COMPARING THE IMPACTS OF BIOFUELS USING SURVEY AND NON-SURVEY DATA

A Thesis

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

---

by

JAMES ROSSI

Dr. Judith I. Stallmann, Thesis Supervisor

DECEMBER 2010

The undersigned, appointed by the dean of the Graduate School, have examined the thesis entitled

COMPARING THE IMPACTS OF BIOFUELS

USING SURVEY AND NON-SURVEY DATA

presented by James Rossi,

a candidate for the degree of master of science,

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Judith I. Stallmann

Professor Thomas G. Johnson

Professor Deanna L. Sharpe

DEDICATION

I dedicate this thesis to my parents. To my father for his continued stressing of the importance of education and setting an example of what I could become if I applied myself. And to my mother for always being there with a sympathetic ear. I could not have done this without your support.

ACKNOWLEDGMENTS

I would like to thank Dr. Judith I. Stallmann for both her great contributions towards my academic development and for her patience through multiple drafts of this thesis. I would like to also thank Dr. Thomas G. Johnson for his timely advice and help with the technical aspects of my research. I would like to thank Dr. Deanna L. Sharpe for her participation on my thesis committee. I would like to thank Dr. Georgeanne M. Artz, who was instrumental in both composing the survey as well as resolving several data issues. Finally, I would like to thank Dave Swenson for his assistance with his modeling methodology used in this paper.

TABLE OF CONTENTS

ACKNOWLEDGMENTS……. ii

ABSTRACT……. vi

Chapter

I. INTRODUCTION……. 1

II. THE MODEL……. 6

The Transactions Table……. 6

Technical Coefficients……. 8

Final Demand Changes and Rounds of Economic Activity……. 10

The Interdependence Coefficients Matrix……. 13

Regional Purchase Coefficients……. 16

Input Coefficients……. 25

Multipliers……. 27

Accuracy in Input-Output Models and Sources of Error……. 28

Selecting a Non-Survey Method……. 34

III. LITERATURE REVIEW……. 38

Scenario Development……. 38

Testing Regiments……. 51

Summary……. 61

IV. METHODS……. 64

Preparing the Data……. 64

Scenario Development……. 69

Comparing Input Coefficients..... 71

Multiplier Testing..... 79

V. RESULTS..... 82

Scenario Results..... 82

Input Coefficient Testing Results..... 84

Wilcoxon Signed-Rank Test..... 84

Simple Regression..... 85

Measures of Distance from the Mean..... 87

Theil's Inequality Coefficient..... 89

Information Index..... 91

Multiplier Testing Results..... 93

Paired t Test..... 94

Spearman's Rank Correlation Coefficient..... 95

Mean Absolute Percentage Error..... 96

VI. CONCLUSIONS..... 98

Limitations..... 102

Implications..... 103

APPENDIX

A. THE SURVEY..... 105

B. IMPUTATIONS..... 112

REFERENCES..... 115

LIST OF ILLUSTRATIONS

Table Page

2.1 Example Input-Output Table in Dollar Values…….. 7

2.2: Example of Technical Coefficients Table……..... 10

3.1: Plant Characteristics…….....…. 48

3.2: Multipliers Reported in Previous Studies……..... 49

5.1: Scenario Employment Effects and Multipliers…….. 82

5.2: Scenario Output Effects and Multipliers…….. 83

5.3: Scenario Total Value-Added Effects and Multipliers…….. 84

5.4: Level of Significance Table under a Two-Tailed Test……. 85

5.5: Regression Results……..... 87

5.6: Critical Values of the t-Distribution for a Two-Tailed Test…….... 87

5.7: Measures of Distance from the Mean……..... 89

5.8: Theil's Inequality Coefficient, U…….…… 91

5.9: Goodness of Fit…….... 92

5.10: Composite Scenario Rankings……..... 93

5.11: Paired t Test……..... 94

5.12: Spearman's Rank Correlation Coefficient…….. 96

5.13: Mean Absolute Percentage Error……..... 96

5.14: Composite Scenario Rankings……..... 97

ABSTRACT

This paper utilizes survey data to compare several non-survey methods of modeling the economic impacts of biofuels plants. It examines differences in the input coefficients derived from the survey versus the trade coefficients generated through the non-survey methods. It finds that of the three non-survey methods examined, the Swenson (2006) scenario input coefficients most closely represent those found in the survey based on the performance of the non-survey scenario input coefficients in a variety of statistical tests. Further, it examines the economic impacts (multipliers) generated by these scenarios compared to those generated from the survey. Based upon statistical tests of the multipliers, the Swenson scenario's estimated impacts most closely represent the impacts derived from the survey.

I. INTRODUCTION

Input-output models are used to predict how an exogenous shock in one or more industries affects a regional economy. That is, it is important to recognize that the data and analysis are specific to a region. The foundations of this technique were outlined in Wassily Leontief's seminal 1936 article, "Quantitative Input and Output Relationships in the Economic Systems of the United States." Leontief sums it up thusly, "Essentially it is a method of analysis that takes advantage of the relatively stable pattern of the flow of goods and services among the elements of our economy to bring a much more detailed statistical picture of the system into the range of manipulation by economic theory" (Leontief, 1986: 4).

Oftentimes, a regional scientist is asked to predict how a particular shock will affect an area. The methods of performing such prognostications range from costly and time-consuming surveys to quicker, packaged datasets. The quicker, out-of-the-box, non-survey datasets provide the distinct advantage that the analyst can give a more timely answer than is possible when a survey is used. At the same time, the survey is likely to give a more accurate answer because is it based on the specific expenditure patterns of the firms in the region of interest. This is particularly true when an industry is undergoing technological change. Thus, the analyst often must make a tradeoff between time and cost of survey data compared with packaged datasets and accuracy.

This paper aims to examine this tradeoff and how this impacts the projections made by analysts.

One software program with accompanying dataset widely used by economists and regional scientists is Minnesota IMPLAN Group's, Incorporated's Input-Output Model for Planning (IMPLAN). IMPLAN originated through a cooperative effort by the United States Department of Agriculture Forest Service, the Federal Emergency Management Agency, and the United States Department of the Interior Bureau of Land Management with the goal of easing the Forest Service's task of land and resource management (Minnesota IMPLAN Group, 2000).

Input-output models all essentially follow a similar framework. They can be described as a mathematical approach to analyze the relationships among various sectors of the economy. A firm that produces a given product purchases inputs from a variety of firms engaged in other sectors, while selling its output as either an intermediate good to other firms or as final goods to consumers. This buying and selling is represented by a matrix in which the column contains the sectors from which the firm buys and the row contains the sectors to which the firm sells. It becomes a matrix because sales by Firm A to Firm B are purchases by Firm B from Firm A.

The accuracy of the results derived from these models is contingent on the accuracy of the data used in their construction. This paper compares the purchases (column) of firms based on survey data with the purchases reported in packaged data

disseminated with input-output software. Further, it tests a non-survey technique of modeling biofuels production derived by Swenson (2006) to see how this technique compares to the other non-survey approaches. Inaccuracy of entries in input-output tables can lead to misinformation for users.

The questions this paper addresses are the following:

- How representative are regional coefficients derived from the national level of actual regional survey data? Does there exist significant statistical variation between the two?
    - Is the rapid technological change being experienced in the biofuels industry leading to changes in the regional coefficients that are not captured by packaged data because of the time-lag for release?
- If there are differences between the input coefficients of the survey and the pre-packaged data, how does this impact the multipliers derived from their application?
- Of the non-survey techniques, which most closely matches the survey data?

This paper will compare survey data from recent Missouri biofuels plants with the input coefficients packaged with the IMPLAN software and the input coefficients using the Swenson (2006) method. This paper will test whether the two sets of coefficients differ using a variety of statistical tests. The paper will also compare the multipliers generated by the survey-based coefficients and the non-survey coefficients.

The potential for inaccuracy of the coefficients in an input-output model has long been known. Unfortunately for packaged datasets, there usually is a substantial lag between when the data are collected, organized and finally constructed into a dataset (Matuszewski, Pitts, and Sawyer, 1964). Further, for any relatively new industry, or an industry undergoing technological change, there is the risk that the time lag between when the data were collected and compiled and when the packaged data are released to the public results in the use of information that is not an accurate reflection of the current firms.

Corn-based ethanol and soy-based biodiesel (both hereafter called biofuels) plants were selected as this industry is rife with technological change, perhaps most easily demonstrated by the dramatic change in plant capacities from 10 million gallon a year (MGY) facilities 15 years ago to 50 MGY facilities 5 years ago to the 100 MGY facilities of today (Johnson, 1995; Swenson, 2006; Swenson, 2008). In addition, some expect to see significant expansion in the coming years. The expansion of the industry could arise from a number of factors: rising petroleum prices, subsidy and use requirements of the Energy Policy Act of 2005, and their attractiveness as a source of value-added agriculture as a rural development policy (De La Torre Ugarte, English, and Jensen, 2007).

This paper contributes to the literature in that it compares several scenario development techniques with actual survey data to determine the accuracy of both the

input coefficients and the multipliers produced by these different scenario development methodologies utilizing several statistical tests. It draws attention to which elements of a forecasted impact are most likely to be misestimated by a researcher and thus, most likely to convey misinformation.

II. THE MODEL

All economies studied via input-output models start with a geographic basis. This geographic partitioning can be done at a national level, state level, county level, municipality level, or any combination thereof. The inputs and outputs of firms in various sectors appear in the transactions table.

The Transactions Table

The first part of an input-output model is a national or regional input-output or transactions table. A simplified example of this is presented below. Each sector within an economy purchases a mix of raw inputs, intermediate goods, and labor in order to produce its output. What the firm purchases to make its output, its production function, is reflected in the columns of the input-output table. Its output can be sold within the sector, sold to other sectors, sold to final consumers, or sold to firms and consumers outside of the region as exports.

For simplicity, this fictional economy has been broken into three producing sectors: an agricultural sector, a manufacturing sector, and a services sector and one extra sector: a final demand sector. The column final demand represents the value of goods and services purchased by households (after savings are deducted), the government and exports. The row for final payments represents payments to households for wages and proprietor income, profits, and payments made for imports

(that is goods and services either not produced in the region or in instances where regional demand exceeds available regional supply). Outputs and inputs for each sector must be equal for accounting reasons (Leontief, 1986; Jones, 1997).

Table 2.1: Example Input-Output Table in Dollar Values

To / From	Sector 1 Agriculture	Sector 2 Manufacturing	Sector 3 Services	Sector 4 Final Demand	Total Output
Sector 1 Agriculture	420	640	250	640	1950
Sector 2 Manufacturing	160	310	590	360	1420
Sector 3 Services	730	110	350	890	2080
Sector 4 Final Payments	640	360	890	250	2140
Total Inputs	1950	1420	2080	2140	

While the transactions table is a helpful tool for analyzing the structure of the economy, its usefulness for analysis and forecasting is less (Jones, 1997). If we are using I-O to see what happens with shocks, we are using it as a forecasting model. It should not be used to forecast when the shock is large, because then technical coefficients are not likely to be stable and the structure of the economy can change—sectors can enter or leave. The transactions table can be used to construct technical coefficients, which represent the production function of an industry, that is the inputs it buys in order to produce its outputs.. A technical coefficient can be thought of as the portion of total inputs required to be purchased from sector j for a firm in sector i to produce one dollar's worth of output.

The transaction table represents the production function of the firm, but input-output requires three key assumptions about what is produced and the production function itself, which are the following:

1) Each sector produces only one commodity, where commodity is defined as the chief output of the firm. All commodities produced are homogenous in nature.
2) Each sector has a fixed ratio of inputs and outputs.
3) Each sector operates under strictly constant returns to scale in production.

