. The matrix S appearing in equation (31) is a standard aggregating matrix, well-known from literature pertaining to aggregation.[7] Rows of S refer to z-based industries, columns to n-based industries. A given column, representing a particular n-based industry, exhibits a 1 in rows representing z-based industries contained in the given n-based industry, zeros otherwise. Proof of (31) is left up to the reader.[8]

$$\underset{(n\times z)}{S'} = \hat{X}_nP' \left\{\hat{X}_z\right\}^{-1} \tag{31}$$

Substituting equation (31) into equation (30) yields equation (32) and completes our immediate task, showing expanded matrix A_{zz} in terms of original matrix A_{nn}.

$$A_{zz} = PA_{nn}S' \tag{32}$$

7.2 A Simple Two and Three Sector Example

Our discussion to this point has been conveyed in general terms. At this point the presentation is facilitated by considering the simplest example. Consider a model that starts with just two industries (i.e., n = 2) and proceed to expand one of these to obtain a model with three industries (i.e., z = 3).

Designating our two original industries 'a' and 'c,' our n-sector vector of total sales appears

[7] See for example Morimoto (1970).

[8] While a general proof is awkward, a particular proof using a specific example is easy. The reader might use the example provided by the n=2, z=3 case examined in the next section.

as follows:

$$X_n = \begin{bmatrix} X_a \\ X_c \end{bmatrix} \tag{33}$$

Let us next define the parallel vector for total sales in our z-sector model:

$$X_z = \begin{bmatrix} X_a \\ \tau_1 X_c \\ \tau_2 X_c \end{bmatrix} \tag{34}$$

Where the elements τ_1 and τ_2 are scalars that sum to unity:

$$\tau_1 + \tau_2 = 1 \tag{35}$$

The expansion matrix P for this simple two-expanded-to-three industry system appears as:

$$P = \left\{ \begin{matrix} 1 & 0 \\ 0 & \tau_1 \\ 0 & \tau_2 \end{matrix} \right\} \tag{36}$$

while the associated aggregation matrix S appears as:

$$S = \left\{ \begin{matrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{matrix} \right\} \tag{37}$$

We are now in a position to compare expanded and original I-O coefficients. Given our chosen industry subscripts, the original I-O coefficients matrix A_{nn} is given by:

$$A_{nn} = \left\{ \begin{array}{cc} a_{aa} & a_{ac} \\ a_{ca} & a_{cc} \end{array} \right\} \tag{38}$$

and applying the mathematics of equation (32) yields the expanded matrix in terms of the original matrix thus:

$$A_{zz} = \left\{ \begin{array}{ccc} a_{aa} & a_{ac} & a_{ac} \\ \tau_1 a_{ca} & \tau_1 a_{cc} & \tau_1 a_{cc} \\ \tau_2 a_{ca} & \tau_2 a_{cc} & \tau_2 a_{cc} \end{array} \right\} \tag{39}$$

7.3 Two Assumptions, One Theorem

We are now in a position to identify two key assumptions and one important modeling feature that is most clearly stated as a theorem. First some language. We designate our original industries 'a' and 'c,' where industry 'c' can be thought of as the "combined industry." Our expansion process splits combined industry 'c' into two component industries, which we will designate as "industry 1" and "industry 2." Our expanded (z-based) system thus appears with three industries: "industry a," "industry 1," and "industry 2."

Assumption 1: component industries (like industries 1 and 2) that result from the expansion of a given n-based industry (industry c in the present case) exhibit the same production function. Columns 2 and 3 of expanded matrix (39) illustrate this assumption.

Assumption 2: The requirement for inputs from an originally combined industry (industry c for example) will be divided among member z-based industries according to their proportion of the original combined industry. This assumption is illustrated by elements in the lower two rows of (39), where requirements for the output of industry c are supplied τ_1 from industry 1 and τ_2 from industry 2 (and recall $\tau_1 + \tau_2 = 1.0$).

Theorem: Where information on a modeled activity is available with z-level detail, use of expanded matrix (39) entails no greater error than use of the original n-based matrix. Demonstrating this important feature requires introduction of the unknown "true" z-based matrix of I-O coefficients which we will designate as:

$$A^*_{zz} = \left\{ \begin{array}{ccc} a_{aa} & a_{a1} & a_{a2} \\ a_{1a} & a_{11} & a_{12} \\ a_{2a} & a_{21} & a_{22} \end{array} \right\} \tag{40}$$

Elements of n-based coefficients matrix (38) are then defined (implicitly) in terms of the unknown disaggregated matrix as shown in equations (41), (42) and (43):

$$a_{ac} = \frac{a_{a1}X_1 + a_{a2}X_2}{X_1 + X_2} \tag{41}$$

$$a_{cc} = \frac{(a_{11} + a_{21})\,X_1 + (a_{12} + a_{22})\,X_2}{X_1 + X_2} \tag{42}$$

$$a_{ca} = a_{1a} + a_{2a} \tag{43}$$

where X_1 and X_2 are the unknown "true" total sales for disaggregated industries 1 and 2, and $X_1 + X_2 = X_c$.