Technical Coefficients

Technical coefficients are derived thusly. The economy can be divided into n+1 sectors, n producing sectors and n+1 final demand sectors. Let x_i be the value of physical output of sector i. Further, let x_{ij} be the value of the output of sector i used as an input by sector j. Thus, the technical coefficients for each sector can be written as a_{ij} and are calculated as follows:

$$a_{ij} = \frac{x_{ij}}{x_i} \tag{2.1}$$

The technical coefficients are the percentage or the cents of every dollar of inputs that firms in sector i buy from sector j. Our example economy's technical coefficients can

then be derived. The calculated technical coefficients for the agriculture sector are as follows:

Agriculture,	$a_{11} = 420/1950 = 0.215$
Manufacturing	$a_{21} = 160/1950 = 0.082$
Services	$a_{31} = 730/1950 = 0.374$
Final Payments	$a_{41} = 640/1950 = 0.328$

These coefficients represent the value of purchased inputs from each sector in the economy that must be used by the agricultural sector to produce one dollar of output. Note that there are intra-sector purchases. For example, in the agricultural sector a dairy likely buys heifers. These technical coefficients are the direct or first round effects of a one dollar increase in the output of the agricultural sector. The technical coefficients of each sector are represented below in table 2.2. It should be noted that transactions and coefficients are in producer prices, rather than purchaser prices. As such, when putting these values into purchaser prices, a portion of these purchases will go to margin sectors such as retail and wholesale. However, this does not change during the calculation of coefficients nor does it change either total expenditures or receipts. In this simplified model, these margin sectors do not exist and as such, it assumes these sectors trade directly with each other.

Table 2.2: Example of Technical Coefficients Table

	Sector 1: Agriculture	Sector 2: Manufacturing	Sector 3: Services	Sector 4: Final Demand
Sector 1: Agriculture	.215	.451	.120	.299
Sector 2: Manufacturing	.082	.218	.284	.168
Sector 3: Services	.374	.077	.168	.416
Sector 4: Final Payments	.328	.254	.428	.117

Final Demand Changes and Rounds of Economic Activity

Changes in an economy are not driven by changes in output per se, but instead by changes in final demand[1]. To measure the effects of a one dollar change in final demand for a sector we multiply the one dollar change by the technical coefficients of the sector. Thus, in the first round of economic linkages a one dollar increase in final demand for the agricultural sector increases agricultural output by $1.215 (*1 x 0.215* plus the one dollar change in final demand). Manufacturing increases by $0.082 and services increases by $0.374. Thus, the first round effect of a one dollar increase in final demand for agricultural products is $1.671 (*1.215+0.82+0.374*).

[1] To avoid any potential confusion between final demand, direct effects, and first round effects let the following terms be defined as such: 1) final demand is the demand for goods and services by external units (such as households, government, and exports) which are not used as an intermediate input in a production process; 2) direct effects are the manifestation of this change in final demand in the sector experiencing an increased demand for its product; and 3) first round effects are the increased output of those sectors (and not their own subsequent input demands needed to meet this new output requirement) which supply the firm experiencing the direct change in final demand (Miller and Blair, 1985).

This change can be traced through subsequent rounds. These subsequent rounds must happen because in order to meet the output increase in the first round, each of these sectors need to increase output in the second round, the third round, etc. In this context, a round is defined as a sequence of economic activity. For example, when firm i buys from firm j then firm j must increase its purchases. Thus, firm j buys more from firm k, which in turn must increase its purchases. These are called rounds of economic activity and lead to the multiplier effect. These effects will continue infinitely with increasingly smaller output effects because of the leakages. Leakages are the amount of purchases by firms in this regional economy that are imported. This can be shown by examining the first round impacts from the one dollar increase in final demand for agriculture. Multiply the effects from the first round (.215, .082, and .374) by their own respective column elements from table 1.2 and summing their products to find the second round impacts.

First round impact		Column Coefficients (from table 2.2)		
$.215	*x*	*(.215 + .082 + .374)*	=	*.145*
$.082	*x*	*(.451 + .218 + .077)*	=	*.061*
$.374	*x*	*(.120 + .284 + .168)*	=	*.214*
Total second round output increase			=	*.420*

Through two rounds the sum of the total regional output to meet the original one dollar of final demand increase in the agricultural sector is $2.091 (*1.671 + .420*). This process will continue *ad infinitum* (Jones, 1997). Eventually, the effects of additional rounds will

approach zero. This decrease is due to a phenomenon known as leakages. Leakages are money that leaves the system. Money leaves the system through two causes: 1) some money is deducted by households to be put into savings for future periods and 2) some money is lost due to imports of goods produced outside of the region.

The effects of rounds of economic activity can be divided into several categories: direct effects, indirect effects, and induced effects. Direct effects are those that take place only within the industry that is directly shocked. For example, if there is an exogenous purchase of $1,000,000 of milk by the government, the dairy industry has a direct effect of an increase in $1,000,000. The indirect effects are the inter-industry transactions necessary for this increased output. In this example, the dairy must buy corn, veterinary services, etc. to meet this increased output. Finally, there are the induced effects, which measure the changes in production due to increases in household incomes. Continuing with the dairy example, some of the sector's expenses are for labor and returns to owners of capital. When output increases these households gain additional income of which they save some and spend the rest. The induced effect measures the extra spending in local restaurants, stores, etc. Fortunately, Leontief demonstrated how the solution to the input-output system could be determined analytically, which generates the interdependence (or Leontief) coefficients matrix (Jones, 1997).

The Interdependence Coefficients Matrix[2]

The interdependence coefficients matrix measures the total (direct and indirect) output required by all producing sectors to meet any individual sector's one dollar change in final demand. Continuing our example, we write the elements of the producing sectors in table 2.1 as a set of simultaneous equations:

$$x_{11} + x_{12} + x_{13} + Y_1 = X_1$$

$$x_{21} + x_{22} + x_{23} + Y_2 = X_2 \qquad (2.2)$$

$$x_{31} + x_{32} + x_{33} + Y_3 = X_3$$

where: x_{ij} = sales from sector i (rows) to sector j (columns)

Y_i = sales from sector i to final demand

X_i = total output of sector i

If we rearrange equation 2.1, we get $x_{ij} = (a_{ij})\ (X_i)$. Thus, the amount of sales from sector i to sector j depends on the amount of output in sector j and the technical coefficient of what sector j buys from sector i. Further, if we substitute this equation into equation 2.2., we can rewrite our three sector economy model as:

$$a_{11}X_1 + a_{12}X_2 + a_{13}X_3 + Y_1 = X_1$$

$$a_{21}X_1 + a_{22}X_2 + a_{23}X_3 + Y_2 = X_2 \qquad (2.3a)$$

$$a_{31}X_1 + a_{32}X_2 + a_{33}X_3 + Y_3 = X_3$$

Now the interdependence of the individual sectors is beginning to take shape. The output in any sector is dependent upon the output of other sectors, the input

[2] This section follows the notation used by Jones (1997).

requirements of every sector, and the amount of final demand (Y_i). Solving for the exogenous final demand, 2.3a must be re-written again:

$$
\begin{aligned}
&X_1 - a_{11}X_1 - a_{12}X_2 - a_{13}X_3 = Y_1 \\
&- a_{21}X_2 + X_2 - a_{22}X_2 - a_{23}X_3 = Y_2 \\
&- a_{31}X_1 - a_{32}X_2 + X_3 - a_{33}X_3 = Y_3
\end{aligned}
\tag{2.3b}
$$

or,

$$
\begin{aligned}
&(1 - a_{11})\, X_1 - a_{12}X_2 - a_{13}X_3 = Y_1 \\
&-a_{21}X_1 + (1 - a_{22})\, X_2 - a_{23}X_3 = Y_2 \\
&-a_{31}X_1 - a_{32}X_2 + (1 - a_{33})\, X_3 = Y_3
\end{aligned}
\tag{2.4}
$$

We can further simplify through the use of matrix notation[3]:

$$
\begin{bmatrix} (\mathbf{1} - \mathbf{a_{11}}) & -\mathbf{a_{12}} & -\mathbf{a_{13}} \\ -\mathbf{a_{21}} & (\mathbf{1} - \mathbf{a_{22}}) & -\mathbf{a_{23}} \\ -\mathbf{a_{31}} & -\mathbf{a_{32}} & (\mathbf{1} - \mathbf{a_{33}}) \end{bmatrix} \times \begin{bmatrix} \mathbf{X_1} \\ \mathbf{X_2} \\ \mathbf{X_3} \end{bmatrix} = \begin{bmatrix} \mathbf{Y_1} \\ \mathbf{Y_2} \\ \mathbf{Y_3} \end{bmatrix}
$$

or **A*X = Y**. (2.5)

The matrix **A*** is the difference between two matrices, an identity matrix (**I**) minus the matrix of technical coefficients (**A**) calculated earlier or **(I – A) = A***. Ergo, equation 2.5 can be written in matrix notation as:

$$\mathbf{(I - A)X = Y} \tag{2.6}$$

where: **X** = the matrix of each sector's output

Y = the matrix of each sector's final demand

[3] As this section frequently switches between math and matrix notation, math notation will be written in *italics* and matrix notation will be written in **BOLD CAPITALS**.

In order to derive sectoral output (X) as a function of final demand (Y) multiply each side of equation 2.6 by $\mathbf{(1 - A)^{-1}}$:

$$\mathbf{(I - A)^{-1}\,(I - A)X = (I - A)^{-1}\,Y} \qquad (2.7)$$

Thus, 2.7 becomes:

$$\mathbf{IX = (I - A)^{-1}\,Y}$$

or,

$$\mathbf{X = (1 - A)^{-1}\,Y} \qquad (2.8)$$

The $\mathbf{(I - A)^{-1}}$ matrix is the matrix of interdependence coefficients. Returning to table 2.2, matrix **A** is:

$$\mathbf{A} = \begin{bmatrix} .215 & .451 & .120 \\ .082 & .218 & .284 \\ .374 & .077 & .168 \end{bmatrix}$$

Therefore, **(I – A)** becomes

$$(I - A) = \begin{bmatrix} .785 & -.451 & -.120 \\ -.082 & .782 & -.284 \\ -.374 & -.077 & .832 \end{bmatrix}$$

Now we invert this matrix to find $\mathbf{(I - A)^{-1}}$. First, we note that the determinant of **(I –A)** is 0.378. To do this first we must find the cofactor matrix **C**:

$$C = \begin{bmatrix} .629 & .174 & .299 \\ .384 & .608 & .230 \\ .222 & .232 & .576 \end{bmatrix}$$

If we transpose **C** to form **C'** and multiply it by the inverse of the determinant of **(I – A)**, we find $\mathbf{(I - A)^{-1}}$:

$$(I - A)^{-1} = \begin{bmatrix} 1.660 & 1.015 & .586 \\ .461 & 1.606 & .614 \\ .790 & .607 & 1.523 \end{bmatrix}$$

It should be noted that the elements in the $\mathbf{(I - A)^{-1}}$ matrix correspond exactly with the elements of table 2.1. That is, the elements of column 1 correspond to sector 1, agriculture. The coefficients in column 1 indicate that if final demand for agriculture increases by one dollar of sales, total output requirements for agriculture are $1.66 ($1 going to final demand and 66 cents of additional indirect output that is needed by the other sectors and the agricultural sector so they can each satisfy their own input needs), $0.46 output increase in manufacturing, and $0.79 output increase in services.

Regional Purchase Coefficients

We can now proceed to a discussion of regional purchase coefficients (RPCs). As noted above the input-output table is specific to a region. Regional purchase coefficients represent the amount of demand that can be satiated by local (or regional) sources. RPCs are used to convert national coefficients into regional coefficients which reflect interregional procurement of inputs, and thus leakages from the region. The amount of leakage at the regional level for a given input likely exceeds the amount of leakage at the national level. For example, the supply of girders manufactured in a given small rural community is likely quite low, but at the national level, steel girders are readily produced by a number of firms somewhere in the nation.

The techniques for deriving these regional purchase coefficients vary amongst input-output models (Schaffer, 1999). One method of estimating RPCs is through location quotients. At their simplest, location quotients compare the local industry mix

to the national industry mix. That is, location quotients measure the concentration of a given industry in a region relative to the national concentration for that industry. The simple location quotient for sector i in region R, LQ_i^R is given by:

$$LQ_i^R = (X_i^R/X^R)/(X_i^N/X^N) \tag{2.09}$$

where: X_i^R = gross output of sector i in region R

X^R = total ouput in region R

X_i^N = gross output of sector i at the national level

X^N = total output at the national level.

The numerator is the proportion of region R's total output produced in sector i. The denominator is the proportion of national total output produced in sector i. Interpretation of LQ_i^R is as follows: if $LQ_i^R > 1$, then sector i is more concentrated in region R relative to the national level because it is a higher percentage of region R than of the nation; if $LQ_i^R < 1$, then sector i is less concentrated in region R relative to the nation; if $LQ_i^R = 1$, then sector i is equally concentrated at the regional level as it is at the national level.

The location quotients technique is then applied to the national technical coefficients,

$a_{ij}^N s$, as follows: let a_{ij}^R denote the regional technical coefficient:

If $LQ_i^R < 1$, then $a_{ij}^R = a_{ij}^R(LQ_i^R)$

If $LQ_i^R \geq 1$, then $a_{ij}^R = a_{ij}^N$.

The interpretation of this is that if $LQ_i^R < 1$, then the region cannot meet its own demands. Conversely, if $LQ_i^R \geq 1$, then it is able to meet its own internal demands and the excess is shipped to external regions as exports (Miller and Blair, 1985).

In addition to the simple location quotient technique presented above, Miller and Blair (1985) highlight two additional types of location quotient techniques that are available: purchases-only location quotients and cross-industry location quotients. The purchases-only location quotients operates on the basis that the production of a given output is only relevant to the region's demand if it is used in that region. Ergo, the production of livestock is not really relevant in determining the region's ability to meet its demand for computer components, for example. Further, this approach ignores the importance of computers to households. If household coefficients are adjusted by the RPC as they typically are, then their demand should be included in X*. The region's purchases-only location quotient, PLQ_i^R, is given by:

$$PLQ_i^R = (X_i^R / X^{*R}) / (X_i^N / X^{*N}) \tag{2.10}$$

where: X^{*R} = total regional output of only sectors that use i as an input

X^{*N} = total national output of only sectors that use i as an input.

The cross-industry location quotients technique holds the chief advantage of comparing the relative size of selling sector i and buying sector j in the regional economy. Let, CLQ_{ij}^R denote the cross-industry location quotient for region R for selling sector i and buying sector j, then:

$$CLQ_{ij}^{R} = (X_i^R / X_i^N) / (X_j^R / X_j^N) \tag{2.11}$$

The cross-industry location quotient is used as follows:

If $CLQ_{ij}^{R} < 1$, then $a_{ij}^{R} = a_{ij}^{N}\ (CLQ_i^R)$

If $CLQ_{ij}^{R} \geq 1$, then $a_{ij}^{R} = a_{ij}^{N}$.

Thus, if the location quotient of selling sector i is larger relative to the location quotient of buying sector j, then all of sector j's needs can be met by local supplies. Conversely, if the location quotient of selling sector i is smaller relative to the location quotient of buying sector j, then all of sector j's needs cannot be met by local suppliers and some must be imported from outside regions.

A major problem with location quotient techniques is its inability to allow for cross hauling. If an industry is concentrated in a region, the location quotient will be greater than one and as such the RPC will be one. This falls apart when considering a region that is a heavy producer of a product, but buys its supply of the product from a manufacturer outside of the region. Put simply, just because a brewery exists in a given county and it exports some of its beer outside the county, it is not entirely plausible that residents of that county only drink that beer and do not import beers brewed outside of the region. Further, this can understate the impact these imports and exports can have on other local industries, such as trucking and wholesale trade. In this case, the location quotient would be greater than one, but the excess demand would necessitate a RPC less than one. Also, the major assumption that the regional and national economies

both contain identical production structures is questionable (Swanson, Morse, and Westeren, 1999). It should seem quite clear that the inputs used to produce electricity are going to vary between the coal power plants of West Virginia and the Hoover Dam on the border of Arizona and Nevada.

Another method of calculating RPCs is the supply-demand pool method developed by Isard (1953). The essence of the supply-demand pool method is that it assumes that all local demand is satiated by local supply first before it begins importing, that is it assumes away cross-hauling. Likewise, all local output is used first by local demanders before it is exported to users outside of the region.

In the supply-demand pool technique, the first step is finding the national technical coefficients denoted a_{ij}^{N}. Next, regional output for each sector is found by multiplying the national technical coefficients by each sector's actual regional output and summing. Let the total regional demand for commodity i in region R be: $\tilde{X}_i^R$, then:

$$\tilde{X}_i^R = \sum_j a_{ij}^N X_j^R + \sum_f c_{if}^N Y_f \tag{2.12}$$

where: $X_j^R =$ the regional output of industry x_i of region R

$c_{if}^N =$ the national final demand input proportions

$Y_f =$ the regional final demand of region R.

From this, the regional commodity balance, b_i is found for industry i by subtracting the total regional demand for commodity i in region R from the total output of sector i in region R, X_i^R, such that:

$$b_i = X_i^R - \tilde{X}_i^R \tag{2.13}$$

The appropriate regional purchase coefficient, a_{ij}^R is then found based on the value of b_i:

$$a_{ij}^R = \begin{cases} a_{ij}^N\left(X_i^R / \tilde{X}_i^R\right) & if\ b_i < 0 \\ a_{ij}^N & if\ b_i \geq 0 \end{cases} \tag{2.14}$$

Thus, if the difference between total regional demand for commodity i in region R and the total output of sector i in region R is positive or zero, the national coefficient is appropriate to use. Conversely, if the difference is negative, national coefficients will overestimate the regional coefficient and they must be reduced by the ratio of total output of sector i and total regional demand so that the difference becomes zero (Miller and Blair, 1985). The chief difference between the location quotient technique and the supply-demand pool technique is that the latter reflects the regional industry mix while the location quotient technique does not (Braschler and Devino, 1993). However, this technique is not without flaws. Because of the need for all local production to be used in-region prior to any exporting, it implies that there are no exports if the regional supply is not large enough to meet regional needs and this assumption may not be entirely realistic (Swanson, Morse, and Westeren, 1999). This ignores the possibility of cross-hauling, that is, when the same good is imported into a region while being simultaneously exported from the region. This can be especially problematic when considering aggregation of products in an input-output table. This ignores imports and exports from a region and only focuses on the net (Jones, Sporleder, and Mustafa, 1972). Again, the earlier mentioned brewery example applies.

The third major technique employed in input-output models is the Regional Science Research Institute (RSRI) approach developed by Stevens and Trainer (1978). The RSRI regional purchase coefficient approach emphasizes consistency across rows. This can be described as "the proportion of regional demand for that sector's output that is fulfilled from regional production" (Miller and Blair, 1985: 301). The regional purchase coefficient for region R and good i is:

$$RPC_i^R = \frac{(X_i^R - E_i^R)}{(X_i^R - E_i^R + M_i^R)} \tag{2.15}$$

where: X_i^R = the total regional output of industry i

E_i^R = the gross amount of the output of industry i that is exported from the region

M_i^R = the gross amount of industry i that is imported.

This ratio represents the amount of output of good i available to local purchasers and users from local production (the numerator) and the total amount of good i available (the denominator) (Braschler and Devino, 1993). It should be clear this ratio will typically take values between zero and one. The value will be unity only if zero amount of good i is imported. Likewise, the value will be zero only if all of the good is exported.

This method was formerly employed by IMPLAN to calculate its RPCs. The former-IMPLAN RPC approach required econometrics in order to calculate its RPCs. While there are likely local characteristics (variables) that can be used to predict the local capacity to produce a given product, the assumption that this relationship is

homogenous across regions of starkly different characteristics is tenuous. (Swanson, Morse, and Westeren, 1999). Finally, concern has been expressed that the estimations employed by this technique do not differ significantly with "best guess" estimations and therefore, hold little to no explanatory power when compared using Theil's U, root mean square errors, and regressions (Stevens, Treyz, and Lahr, 1989).

A method proposed by Swanson, Morse, and Westeren (1999) is based on the use of the value-added tax data in Norway. The authors collected data from individual firms on: 1) total sales, 2) sales to firms or final consumers within the region, 3) sales to the rest of Norway, and 4) exports outside of Norway. They calculated their RPCs as the ratio of total local sales to total local demand at the consumer price level. While, the authors tout this method as theoretically superior, its use in the United States and other countries that do not have value-added tax schemes is highly limited.

IMPLAN now utilizes a different method, the National Trade Flows Model. The method developed by IMPLAN is a

> double constrained gravity model where the import and export flows between regional economies are modeled as flows between county nodes. In general terms, the import and export flows between regions are thought to be proportional to the "mass", "attractiveness" or "size" of an economy and inversely proportional to the "distance" or cost of moving goods and services between them (Lindall, Olson, and Alward, 2006: 76).

Thus, as the size of the economy increases, the greater the amount of commodities flowing to and from that region, *ceteris paribus*. Likewise, as distance between regions

becomes smaller, the greater the amount of trade between the two regions and as distance increases the amount of trade decreases. Gross domestic supply and demand for all regions must add to national domestic supply and demand for the nation. Let **M** be the byproducts matrix for region r, *x* be the industry output for region r, *z* is non-industry output for region r, and *s* is the domestic commodity supply for region r. "The byproducts matrix represents the proportion of each commodity an industry produces" (IMPLAN, 2000: 271). For example, a dairy's chief output is milk, but it likely also sells some heifers. Thus:

$$\mathbf{M}_r x_r + z_r = s_r \tag{2.16}$$

Further, for regional gross domestic commodity demand, let **A** be the gross absorption matrix for region r, *y* be gross final commodity demand for region r, *f* is foreign exports from region r, and *d* be gross commodity demand for region r. Then:

$$\mathbf{A}_r x_r + y_r - f_r = d_r \tag{2.17}$$

Summed over all regions this requires that all demand and supply be accounted for as interregional trade, foreign exports or imports:

$$\sum_r s_r = \sum_r d_r \tag{2.18}$$

In order to estimate the flows of commodities an interaction model is needed. The form utilized by IMPLAN is:

$$T_{ij} = A_i B_j O_i D_j d_y^{-b} \tag{2.19}$$

where: T_{ij} = trade flow between regions i and j

O_i = total commodity supply originating in region i

D_j = total commodity demand originating from and used within region j

$$A_i = (\sum_j B_j D_j d_{ij}^{-b})^{-1}$$

$$B_j = (\sum_j A_j O_j d_{ij}^{-b})^{-1}$$

d_{ij}^{-b} = the distance function

B_j requires the satisfaction of two constraints:

$$\sum_j T_{ij} = O_i \quad \text{and} \quad \sum_i T_{ij} = D_j \tag{2.20}$$

Because A_i includes B_j and B_j includes A_i they must be solved by iteration (Lindall, Olson et al., 2006).

Input Coefficients

The input requirement, x_{ij} of industry i will be partially satisfied by local suppliers. The proportion of this met by local suppliers forms the RPC as detailed above. "A set of coefficients based on inputs supplied from firms within the region for outputs of firms in the region would reflect both regional production technology and the input amounts to be expected from inside the region" (Miller and Blair, 1985: 50). If z_{ij} is the dollar flow of goods from sector i in the region to sector j in the region and X_j is the gross output in sector j in the region then the input coefficient for the region can be calculated as:

$$t_{ij} = \frac{z_{ij}}{X_j} \tag{2.21}$$

As to be expected the input coefficients can hold any value between zero and unity depending on both the input requirements of the industry and the amount that is supplied by in-region producers. As the values for regional purchase coefficients range from 0 to 1 (and thus, the amount supplied by local producers can range from 0 to 1), the amount of local demand that can be met through local sources directly impacts the t_{ij}'s of our model. When the RPC for a given industry is unity and thus all local demand is met by local supply, the z_{ij}, is the x_{ij} as described above. Conversely, when the RPC is zero and none of the local demand is met by local supply, the resulting z_{ij} is zero and therefore the t_{ij} is also zero. The z_{ij} can only be unity when both the entire local demand is met by local supply and it is the only input used in the production process. The column sum of these input coefficients can hold any value between zero and unity, the latter occurring only if the RPCs for all inputs required by the industry are 100%.

Thus, in a small regional economy, even though an input may be an important input to the production process, its impact on multipliers may be minimized or negated altogether. The logic behind this is relatively straightforward. As the supply of a given input in a regional economy approaches zero, multiplying a given technical coefficient by this will produce smaller values. If the available amount is zero, then any technical coefficient multiplied by zero will produce zero. As such, even though the input is crucial to the production process, the dollars allocated to this input will be lost as leakages to economies outside of the region.