Now imagine a modeling application where the initial effect (i.e., the effect being modeled) reflects z-level detail, for example a change in the output of a given NAICS 6-digit industry. For our theoretical consideration, let us imagine an initial change equal to the total sales of industry 2, given by X_2. Enabling a direct side-by-side comparison, let us consider the direct output response of our original industries a and c, first using our "true" input-output coefficients matrix (equation (40)), next using the original n-based coefficients (equation (38)), and finally using our expanded z-based coefficients (equation (39)). The results of our three applications are as follows:

True Response

$$\begin{aligned} \Delta_a &= a_{a2} X_2 \\ \Delta_c &= (a_{12} + a_{22}) X_2 \end{aligned} \tag{44}$$

Response Using Original n-based Model

$$\begin{aligned} \Delta_a &= a_{ac} X_2 \\ \Delta_c &= a_{cc} X_2 \end{aligned} \tag{45}$$

Response Using Expanded z-based Model

$$\begin{aligned} \Delta_a &= a_{ac} X_2 \\ \Delta_c &= (\tau_1 a_{cc} + \tau_2 a_{cc}) X_2 = a_{cc} X_2 \end{aligned} \tag{46}$$

We can summarize our finding in two parts. First, where information on the initial effect exhibits finer detail than the unexpanded (or n-based) model, neither the original n-based model nor the expanded z-based model capture the true effects, i.e., the effects indicated by the unknown "true" model. Second, use of the expanded z-based model entails no greater error than use of the original n-based model: the two models produce mathematically identical results.

8. Expansion Process Discussion

EMSI Analyst conveys roughly the 1,000 sectors of the NAICS 6-digit industrial classification system and for EMSI's I-O model to sync it must convey these same 1,000 sectors. The US model on which EMSI regional models are based convey fewer sectors, approximately 400. EMSI expands the national model's 400 sectors to the EMSI Analyst 1,000 sectors following the procedure formally described in section 6 above. Section 7 disentangles the expansion process to disclose its implications, which are then summarized in two assumptions and one theorem. Let us now consider these implications in less formal language.

8.1 Using a Single Production Function for Multiple Sectors

The simplest exercise in applied I-O impact analysis starts with a known change in some regional industry, called the "initial effect," and proceeds to estimate related changes to all the other regional industries, called "direct," "indirect," and "induced" effects. The initial change will normally be known at the NAICS 6-digit level and this makes application using the EMSI I-O model easy — the user simply enters the initial change and the model does the rest, applying the indicated production function and making all the various input-output calculations. How does this differ from the user of a non-expanded model?

Using a non-expanded model, the first step is to consult the "NAICS to BEA I-O Model Bridge" and locate the National I-O Model sector containing the initial impact (NAICS 6-digit) industry. Imagine, for example, being asked to model the impact of a new 100 job "Nitrogenous Fertilizer Manufacturing" plant (NAICS 325311). The user finds this NAICS 6-digit industry contained in National I-O Model sector BEA 32531, an aggregate of three

NAICS 6-digit sectors: "Nitrogenous Fertilizer Manufacturing" (NAICS 325311), "Phosphatic Fertilizer Manufacturing" (NAICS 325312), and "Fertilizer (mixing only) Manufacturing" (NAICS 325314). The user of the non-expanded model proceeds to enter the 100 job change in nitrogenous fertilizer employment (NAICS 325311) into the National I-O Model's aggregate fertilizer manufacturing sector (BEA 32531), thus invoking the aggregate sector production function, and completing all the usual I-O calculations based on this aggregate function. Short of building a custom survey-based nitrogenous fertilizer production function, there is no other choice.

Now consider the same exercise using the EMSI I-O model. The EMSI I-O model is constructed on a NAICS 6-digit platform so the user simply enters the 100 jobs in the model's Nitrogenous Fertilizer Manufacturing sector (NAICS 325311). In computing the associated impacts, the EMSI model effectively borrows from the US model and employs the production function that includes NAICS 325311, which is National I-O Model sector BEA 32531. The net result is the same as using the non-aggregated model. This is a non-algebraic statement of "assumption #1" above.