Multipliers

At their simplest, multipliers can be thought of and calculated as the total effects divided by the direct effects (Miernyk, 1967). Thus, the ith multiplier is calculated as:

$$Multiplier_i = Total\ Effect_i / Direct\ Effect_i \tag{2.22}$$

where: $Total\ Effect_i = Direct\ Effect_i + Indirect\ Effect_i + Induced\ Effect_i$

As to be expected, multipliers decrease as the amount of leakage in an economy increases. Thus, multipliers at the national-level are higher than those at the state-level and the state-level multipliers obviously trump those at the county-level. Multipliers can further be distinguished by what they are measuring. Multipliers can be calculated for employment effects, output effects, value-added effects, income effects, or at the sectoral level (Miller and Blair, 1985).

Multipliers can further be distinguished by type. Type I multipliers are the simplest, only measuring the relationship between direct and indirect changes in income from a one dollar increase in the output of all the industries in the processing sectors. Type II multipliers take into account the direct and indirect as well as the induced effects and thus, are much more frequently of interest. As to be expected, the magnitude of Type II multipliers is larger than those of Type I multipliers *ceteris paribus*. Further, it should be noted that larger direct effects do not automatically translate to larger multipliers (Miernyk, 1967).

Accuracy in Input-Output Models and Sources of Error

Accuracy in input-output models is an elusive goal. First, we must distinguish between accuracy in the tables and matrices of the model and model accuracy. From the former two major sources of error arise: data error and errors in compilation. Data error refers to the ability to collect accurate data. Compilation errors arise from the multitude of ways researchers can construct input-output tables from the same data, providing differing results. Model accuracy measures how closely the input-output model reflects the "real world" economy it seeks to emulate (Jensen, 1980).

Jensen (1980) further distinguishes between "partitive" and "holistic" accuracy. Partitive accuracy focuses on the individual elements of the tables. Holistic focuses on the accuracy of the model as a whole, even more specifically how the model represents the most important and largest elements of the economy it is to represent. Further, Jensen concludes that only holistic accuracy is a reasonable goal.

The technical coefficients at a national level can be seen as a blueprint of the production process. When reduced to a regional level this is not the case.

> Technical coefficients of a regional model are not intended to describe the production processes of a sector within the region. Rather these coefficients show the ratios of inputs of locally produced goods and services to output by local sectors. The prime difference between the two is the magnitude of exports and competitive imports for the region (Jones, Sporleder, and Mustafa, 1972: 10).

The magnitude of these imports and exports can significantly impact the results generated by an input-output model. This will lead to bias in the interdependence

coefficients matrix. The direction of this bias will be determined by the relative size of exports and imports as well as how economically integrated the region in question is. This bias arises because trade flows are often estimated on the net, where positive balances represent net exports and negative balances represent net imports. Net imports lead to a reduction in the row coefficients for that sector. However, counterbalancing exports and imports are ignored. This leads to regional inter-industry trade to be overstated as "all products imported or exported in excess of those represented by net import or net export values to be treated as purchases from sectors within the regional economy" (Jones, Sporleder, and Mustafa, 1973: 68).

In addition, there is concern if the procedure to adjust national coefficients to the regional level is done haphazardly and is merely a modification of national tables using only local imports and exports to modify the column totals that the resulting tables will be inaccurate due to fundamental differences in regional versus national production processes (Su, 1970). This concern is due to the reliance on the results of input-output analysis by policymakers and the fact that flawed tables could potentially result in misallocated dollars and a loss in societal efficiency.

The potential magnitude of errors in the calculation of technical coefficients when the tables are modified from national tables is inversely related to the number of industries in the model (Kimura, 1955). Because IMPLAN encompasses 440 sectors, we

can expect the magnitude of these errors to be small relative to models with fewer sectors and large when compared to more disaggregated models.

Sources of technological change in the technical coefficients can be divided into two broad categories: endogenous and exogenous changes. Endogenous changes can be policy-induced or crisis-induced. Exogenous changes can be subdivided into output-related, price-related, innovation-related, and productivity changes. Sources of output-related changes would be scale effects and cumulative output effects. Price-related changes would be technical substitution and process substitution effects. Innovation-related changes could be spurred by process changes and product changes. Finally, productivity changes would be the result of input change and external effects (Rose, 1996). These changes can be of critical importance when dealing with the accuracy of the technical coefficients for emerging industries such as biofuels production.

A cited flaw of non-survey based input-output models' RPCs is their uniformity across rows. What this means is that the same proportion of a given input is purchased locally amongst all users of the input, that is, all sectors of the regional economy (Ralston, Hastings, and Brucker, 1986). The study used primary data from manufacturing firms in Delaware and compared it to non-survey estimated RPCs. Overall, the major finding of the study was that RPCs based on the supply-demand pool method were consistently higher than those found by survey. The difference was a

huge variation in the survey, with RPCs ranging from 0% to 99%. Modifying the RPCs based on the survey data, tended to decrease output multipliers.

A survey-based study of farrow-to-finish hog operations in Minnesota uses various methods to compare the effects of the regional purchase coefficients versus technical coefficients and how the two affected the multipliers calculated in IMPLAN 1998. They tested the regional purchase coefficients by assuming the purchase location was the same as production location, but this tended to overestimate the RPCs compared to the survey-based coefficients. They corrected for this by accepting supply-demand pool estimates as the maximum feasible. When RPCs exceeded the maximum of IMPLAN, they tested three methods: 1) forcing the model to accept the higher RPC by modifying regional information for output until it could meet the local demand, but this overestimated RPCs compared to the survey; 2) using the maximum RPCs as constrained by supply-demand pools and ignoring any excess, but this underestimated RPCs as compared to the survey; and 3) using the maximum allowed by supply-demand and forcing the excess to wholesale trade. This last option allowed only the margins of wholesale trade to be captured regionally. The authors concluded that the technical coefficients had a larger impact than RPCs, but that the RPC of a major input could dwarf the combined effects of all other RPCs and technical coefficients based on the relative magnitudes of the effects of each. The authors attribute this to the constraining of the changes in the RPCs by the supply-demand pool ratio which significantly decreased almost one-half of the survey RPCs (Lazarus, Platas, and Morse, 2002).

However, a Utah study finds that outputs and multipliers are more sensitive to errors in the calculation of RPCs than in technical coefficients. The study uses a simulation of the survey-based 1963 Utah input-output model. The authors applied error variations to both the regional purchase coefficients and the technical coefficients of the 1963 table. The authors generated random errors from a normal distribution and a random number generator. Then, sectoral output and multipliers were calculated based on the original final demand vector. The authors attribute their finding that RPCs are more important than technical coefficients to "mutual cancellation of errors" in the technical coefficients; an increase in one sector's technical coefficient is offset by a decrease in the technical coefficients of other sectors because the technical coefficients must sum to one whereas the RPCs are not related to each other (Park, Mohtadi, and Kubursi, 1981).

The potential for inaccuracy in the calculation of RPCs and its impact on multipliers using the RPC and the supply-demand pool technique was documented in a study of the impacts of tourism in a four county region in north-central Wyoming. The authors focused their study on tourist expenditures for both the lodging and the eating and drinking sectors. These sectors were chosen to represent the bulk of tourist expenditures. Also, there was a large amount of cross-hauling perceived in the lodging sector (Taylor and Fletcher, 1993).

Taylor and Fletcher first used survey data to construct an intraregional commodity sales coefficient (RSC), for both sectors. The RSC represents the proportion of regional commodity supply sold to regional demand. This was used to calculate "observed" RPCs where:

$$RPC = \frac{Net\ Commodity\ Supply\ x\ RSC}{Gross\ Regional\ Commodity\ Demand} \tag{2.23}$$

where: *Net commodity supply = total commodity supply – foreign exports*

Gross regional commodity demand = intermediate + final demand

The results indicated that the observed RPC for the lodging sector was 0.03, while IMPLAN calculated the RPC to be 0.82 and the supply-demand pool method calculated the RPC to be 0.91. In the eating and drinking sector, the observed RPC was 0.51 while IMPLAN was 0.90 and the supply-demand pool method was 0.91. Because regional demand is fairly constant, the net effect of the lower RPCs was to increase imports. Likewise, because regional supply is also constant, the net effect was to increase regional exports. The effects of increased imports and exports are that they decrease the indirect and induced effects. The magnitude of this change is relatively small, but noticeable as the economic impact decreased by 4%. The authors attribute this small change to the fact that they adjusted two RPCs. If a similar change was applied throughout the regional economy the change would be more significant (Taylor and Fletcher, 1993).

Selecting a Non-Survey Method

The choice in the current study to compare the survey data versus the coefficients in IMPLAN as opposed to other models (e.g. REMI, RIMS II, etc.) was made based on several factors. First, IMPLAN is a static model while REMI is a dynamic model. While there are advantages to dynamic models, the survey data is only for the year 2007 and as such, the use of a static model makes the most sense. Additionally, the relatively high cost of REMI datasets may discourage potential users (Rickman and Schwer, 1995a).

One study compared REMI versus IMPLAN and their predictions based on the shock of opening a new automobile assembly plant in Bloomington-Normal, Illinois (Crihfield and Campbell Jr., 1991). The authors are careful to note two potential problems: first, the lack of a direct benchmark for comparison; and second, the direct comparison of before and after effects in an economy is difficult due to the high degree of noise from other events not directly under the influence of a given shock. Both of these points are valid and are likely endemic to any study aiming to measure predictions versus actual, real-life results. To resolve this, the authors compared the effects generated by the models to indirect benchmarks such as BEA-measured values. Another problem facing the authors is the difference in how shocks are entered into the models. REMI defines shocks in terms of employment change and converts that into output changes. Conversely, IMPLAN defines shocks as changes in final demand. In

order to correct for this discrepancy, the authors modified the REMI data to fit IMPLAN and the IMPLAN data to fit REMI. Then the authors ran both programs for the program's own data and the modified data. After all the scenarios were run, the results were: 1) REMI's employment multipliers were higher than IMPLAN's (which, is surprising as IMPLAN assumes perfectly elastic supply of resources, while REMI does not); and 2) stark differences between output-per-employee ratios—REMI's was over twice the Census of Manufacturing data while IMPLAN's was 17% less. Major critiques of the two models are as follows: 1) REMI requires considerable slack in local labor markets, an assumption that may not mirror local conditions; and 2) "IMPLAN's employee-to-output ratio exhibits erratic spatial variation, at least for the motor vehicle industry, which contributes to IMPLAN's underestimation of output impacts for subregions in the state" (Crihfield and Campbell Jr., 1991: 14). While IMPLAN's predictions were closer than REMI's, when compared to BEA values, the authors caution that these results might differ for different industries and regions.

Another study tested the 1992 versions of IMPLAN, REMI, and RIMS II for Clark County, Nevada. The authors benchmarked the results of IMPLAN and REMI using RIMS II as a benchmark. Comparing the default programs of IMPLAN versus RIMS II and REMI demonstrated that IMPLAN estimated the largest multipliers. However, after modification of the three models to control for differences in closure techniques, the three models were statistically indistinguishable from each other (Rickman and Schwer, 1995a).

Brucker, Hastings, and Latham III (1990) did a comparison of five models: IMPLAN, RIMS II, Lamphear and Konecny's ADOTMATR, Stevens' RSRI, and Schaeffer's SCHAEFFER models. They compare the five models' estimated impacts across seven final demand change scenarios. Two of the scenarios were specifically chosen to test against the Texas semisurvey-based 1979 input-output table. The other scenarios had controls for the sector while estimating changes to the region. Because of the lack of true benchmarks, the authors use closeness to the Texas model as a proxy for accuracy. Recognizing that this alone is not sufficient for testing, the authors also compared each model's results to the five-model mean using average deviation as a measure of dispersion.

Given the measures of accuracy, none of the models performed consistently well with respect to the Texas model. As such, the authors also investigated monetary cost, time requirements, flexibility, etc. in order to evaluate the models. While different models excelled at different things (and suffered at different things, as well), the authors conclude that

> If a user puts very heavy weight on monetary and time costs but understand input-output and a spreadsheet well enough to use a total requirements table for repeated analysis, then the better model would be RIMS II. However, if time costs are more heavily weighted than money and if the user wants round-by-round impacts, then RSRI or SCHAEFFER would be the best model.
>
> IMPLAN, a middle-of-the-road model, is less costly and more flexible than RSRI, RIMS II, or SCHAEFFER, but requires less user time, regional data, and knowledge of input-output and computers than does ADOTMATR. For the user with some time, money, computer, and input-output

> literacy, and regional data, IMPLAN offers a reasonable alternative (Brucker, Hastings, and Latham III, 1990: 134).

Thus, each of the models offers some positives, but none is without drawbacks and the selection of model recommended by the authors is based upon user preferences and resources.

The latest version of IMPLAN (v.3) is still new and as such, extensive testing of it versus other nonsurvey programs has not been conducted.

III. LITERATURE REVIEW

Scenario Development

The previous research on the economic impacts of biofuels has centered on corn-based ethanol facilities. Different studies have produced a wide-range of estimated impacts based on model selection and the assumptions employed by the researchers. Hasty estimates based on spurious assumptions and lack of expertise with the models has produced much misinformation in this field (Swenson, 2006). Further complicating research is that biofuel production is categorized as "other basic organic chemical manufacturing" in the IMPLAN model. This sector includes a vast array of products besides biofuels including, but not limited to: gum and wood chemicals, cyclic crude, and solvents (Swenson and Eathingon, 2007). Because of the wide variety of other products produced in this sector (and the myriad of inputs required by these other types of firms), there exists a possibility of a great deal of noise.

Tiffany and Eidman (2003) provide a very thorough discussion of the operations of a fuel ethanol facility. While it is bereft of estimated impacts and regional analysis on the operations of an ethanol facility, it provides a sound, technical explanation of the manufacturing process and cost structure of an ethanol plant. Based on the required inputs to produce a given quantity of ethanol, it allows a plant owner to project cost structure based on adjustable prices for inputs such as corn and natural gas. By measuring the projected cost structure against current market prices for ethanol and

desired return on investment, etc., it allows the plant owner to determine the feasibility of a plant. Readers interested in a more thorough discussion are directed to this report.

Because of the dependence on ever-changing input prices, the Tiffany and Eidman (2003) report does not allow direct projection to technical coefficients. If the user specifies fixed prices, it could potentially be used, but this is highly dependent on the user-specified values. If even a single major input's price changes significantly, this will in turn change all of the technical coefficients because technical coefficients are values based on both price and quantity. The report provides the quantity of inputs needed per gallon of ethanol output, but a set of prices are needed to calculate the technical coefficients.

When modeling the estimated impacts of a biofuels plant's operations on a region, a regional scientist must first construct a scenario. This process includes planning the scope and length of the project, determining key assumptions, etc. The assumptions of the analyst are critical in determining the projected outcome(s) of the scenario. As such it is important that the methodology of the research project be as analytically correct as possible. In addition, it is important that the data reflect the actual operation as closely as possible.

Swenson (2006) is perhaps the most trenchant critic of the scenarios employed by other researchers for estimating the impacts of biofuels production. In light of his criticisms, previous studies on biofuels production will be evaluated.

First, Swenson (2006) is censorious of the way certain factors of production are modeled. Specifically, he is critical of researchers who include increased corn production in their projections. Because the corn was being grown prior to the opening of the ethanol plant, corn production is not a response to the opening of the plant and including it overstates the economic impact. In fact, plants are sited because of the existing corn production. If the corn is legitimate new production, then it is safe to assume that the land was previously employed to grow another crop and this is a substitution from one commodity to another. While local prices for feedstocks may increase because of the increased demand and individual growers may choose to change to the new more profitable crop, only the net should be included, not the gross.

Johnson (1995) conducted a study of the potential impact of a hypothesized soycrush biodiesel facility in Virginia's Northhampton County. Johnson excluded the soybeans used in production. Ma, Scott, and Johnson (1996) conducted a similar study on the operations of a biodiesel facility in Buchanon County, Missouri and they also excluded the soybeans in their analysis. Altman (2002) projected the doubling of capacity for the Macon County ethanol plant and its impacts on Macon and Adair County, Missouri, and excluded corn from his analysis. Petersan (2002) estimated the impact of a corn ethanol facility in Ravenna, Nebraska. His impacts include those of Buffalo County (where Ravenna is located) as well as the 8 surrounding counties. Like the previous researchers, Petersan assumed the creation of the facility did not result in additional agricultural production in the region. Pierce, Horner, and Milhollin (2007)

studied the economic contributions of four corn ethanol plants in Missouri. They excluded the corn production from their analysis. Low and Isserman (2009) follow the steps laid out in Swenson (2006) to model the impacts of four ethanol plants in Illinois. Thomassin and Baker (2000) conducted a study of the impact of a corn ethanol facility in Ontario. They considered three scenarios: one of the scenarios treated the corn as new production while the remaining two treated only treated 1/5 of the corn as new production while the remaining 4/5 was existing production. The first scenario assumes that ethanol produced by the facility replaces gasoline that likely would have been imported. In the second scenario while 4/5 of the corn used comes from existing production, some new production does take place. In addition to increased corn production, barley production increases to meet the needs of livestock producers. In the final scenario, the gasoline sector impact continues the assumption of the second scenario, that is, that the corn used is not new production. It further extends the assumptions to assume that the petroleum sector would see a decrease in demand as ethanol production replaces gasoline. Further, they assume the ethanol would be shipped to blending plants or distribution centers and the cost of this transportation would be borne by the ethanol facility. Any costs after arriving at these destinations would be borne by the retail outlets. Retail outlets would cover these extra costs by charging higher prices. The authors assume these costs to be similar to existing petroleum retailers.

Swenson (2006) also disparages previous researchers' tendencies to fail to acknowledge other regional offsets. First, he finds fault with the way the transportation sector is modeled. Many researchers increase the trucking sector because growers shipping to a local ethanol plant may employ local trucking firms. But prior to the opening of the plant, the corn still had to be transported to some destination. Furthermore, if the corn had originally been trucked to a rail yard for export, then there is a decrease in the trucking sector that many economists do not account for. It should be noted that if the corn or soybeans are excluded from the analysis (as many of the studies cited above did), then this issue might not apply. Furthermore, other services, such as grain handlers, which have traditionally existed to transport corn outside of the region, will surely see a decrease in their business as more and more grain is converted to ethanol.

Thomassin and Baker (2000) acknowledge the impact on alternative users of corn by including an increased demand for barley in their second and third scenarios as livestock producers switch from higher-priced corn to lower-priced barley. However, they do not discuss how they handled the impacts on the transportation sectors, grain handling sectors, etc. Low and Isserman (2009) discuss how ethanol facilities will increase cattle production because of the access to cheap distiller's grains. Swenson (2006) acknowledges this possibility, but claims the distiller's grains can only be used as a supplement and in limited quantities. Low and Isserman do not discuss any of Swenson's other regional offsets or whether they accounted for them. Johnson (1995),

Ma, Scott, and Johnson (1996), Altman (2002), Petersan (2002), and Pierce, Horner and Milhollin (2007) do not discuss if they accounted for these offsets in their studies or not as the level of detail is not provided in their reports.

Swenson (2006) also denounces the way in which declining cost industries are modeled. Many of the inputs of an ethanol plant (electricity, natural gas, rail transportation) come from industries where increases in demand only affect the firm marginally. An increase in the output of these industries does not yield significant increase in employment and salaries. This is due to fact that input-output is linear while these industries are common examples of economies of scale. Thus, the effects of these purchases should yield small multiplier effects. Rail transportation provides an even greater difficulty for researchers as corn and ethanol are both shipped by rail, but ethanol is a volume-reducing output of corn. While there may be higher per car prices for a hazardous material such as ethanol than there is for corn, this has not been addressed in the literature. None of the studies presented excluding Swenson's studies appear to account for this or discuss it in any manner. The exact magnitude of these industries' impact on projections is unknown and is relative to the quantity (and price) of these purchases for a biofuel plant.

Additionally, Swenson (2006) cautions against the treatment of producer premiums. The prevailing belief is that corn growers receive a higher price or save money on reduced transportation costs. Previous researchers have translated this

windfall into household income (e.g. Altman, 2002). But this ignores other factors: if returns are higher, then land rents increase and this can be problematic for tenant farmers. This may not be as much of an issue if the land owners live in the region. Also, if prices increase, then this means lower agricultural subsidies and may nullify the potential increased income. Finally, if the price of corn increases due to the demand created by an ethanol plant, then competing users of corn in the area (such as hog farmers) must pay more for their inputs and this can reduce their incomes.

Johnson (1995) and Ma, Scott, and Johnson (1997), Petersan (2002) acknowledge that increased demand would yield higher prices for soy producers, but do not include higher prices in their analyses. Thomassin and Baker (2000) note that the increased demand for corn will result in decreased subsidies for corn producers, but do not incorporate an increased price in their impacts. Altman (2002) includes a 15 cent per bushel premium to growers in his estimates of direct impacts (and subsequently, this impacts indirect and induced results). Similarly, Pierce, Horner, and Milhollin (2007) include a 10 cent per bushel premium in their estimates. They argue that for each $0.01 per bushel rise in the regional corn market price, the total value of Missouri's corn crop rises by over $2 million. The problem is that if this premium exists, it is local to the firms and results from reduced transportation costs on the corn or soy purchased, not on higher prices. They include this increase in their estimates and this lead to substantially higher multipliers. However, they recognize the competing users of corn prior to the operation of ethanol facilities and account for these users by subtracting "the economic

impacts of the new ethanol industry that displaced those bushels flowing into traditional channels" (Pierce, Horner, and Milhollin, 2007: 6). Low and Isserman (2009) discuss how the difference in ownership of the farmland can affect who receives the lion's share of price premiums. They treat the price premium as proprietor's income instead of household income as they feel the money is more likely to be used to pay debt or reinvested in the property. They note that the effect of this choice is minimal on the outcome.

Finally, Swenson (2006) is critical of the treatment of construction impacts. First, previous researchers appear to assume the opportunity costs of alternative investment projects are zero. Not considering the opportunity costs of alternative projects biases benefit-cost analyses downwards. Second, the construction impacts are at best one or two year benefits and likely the specialized labor, engineers, etc. required for such an undertaking will need to be hired from outside of the region.

Johnson (1995) and Ma, Scott and Johnson (1996) both mention the construction providing a two or three year stimulus to the construction industry, but the benefits are short-lived and they do not include them in the economic impact analysis. Thomassin and Baker (2000) do include the construction impacts for a period of 18 months. Because Thomassin and Baker are modeling for the entire Canadian economy, it should come as no surprise that the majority of the benefits can be and are retained within the economy. Likewise, Pierce, Horner, and Milhollin (2007) include construction impacts,

but switch the analysis from the state of Missouri to the national economy as they project the impacts of future expansions of the ethanol industry in Missouri. They do not provide a time frame for this expansion. Petersan (2002), Altman (2002), and Low and Isserman (2009) do not discuss construction impacts and limit their study to only include the operational phase of an ethanol facility.

Swenson's (2006) modeling approach is to modify the wet milling sector in IMPLAN to create a dry milling sector. He adapted the University of Minnesota model developed by Douglas Tiffany (2003) to Iowa to best fit the technology employed by corn-based ethanol plants. After determining the appropriate purchases based on plant capacity, certain values were modified: the corn production sector's direct impact was reduced to zero; in light of interviews with producers in the declining-cost industries, the labor and labor earnings in natural gas, electricity, water, and rail were reduced to 25% of the levels reported in IMPLAN; and the finance sector was also reduced to 25% of IMPLAN levels under the assumption "that the financing of a plant will not have a durable effect on the region's financial sector" nor does he believe that the debt financing will be local. A final adjustment is to reduce the regional trucking purchases under the assumption that the farmers' paid the cost to ship the corn to the plant and the plant shipped their product by rail (Swenson, 2006: 13). Follow up e-mail exchanges with Swenson confirmed the methods employed by Swenson for the testing of his coefficients and multipliers (Swenson, 2010).