8.2 Distributing Impacts Across Expanded Sectors

The EMSI I-O model uses the same production functions as the traditional non-expanded model and will therefore indicate the same impacts. If the traditional non-expanded model indicates a 10-job impact in wholesale trade (BEA 42000), then the EMSI model will likewise indicate 10 jobs in wholesale trade. The difference is that the EMSI model will show these

10 jobs spread across the several NAICS 6-digit wholesale trade sectors present in the region. This feature of the expansion process is conveyed formally as "assumption #2" above.

Considering a deliberately simple hypothetical, suppose the region has but two wholesalers: a lumber wholesaler (NAICS 423310) and a refrigeration wholesaler (NAICS 423740). Suppose further that the lumber wholesaler accounts for 80% of combined wholesaler earnings while the refrigeration wholesaler accounts for the remaining 20%. A 10-job wholesale trade impact indicated in a non-expanded model for the region will appear in the EMSI expanded model as 8 jobs in lumber wholesaling and 2 jobs in refrigeration wholesaling.

The distribution of impacts across expanded sectors according to their base earnings provides an admittedly crude approximation. In reality, the true impact may all occur in lumber wholesaling, or in refrigeration wholesaling, or any other combination. Note that the user of the non-expanded model will have no information on the distribution of the impacts — the model indicates 10 wholesale trade jobs, and the user is on his or her own to allocate them to lumber wholesaling, refrigeration wholesaling, or some combination of both.

Does the expanded-model's "crude approximation" of specific NAICS 6-digit impacts represent an improvement over the estimate (guesstimate?) made by the user? Unfortunately this question cannot be generally answered in the affirmative; the user may have some firsthand knowledge of the 6-digit industries in question. We can say, however, that where NAICS 6-digit data are available (as they are for users of EMSI Analyst), ready allocation of impacts to the 6-digit level of the base data is a clear convenience. And that is ultimately EMSI's reason for adopting the expansion process: to synchronize I-O impacts to NAICS 6-digit baseline data.[9]

8.3 The Expanded Model Introduces No Error Relative to the Non-Expanded Model

Let us return to the fertilizer manufacturing example discussed above. The National I-O Model for this sector contained three NAICS 6-digit sectors: "Nitrogenous Fertilizer Manufacturing," "Phosphatic Fertilizer Manufacturing," and "Fertilizer (mixing only) Manufacturing." An unknown "true" production function exists for each of these three sectors. At the same time, a user wishing to model one of the individual sectors is obliged to use the production function representing the aggregate of the three. This is true whether the user employs EMSI's expanded model or one of the I-O modeling packages based on a non-expanded model. Both approaches produce error, measured by the difference in results between the unknown true production function and the aggregated production function. The point is that the use of the expanded model produces no greater error than use of the expanded model, and this is the thrust of the "Theorem" stated above.

[9]Of course the true (but unknown) impact may occur in a wholesale trade sector non-existent in the region, perhaps electrical equipment wholesaling (NAICS 423610) or farm machinery wholesaling (NAICS 423820). In this case, the 10 job wholesale trade impact in the region is entirely erroneous, a result of the data reduction process that uses a single (i.e., aggregated) national model wholesale trade sector. Note that this traditional "aggregation error" has nothing to do with the expansion process as discussed: It afflicts expanded and non-expanded models alike.

9. Summary

The EMSI I-O model is constructed according to a data reduction method that regionalizes the US National I-O model using a modern gravity model algorithm. Prior to regionalization, a number of model reconfirmations are required. Key among these is ridding national model production functions of foreign imports, configuring matrices to an industry-by-industry basis, and relocating subsidies. The most notable reconfiguration, however, is expansion of the approximately 400 national model sectors to the approximately 1,000 NAICS 6-digit sectors of EMSI Analyst.

The expansion of sectors is a bold move needed to synchronize EMSI I-O and Analyst sectors. Importantly, however, in I-O applications results obtained from the expanded model are no different than those derived using the standard non-expanded model. Both models rely on the same 400 some-odd production functions of the national model. EMSI expanded model results do show more detail (1,000 sectors) than the non-expanded model, though this detail can be seen to reflect convenience rather than sectoral precision.

10. References

Crapuchettes, Jonathan, Hank Robison, and Deacon James (2012), EMSI Multi-Regional Social Accounting Matrix, (MR-SAM) Modeling System. Available at economicmodeling.com.

Lahr, Michael L. (2001), "Reconciling Domestication Techniques, The Notion of Re-exports and Some Comments on Regional Accounting." Economic Systems Research,13(2).

Miller, Ronald E. and Peter D. Blair. (2009). Input-Output Analysis: Foundations and Extensions. Cambridge , UK: Cambridge University Press.

Morimoto, Y. (1970). "On Aggregation Problems in Input-Output Analysis," Review of Economic Studies, 119-126.