The potential for technological change in the ethanol industry is well-documented. Part of this is attributable to changes in the prices of the feedstocks. As noted by Low and Isserman (2009: 80), "Changing prices confound any effort to measure dollars of inputs divided by dollars of ethanol output, the production function at the heart of an input-output model." This change in prices can be seen merely by a review of the Swenson studies: in 2005 corn accounted for one third of an ethanol plant's total costs (Swenson, 2006; Swenson and Eathington, 2006) whereas in 2007 corn was 60% of an ethanol plant's costs (Swenson, 2008).

Also, gallons-to-employee ratios show a dramatic shift in the cost structure consistent with scale economies, which the linear production function of an IO model cannot capture. Johnson (1995) and Ma, Scott, and Johnson (1997) estimate 46.6 thousand gallons per employee. Thomassin and Baker increased this ratio to 347.6 thousand gallons per employee. The other studies (Altman, 2002; Petersan, 2002; Swenson, 2006; Swenson and Eathington, 2007; Pierce, Horner, and Milhollin, 2007; Swenson, 2008; Low and Isserman, 2009) report ratios between just under 1 million gallons- to 2.5 million gallons per employee. Part of this is likely attributable to changes in the capacity of plants. Whereas the Johnson (1995) and the Ma, Scott, and Johnson (1997) studies estimate impacts of a 3.776 million gallon/year (MGY) facility, this number skyrocketed to 50 MGY in Thomassin and Baker's (2002) study and to 60 MGY and 100 MGY plants in Low and Isserman (2009). Altman's (2002) study of the Macon County ethanol plant modeled the plant's capacity doubling, but employment only

increased from 31 to 36 persons. The fact that plant capacities are rising significantly faster than plant employment highlights the economies of scale at play in the biofuels industry due to economies of scale. This, in and of itself, further highlights why great caution must be exercised when using input-output models to estimate the effects of biofuels production as input-output models assume constant returns-to-scale. Further, they are based on a given scale, while particularly new plants are likely to be on a different scale. The operating capacities, employment and gallons-per-employee ratios are reported below in table 3.1.

Table 3.1: Plant Characteristics

Study	Type	Operating Capacity	Employees	Gallons-per-Employee
Johnson (1995)	Soy Biodiesel	3.776 MGY	81	46,617
Ma, Scott, Johnson (1997)	Soy Biodiesel	3.776 MGY	81	46,617
Thomassin and Baker (2000)	Corn Ethanol	52.834 MGY	152	347,592
Altman (2002)	Corn Ethanol	44 MGY	36	1,222,222
Petersan (2002)	Corn Ethanol	80 MGY	48	1,666,666
Swenson (2006) Swenson and Eathington (2007)	Corn Ethanol	50 MGY	35	1,428,571
Swenson (2008)	Corn Ethanol	100 MGY	46	2,173,913
Pierce, Horner, and Milhollin (2007)	Corn Ethanol	160 MGY (4 plants)	161	993,788
Low and Isserman (2009) - 100 MGY	Corn Ethanol	100 MGY	39	2,564,102
Low and Isserman (2009) - 60 MGY	Corn Ethanol	60 MGY	35	1,714,285

Table 3.2: Multipliers Reported in Previous Studies

Study	Geography	Employment	Output	Value-Added
Johnson (1995)	County	1.67*	2.32*	**
Ma, Scott, Johnson (1997)	County	2.00*	3.22*	**
Thomassin and Baker (2000) - Scenario 1	National	15.99*	3.95*	7.25*
Thomassin and Baker (2000) - Scenario 2	National	**	**	**
Thomassin and Baker (2000) - Scenario 3	National	**	**	**
Altman (2002) - Doubled Capacity	County	7.09*	1.39*	6.80*
Petersan (2002)	Region	3.40*	1.31*	1.74*
Swenson (2006); Swenson and Eathington (2007)	Regional	3.80	1.13	1.38
Swenson (2008)	State	3.70	1.12	1.35
Pierce, Horner, and Milhollin (2007)	State	18.77*	1.83*	6.04*
Low and Isserman (2009) - Hamilton County (100 MGY)	County	3.92*	1.08*	**
Low and Isserman (2009) - Coles County (60 MGY)	County	4.34*	1.14*	**
Low and Isserman (2009) - Kankalee County (100 MGY)	County	6.41*	1.15*	**
Low and Isserman (2009) - Hardin County (60 MGY)	County	2.83*	1.06*	**

* Calculated by the author of the present study
** Could not be calculated from data given

Table 3.2 above summarizes the multipliers calculated in the various studies. In reviewing the table it should be pointed out that studies of a larger geographic scale or of a larger economic region will provide larger multipliers. Johnson (1995) calculates an employment multiplier of 1.67 and an output multiplier of 2.32. Ma, Scott, and Johnson's (1997) employment multiplier is 2.00 and the output multiplier is 3.22. The reason for the higher multipliers in the Ma, Scott, and Johnson study relative to the

Johnson study despite similar assumptions is that the latter study featured a more economically-integrated region than the former. Neither study however discusses the value-added multipliers. The Thomassin and Baker (2000) study was unclear on the reporting of their impacts and as such, multipliers could not be calculated except for the first scenario. Further, it should be recalled that the Thomassin and Baker multipliers are for all of Canada. Altman (2002) reports higher employment and value-added multipliers of 7.09 and 6.80, respectively, with a lower output multiplier of 1.39. The reason for Altman's higher employment and value-added multipliers is the premium paid to the producers of corn. Petersan's (2002) employment multiplier is 3.40 and his output multiplier is 1.31, both relatively in line with the multipliers calculated in Swenson (2006), Swenson and Eathington (2007), and Swenson (2008). Petersan (2002) separates his value-added multipliers between personal and other property income effects and indirect business taxes effects, but when they are aggregated, the total value-added multiplier is 1.74. The reason this multiplier exceeds the multipliers reported in the Swenson studies is likely Swenson's assumption that the ownership of the plant is external to the region. Pierce, Horner, and Milhollin report a significantly higher employment multiplier of 18.77 reflecting the per bushel premium paid to growers and their estimate of its impact on the total value of the entire Missouri corn crop because they are doing a state-level analysis. Low and Isserman (2009) report employment multipliers ranging from 2.83 to 6.41, output multipliers ranging from 1.06 to 1.15, but fail to report value-added multipliers. The reason for the large variation in

employment multipliers is that Kankalee County is part of the Chicago-Naperville-Michigan City combined statistical area, while the remaining three counties are significantly smaller and located in the less-densely populated southern portion of Illinois.

Testing Regiments

One of the first formal tests of adapting national-level coefficients to the regional level was done by Czamanski and Malizia (1969). Using a case study approach, the authors used regional income and product accounts to modify the National Input-Output Table for 1958 in order to construct the Washington State Input-Output Table for 1963 (hereafter, estimated) and then compared it to the Washington State table based on direct field data (hereafter, benchmark). They compared the estimated table and the benchmark table using a variety of techniques. First they transformed the deviations into percentages and calculated the mean, standard deviation, and distribution by deciles of error. This allowed for some assessment, but had the major drawback that deviations from smaller benchmark coefficients held much more weight than equal deviations from larger benchmark coefficients. Because weighting the deviations by the absolute size did not correct this weighting issue, Czamanski and Malizia borrowed from information theory to assess the accuracy of their estimated coefficients. Czamanski and Malizia tout the advantages of this as the accuracy of the estimated table being able to be viewed as a function of the benchmark table. If the

benchmark coefficients are correct, then the information content of the estimated table would be small. Overall, Czamanski and Malizia found that industries that are highly locally-concentrated cannot accurately be estimated using short-cut methods. The exception to this was the aerospace industry, where the national average follows the Washington state average as the national aerospace industry is heavily based in Washington state (11% of total national employment in the aerospace industry was in Washington). When they removed the most locally densely-concentrated sectors from the analysis, the information value fell significantly, indicating a better goodness of fit.

Schaeffer and Chu (1969) tested the ability of non-survey RPC estimation techniques to effectively estimate the 1963 Washington state table and compare it to the survey-based table. In order to test the effectiveness of the techniques, Schaeffer and Chu used a chi square test of each column in the direct requirements table. Overall, the authors found that none of the techniques tested were sufficient to replace a survey-based table.

Hewings (1971) followed the information theory approach laid out in Czamanski and Malizia (1969) in comparing their attempts to adopt the United Kingdom tables for West Midlands using a variety of non-survey estimations. He modified the approach, however, such that if either the national or regional coefficient was equal to zero and the corresponding coefficient in the other table was greater than zero, the zero coefficient was raised to 0.000001, which while not significantly different than zero, was

a non-zero value, nevertheless. No such modification was made if both elements were zero. He concluded that different non-survey techniques may be capable of effectively estimating regional tables, however, because of the range of the sectoral multipliers, one cannot suggest that any given method is preferable.

McMenamin and Haring (1974) developed a technique of updating coefficients and tested their method against the RAS method developed by Stone and Brown and a naïve method which assumed no change. The method proposed by McMenamin and Haring "adjusts the regional table at Year 0 for changes in prices, the effects of substitution, and the effects of fabrication that have taken place between Year 0 and Year 1" (McMenamin and Haring, 1974: 194). The RAS procedure on the other hand modifies the Year 0 coefficients to Year 1 coefficients "by the same substitution multiplier, which we shall denote r_j; and each of the coefficients a_{1k}, a_{2k}, …, a_{nk}, relating to the intermediate inputs into k, is modified by the same fabrication multiplier, which we shall denote by s_k. This means that ${}_1A$ is related to ${}_oA$ by the equation ${}_1A = \hat{r}_0 A \hat{s}$" (Bates and Bacharach, 1963: 28). The McMenamin and Haring procedure differs from the RAS method of Stone and Brown in that it eliminates the need "to estimate the total intermediate output and total intermediate input vectors" (McMenamin and Haring, 1974: 194).

They tested each of the techniques' abilities to update the full-survey 1963 Washington State table to the full-survey 1967 Washington State table. In order to

gauge the effectiveness of their technique, they tested the absolute percentage deviation, the mean of absolute percentage deviation, and its standard deviation. Finally, they tested the techniques using a chi-square test similar to Schaeffer and Chu (1969). They draw two major conclusions:

> In the first place, in estimating an input-output table for Year 1 based on one in Year 0, the accuracy obtained will vary with the amount of structural change actually experienced in the economy from Year 0 to Year 1.…In the second place, if a full-survey table is available for a region at a particular date, it may be most worthwhile to estimate a table for a second date based on this table than to construct a second full-survey table or try to derive a regional table from an existing national table (McMenamin and Haring, 1974: 204).

Morrison and Smith (1974) tested various non-survey reconciliation techniques against the regional model for Peterborough, UK. They used the Isard and Romanoff mean similarity index (Isard and Romanoff, 1968), the chi-square test (Schaeffer and Chu, 1969), Czamanski and Malizia's (1969) information theory approach, the mean absolute difference, and the correlation coefficient. They conclude that the RAS technique is superior due to it performing best in each of the tests.

Stevens and Trainer (1978) tested the effects of random errors in the technical coefficients and in the RPCs to see which had a larger impact on the multipliers. They first generated hypothetical input-output tables (orders of 10, 25, 50, and 100) and ran the models to generate benchmark multipliers. Then they applied random errors drawn from a normal distribution and renormalized the column's so that they would still add to one. Finally, they returned to the original tables and applied random errors to the

hypothetical RPCs. In order to compare the effects on the multipliers, Stevens and Trainer tested the multipliers of the different estimated tables against the hypothetical table using Theil's inequality coefficient. They conclude that the errors in the RPCs had a larger impact on the accuracy of the estimated multipliers than errors in the technical coefficients. The logic behind this conclusion is quite simple—errors in the technical coefficients that lead to overestimation of a given technical coefficient are offset by a reduction in other technical coefficients, however, the overestimation of a RPC is not constrained by the other sectors' RPCs and thus, there is nothing to offset the overestimation.

Butterfield and Mules (1980) move from a few simple tests to a thorough and extensive testing regiment to evaluate the accuracy of non-survey reconciliation methods. It should be noted that this regiment is an iterative process in which failure at a single stage can be interpreted as sufficient grounds to reject the hypothesis that the benchmark and estimated coefficients are the same. First, Butterfield and Mules employ a simple sign test to see if there is consistent over- or under-estimation of the coefficients. Second, a simple regression is used to test if the benchmark coefficient is the same as the estimated coefficient. Third, a chi-square test (Schaeffer and Chu, 1969) is employed. Next, measures of distance from the mean are obtained. These include the mean absolute difference, the standardized mean absolute difference, and Isard and Romanoff's (1968) mean similarity index. Finally, Butterfield and Mules use the information index developed by Czamanski and Malizia (1969). They conclude that

the RAS method is superior to the McMenamin and Haring (1974) and a naïve reconciliation techniques. Further,

> In applying the test routine, it was clear that there are benefits in using a series of complementary tests rather than relying on one statistical method in such applications. In this particular case, the non-parametric test indicated little or no bias. However, subsequent tests suggested that this was accompanied by some under- and over-estimating which tended to cancel out in the non-parametric test….Finally, it should be noted that the testing routine is quite time consuming. However, it is exhaustive and provides much more information about an estimated matrix than does any single statistical test (Butterfield and Mules, 1980: 307).

Harrigan, McGilvray, and McNicoll (1980) compared reconciliation techniques to test how accurately those techniques could be used to adapt the UK table to Scotland and then compared the results to the actual survey-based Scotland tables. They employed a series of tests: the mean similarity index, the Euclidean metric difference, the Isard and Romanoff (1968) mean similarity index, the chi-square test (Schaeffer and Chu, 1969), the absolute mean relative difference, Czamanski and Malizia's (1969) information index, and finally the correlation coefficient. These tests were compared to the entire table and individual columns to determine not just the holistic accuracy of the tables, but also which sectors were best simulated. The authors conclude the RAS technique provided superior performance to the other non-survey reconciliation techniques, however, the results of the absolute mean relative difference test indicate "significant relative differences between individual coefficients, even for the RAS technique" (Harrigan, McGilvray, and McNicoll, 1980: 932).

Park, Mohtadi, and Kubursi (1981) in their testing of the effects of RPCs versus the effects of technical coefficients (detailed above in chapter 2) employed Theil's U to test the effects of each. The authors found that RPCs had a larger impact than technical coefficients on the accuracy of the multipliers.

Stevens *et al.* (1983) developed a new method of calculating RPCs (see the RSRI method detailed in chapter 2). In order to test it they employed the mean absolute relative difference, the Euclidean metric difference, the Isard and Romanoff (1968) mean similarity index, and the regression employed by Butterfield and Mules (1980). They tested this RPC estimation technique using the survey-based 1977 Washington state and the semi-survey-based 1975 West Virginia state models.

To cope with zeros in the original Washington state survey table, they took two steps. First, whenever any transaction was less than $50,000 in the estimated table, they rounded it down to zero. Then, whenever the benchmark table had a zero, but the corresponding RPC was nonzero, a random number between $1 and $49,999 was entered in the table. These steps were taken because in the original Washington survey table, transactions less than $50,000 were rounded down to zero. The authors found, "that the non-survey coefficients vary widely from industry to industry in their ability to simulate the survey-based coefficients" (Stevens *et al.*, 1983: 280). However, they find over all the results to be satisfactory. When the authors compared their technique to the West Virginia semi-survey model, "the results were at best mixed" (Stevens *et al.*,

1983: 283). Stevens *et al.* defend their model's performance on the basis that the largest differences were not in the regional coefficients, but instead in the labor rows and household columns.

Stevens, Treyz, and Lahr (1983) expanded on the testing of the RSRI RPC method by using a simple regression, a linear regression with a logarithmic transformation of the benchmark and estimated RPCs, root mean squared error, and Theil's U. They find that the RSRI RPC method is superior to the other RPC estimation techniques. Further, the other RPC estimation techniques tended to systemically overestimate the RPCs based on the regression results.

Lahiri (1984) tested his reconciliation procedure of using endogenously-weighted cells against the Jensen and McGuarr (1976) reconciliation procedure. In order to gauge the effectiveness of his procedure, which he claimed to be less-time consuming, he estimated using both procedures and then compared the results. He employed Theil's U, average absolute percentage deviation, and root mean squared percentage deviation. Lahiri concludes that his technique performed similarly to the Jensen and McGuarr technique based on the comparisons of the multipliers, however, his proposed technique would be significantly less time-consuming.

Brucker, Hastings, and Latham (1990) in their comparison of various models to the survey-based Texas models (detailed above in chapter 2) used eyeball comparisons and the percentage difference from the Texas model mean. Finally, they looked at the

dispersion using the coefficient of variation. Once again, the major finding was that RPCs based on the supply-demand method were consistently higher than those based on a survey.

Rickman and Schwer (1995a) in their comparison of IMPLAN, REMI, AND RIMS II (detailed above in chapter 2) used a matched t test, a simple sign test, the Spearman rank correlation coefficient, and the parametric correlation coefficient to compare the multipliers. In Rickman and Schwer (1995b), they employed fewer tests limiting their analysis to a matched t test, the Wilcoxon sign test, and a simple sign test to compare the multipliers of IMPLAN and REMI. Rickman and Schwer are not clear on why they employed fewer tests in the later study. The second study examines multiplier stability across versions. The major finding of the second study was that the multipliers of the default versions of the models had changed significantly, but that this change "was found to be caused by undocumented or poorly documented changes in the default status of the models across versions, not by changes in the underlying structure" (Rickman and Schwer, 1995b: 150).

Willis (1997) compared the multipliers derived from non-survey techniques versus those of the survey-based input-output tables for Wales and Staffordshire. He employs the absolute difference (minimum and maximum), the R^2, Spearman's rank correlation coefficient, and finally the arithmetic mean ratio for all the 33 sectors of the table based on frequency ratios. Willis finds that the simple location quotient technique

performed poorly, overestimating the Staffordshire type-I employment and output multipliers by 21.8% and 21.3%, respectively. The simple location quotient technique for estimating RPCs performed slightly better for the Welsh table, overestimating the type I output multiplier by 12.9%. They attribute this better performance to the use of a national input-output table for the Welsh table that did not exist for the Staffordshire table. The RSRI RPC estimation technique performed much better only overestimating the Staffordshire output and employment multipliers by 14.3% and 13.5%, respectively and the Welsh output multiplier by 5.9%.

Jalili (2000) compared various reconciliation techniques to update the survey-based 1966 USSR tables to the survey-based 1972 USSR tables. It should be noted that Jalili cautions about the accuracy of both tables as both tables may have been deliberately manipulated for propaganda purposes during the Cold War. He too employs a battery of tests including the number of negative coefficients; the mean absolute deviation; the mean, standard deviation, and maximum value of the coefficient of equality; Theil's U; the degree of approximation (at 5%, 10%, and 20%); the root mean square; the mean, standard deviation, and maximum value of the estimated coefficients; and finally, the standardized total percentage error. Taking Jalili's cautionary note on the accuracy of the Soviet tables into account, Jalili states that the RAS method performed superior to the other reconciliation techniques based on their rank in the battery of tests. However, he further cautions that the RAS performed superiorly in a holistic sense, as opposed to a partitative sense.

Crihfield and Campbell (1991), Taylor and Fletcher (1993), Oosterhaven, Knijff, and Eding (2003), and Bonn and Harrington (2008) only used eyeball comparisons and none employed any formal testing routines.

Summary

As noted by several authors (Butterfield and Mules, 1980; Harrigan, McGilvray, and McNicoll, 1980; Jalili, 2000), none of the individual statistical tests can provide a meaningful measure of the closeness of two tables. As such, a battery of tests must be employed with each test designed to test a different aspect of the closeness between the tables. As previously discussed, one of the first attempts to test short-cut estimation techniques was the work of Czamanski and Malizia (1969). While one of their short-cut techniques revealed an information theory (I) value of 0.779, which was substantially preferable to the other techniques whose values ranged from 6.408 to 54.262, it still retained a 38.93% mean percentage error in the coefficients clearly demonstrating that a low value of I alone was not sufficient in order to gauge the accuracy of an input-output table (Miernyk, 1969). Other measures of distance from the mean such as the mean absolute deviation, the standardized mean absolute deviation, and mean percentage error can be used to reveal if there is any skewness in the coefficients, but do not provide an objective basis with which to test the values (Jalili, 2000).

The Chi-Square test employed by Schaeffer and Chu (1969) and subsequent authors (e.g. Butterfield and Mules, 1980) measures goodness of fit between sets of coefficients and provides for testable hypotheses. Round (1983) and Morrison and Smith (1974) are both condemning of the usefulness of the Chi-square statistic because it cannot account for situations in which a_{ij} is zero and a_{ij}^* is nonzero. Morrison and Smith propose omitting such elements in the calculation, but this removes some of the credibility of the remaining values. Harrigan,, McGilvray, and McNicoll (1980) take this further to suggest omitting any pair of coefficients in which a_{ij} or a_{ij}^* are zero values. McMenamin and Haring (1974) offer an alternative workaround of any cell in which the gross flow is less than $50,000 to be set equal to zero. Then, the zero cells are given a gross flow of $49,000, combined into one cell, and a new technical coefficient is computed for the combined cell. The chi-square test provides the same information (goodness of fit) as the information index developed by Czamanski and Malizia (1969), but doesn't require independence between the variables.

A regression of the benchmark, survey-based coefficients on the estimated coefficients can reveal if the two sets of coefficients are essentially the same, but cannot reveal over- or underestimation. Round (1983) is critical of this approach and his criticisms will be addressed in the following chapter.

A Wilcoxon signed-rank test can reveal if there is consistent over- or underestimation of the coefficients, however, it cannot reveal the magnitude of this

over- or underestimation. In addition, it requires independence of the variables, something that is not possible when comparing matrices of technical coefficients which must sum to one.

Theil's U can be employed to measure overall forecasting error, however it is most useful when it is used in conjunction with its components which "provide some insight into the variations due to bias of estimation, and the variance and covariance of U" (Jalili, 2000: 231).

The Isard-Romanoff (1968, cited in Morrison and Smith, 1974) mean similarity index measures the magnitude of relative changes, but is subject to the weaknesses that it assumes a normal distribution and that it cannot accommodate situations where the survey and non-survey coefficient are both zero.

Correlation measures such as the Spearman rank correlation coefficient can measure statistical correlation between variables, but cannot reveal the size of difference in errors as can the mean percentage error and other measures of difference from the mean.

In sum, it should be stressed that no single test will reveal every difference between two tables, two vectors of coefficients, two sets of multipliers, and the like. In order to properly assess the accuracy of tables a variety of tests must be employed to measure different elements. Of equal (if not, arguably greater importance) is the accuracy of the output of these models for both researchers and policymakers.

IV. METHODS

Preparing the Data

Data on recent Missouri biofuel plants were gathered by e-mail survey using an EXCEL spreadsheet. To maintain confidentiality we will not be revealing the number of firms. The survey for the year 2007 included: total sales, value of purchased inputs, the percentage of the inputs purchases from in-state sources, and profits. The survey can be found in Appendix A.

One of the drawbacks of surveying firms is their lack of willingness to divulge information, a point raised by Swenson and Eathington (2006: 8), "A reader may ask why we do not just ask or survey modern ethanol operations and get actual production information. The answer is simple: industries loathe revealing cost and income information, especially in the kind of detail that allows for reliable input-output modeling." This should come as no surprise—firms are reluctant to release information which may be usable by a competitor.

When conducting surveys of firms tradeoffs must be made between the ease of the respondent and technical accuracy in order to increase response rates. We worked with representatives from the industry to develop a questionnaire that would be understandable and easy for plant managers to complete. While it would be useful to have the North American Industry Classification System code for the firms' input

suppliers, this information is likely outside the expertise of the manager completing the survey. To lessen the burden on the respondents, some of IMPLAN's 440 sectors were aggregated into a smaller number of broad categories. For example, utilities were a single category instead of breaking it into electricity, natural gas, water, etc. Similarly, while we ask the percentage of purchases made in-state, we don't know if the items in question are sold by an in-state producer or are purchased from a wholesaler who happens to be in-state. As such, while the survey can give us a volume of useful information, it is not without limits. But direct, primary surveys provide data that are more likely to reflect the actual purchases of firms in a rapidly changing industry than pre-packaged data, which by its nature is an average of firms in the sector and is several years old.

IMPLAN employs a rectangular accounting system and thus reports transactions in two formats: 1) industries and 2) commodities. Industries contain firms with similar input cost patterns. Industry purchases are used as a proxy for changes in final demand. For example, in this case the changes in final demand are the purchases by the biofuel firms. Multiple industries can contribute to the supply of a commodity. Because industries are categorized into sectors based on their chief output, many firms produce co-products or secondary products in addition to their primary output (Minnesota IMPLAN Group, 2010).

The firms' purchases were entered into IMPLAN's commodity scheme. Sometimes, imputations of the ratios of these purchase categories were created relative to relevant IMPLAN sectors. Continuing with the utilities example, utilities was divided into four commodities: power generation and supply, natural gas distribution, water and sewage systems, and waste management. Mathematically, the percentage allocated to a particular commodity i, x_i is given by the formula:

$$x_i = \frac{X_i}{\sum_{i=i}^{n} X} \tag{4.1}$$

where: X_i = technical coefficient of commodity i in IMPLAN

n = the number of relevant IMPLAN sectors aggregated in the survey.

This is multiplied by gross spending in the survey-aggregated category to determine the firm's spending for a given commodity. For a full list of aggregations consult Appendix B.

Next, the percentages of in-state purchases were calculated for the respondents. While IMPLAN terms these RPCs, to distinguish between the two, we will refer to the survey-based RPCs as regional supply coefficients (RSCs). When the firm specified a percentage of in-state spending for a given commodity, the survey response was used. In an aggregated category, this percentage was applied to all relevant commodities. Unfortunately, not all of the respondents provided the in-state purchase percentage, because this is not information that firms generally track. When a firm did not supply a sector's percentage of in-state spending, the regional purchase coefficient specified in

IMPLAN 2007 was used. When the RSC was not supplied for an aggregated category, IMPLAN 2007's RPC for each individual commodity in the aggregated category was used.

To compute the average RSC for all of the firms the in-state spending percentage for each sector was multiplied by total outlays for that sector for each firm, in other words a weighted average based on firm purchases The dollar value of these outlays were then divided by total (in-state and out-of-state) spending in that sector to produce a RSC. Thus, the RSC for a given commodity i is:

$$RSC_i = \frac{\sum_{j=i}^{n} L_{ij}}{\sum_{j=i}^{n} X_{ij}} \tag{4.2}$$

where: L_{ij} = in-state spending by firm j in sector i

X_{ij} = firm j's total expenditures in sector i.

In order to derive our input coefficients, the firms' purchases are entered into the appropriate commodity along with the RSC for that commodity. The scenario is then run in IMPLAN and the input coefficients are formed from the direct impacts. The scenario will separate out margins and apply the appropriate leakages to margin sectors. The input coefficients are then formed by the ratio of direct impact in sector i to total spending by the firms. Thus, the input coefficient for a given sector i is:

$$t_{ij} = \frac{direct\ impact\ in\ sector\ i}{sum\ of\ firms' expenditures} \tag{4.3}$$

Finally, a discussion of the treatment of employee salaries and wages, employee benefits, proprietors' income, retained earnings, and depreciation is in order. Employee salaries and wages and employee benefits were first adjusted based on the percentage of employees residing in-state. One firm did not report this percentage and because the firm was located close to the state line, we did not feel we could assume that all employees lived in-state. As such, this percentage was derived as the percentage of the county's workers residing in-state using the 2000 U.S. Census Bureau's commuting patterns (U.S. Census Bureau, 2000). Much like the treatment of material inputs, the percentage of in-state employees was averaged over the firms using a weighted average. Employee salaries and wages and benefits were treated as a household income change in the $50-75,000 salary bracket consistent with the average salaries as determined by our surveys.

Proprietors' income, retained earnings, and depreciation were all adjusted for the amount of in-state spending following the logic of earlier steps. Retained earnings and depreciation were included as forms of proprietors' income under the logic that retained earnings increase the assets of the proprietors and that it will eventually be spent like normal income; and depreciation is a return of previously invested money. These sums were entered into IMPLAN as a household income change in the $75-100,000 salary bracket.

With a change in household income, IMPLAN first deducts personal taxes and savings to get disposable income. It then proportions out spending based on the Personal Consumption Expenditures patterns as reported by the U.S. Bureau of Economic Analysis. Further, it should be noted that the firm's state and federal tax payments were excluded from the analysis as this is not a test of Type III multipliers (however, they were retained in the sum of firm expenditures).

Scenario Development

Following the arguments laid out in the previous chapter, the purchases of corn and soy were dropped from the analysis. After these were zeroed out in both IMPLAN and our survey's data, the scenario was shocked with a 100 million dollar final demand to the other basic organic chemical manufacturing sector. This sector represents the appropriate sector for biofuel production within IMPLAN's sectoring scheme. While $100 million may seem like a large shock, the average output of our firms was roughly $98 million.

First, the shock was run for the survey data. It should be noted that despite zeroing out the direct purchases of corn and soy by the firm, some corn and soy does re-enter the scenario as a first round effect. This is because ethanol plants purchase methanol and other catalysts, some of which are purchased from the other basic organic chemicals manufacturing sector, which does purchase corn and soy to make the variety of outputs it produces. Zeroing out these coefficients in the sector's purchasing

scheme was debated, but without extensive knowledge of the production processes of all other firms producing in this sector, this may potentially bias the results. Further, modifying the coefficients of the sector would carry through to later rounds of economic activity.

Second, the shock was run using IMPLAN's standard technical and regional purchase coefficients for the other basic organic chemicals manufacturing sector without corn and soy. The third scenario used IMPLAN;s other basic organic chemicals manufacturing sector, but removes all agricultural and forestry purchases. This was done because as noted earlier, the other basic organic chemical manufacturing sector produces a wide range of outputs many of which have input needs that may differ drastically from the input needs of biofuels. A fourth scenario was run using an approximation of Swenson's (2006) modified wet milling sector (detailed in chapter 3). To maintain consistency, the other property-type income indicated by IMPLAN in the non-survey methods was added to proprietor income but no assumptions were made regarding as to which households would receive the income changes and the allocation was performed by IMPLAN.

Thus, the scenarios, we are testing can be described as follows: first, the benchmark survey-based scenario (hereafter, Survey), IMPLAN's other basic organic chemical manufacturing scenario with zeroed out corn and soy coefficients (hereafter, Base), IMPLAN's other basic organic chemical manufacturing scenario with all

agricultural- and forestry-based purchases zeroed out (hereafter, No Agriculture), and the procedure developed by Swenson (2006) (hereafter, Modified Wet Corn Milling).

Comparing Input Coefficients

As noted by Harrigan, McGilvray, and McNicoll (1980: 930), "There is no unique measure for estimating the 'closeness' or otherwise of input/output matrices." As such in order to evaluate the accuracy of the coefficients a testing routine must be developed. In order to properly assess the accuracy of the various non-survey-based estimated coefficients, we will directly test the input coefficients themselves as well as the multipliers derived from them. The comparison of multipliers is equally important, if not more important, as most often a policy maker will solely be interested in the accuracy of the results of the model, not in the accuracy of the coefficients themselves (Willis, 1997).

Butterfield and Mules (1980) provide a framework for testing the input coefficients, however, in a break from Butterfield and Mules, instead of following a specific routine with conclusions that can be made at each step, five statistical measures will be used and the relative performance of each scenario will be evaluated. The justification for this break is that the procedure detailed by Butterfield and Mules is an iterative process and designed to either accept two tables as identical or reject it based upon the outcome of any given test. As such, the process may lead to the conclusion

that the tables are not the same which while it may be correct, does not tell us detailed information about how and why the tables differ.

Each of the tests employed in the testing routine in this paper assesses a different criteria, thus no single test is given greater weight relative to another test. Therefore, the scenario's accuracy will be gauged based on the overall performance of each non-survey scenario relative to the survey-based scenario with the *a priori* decision criteria that the scenario with the highest performance on at least three or more of the five tests will be deemed to be closest to the survey-based scenario.

The first two tests to be employed, the Wilcoxon signed-rank test and a simple regression both have significance tests to test hypotheses. This is not the case with the remaining tests. The third testing routine measures the distance from the mean employing the standardized mean absolute difference, the mean absolute difference, and the mean similarity index. The fourth routine is Theil's inequality coefficient, U, and its composite measures which will gauge overall scenario accuracy and if inaccuracy exists, where it originates from. Finally, the information theory approach of Czamanski and Malizia (1969) will be utilized to judge overall goodness of fit. It should be noted that we are considering the direct survey-based coefficients to be the benchmark and the non-survey-based coefficients to be the estimates.

First, a Wilcoxon signed-rank test is performed to see if there is consistent over- or under-estimation of the coefficients. Here, the null hypothesis is that the set of

estimated coefficients is not significantly above or below the set of coefficients derived from the survey. If the null hypothesis is rejected, we can conclude that the estimate is poor. Round (1983) is highly critical of this step as the Wilcoxon requires independence of the variables, which in the case of the input coefficients is not possible as they are drawn from a proportion. While this removes some of the credibility, when combined with the results of the measures of distance, which can tell us if any skewness occurs, and in which direction the results are skewed. The null hypothesis and alternative hypothesis are as follows:

H_0: $t_{ij} = t_{ij}^*$

H_A: $t_{ij} \neq t_{ij}^*$

where: t_{ij} = the survey input coefficient for industry i's purchase from sector j

t_{ij}^* = the estimated input coefficient for industry i's purchase from sector j

Because of the lack of independence, additional testing is necessary. Butterfield and Mules (1980) suggest a simple regression of the form:

$$t_{ij}^* = \alpha + \beta t_{ij} \tag{4.4}$$

where: α = the intercept term

If α = 0 and β = 1, we can conclude that the estimated coefficient is a fair approximation of the benchmark coefficient. However, if α > 0 and β < 1 or α < 0 and β

> 1, we need to test further to draw any conclusions about the accuracy of the coefficients.

Round (1983: 202) is critical of this approach for three reasons: 1) The impact that zero and near-zero coefficients will have in weighting; 2) the lack of independence in the technical coefficients; and 3) "the spurious association that might be construed if a comparison is made between regional input-output tables rather than matrices of trade or technical coefficients alone." A work-around to the first criticism is to simply drop the coefficients if both the benchmark and estimated coefficients are zero much like Morrison and Smith (1974) did with zero values when employing the standardized mean absolute difference, a step employed in this paper. The zeros in this case merely indicate that this sector is not part of the production function, so there is no reason to include that sector. The second criticism is valid, but unfortunately, nothing can be done to give independence to the technical coefficients. The third point only applies in situations in which tables are being tested, not in instances of testing actual (survey) coefficients against estimated coefficients.

The third step in our routine is testing using measures of distance from the mean. Here, three tests will be employed: 1) standardized mean absolute difference (SMAD); 2) mean absolute difference (MAD); and 3) mean similarity index (S_{ij}). The first two tests can tell us about the skewness of our errors (Butterfield and Mules, 1980).

The third test, the mean similarity index developed by Isard and Romanoff (1968, cited in Morrison and Smith, 1974) measures the magnitudes of relative changes.

The standardized mean absolute difference (SMAD) holds the chief advantage that the errors do not cancel each other out. Also, it shows the magnitude of overall errors without penalizing large errors. However, it does not tell the direction of the errors (Butterfield and Mules, 1980). Fortunately, the Wilcoxon can tell us the overall direction of these errors. This test standardizes the error relative to the size of the a_{ij}. (Morrison and Smith, 1974). The major drawback is that the test cannot accommodate situations in which t_{ij} is equal to zero. Such instances were not included in the calculated values reported in the following chapter. However, Harrigan, McGilvray, and McNicoll (1980) propose omitting any results in which t_{ij} or t_{ij}^* are zero. For the SMAD, smaller values are more desirable than larger values. The formula for the SMAD is:

$$SMAD_j = \frac{1}{n}\sum_{i=n}^{n}\left|\frac{t_{ij}^* - t_{ij}}{t_{ij}}\right| \tag{4.5}$$

The second test, the mean absolute difference provides some of the same benefits as the SMAD such as the errors not cancelling each other out and not penalizing large errors. It complements the SMAD in that it has the added advantage of being able to assess the impact of zeros in the t_{ij}'s. Also, it keeps errors in their original units and those errors do not cancel each other out. Like the SMAD, the MAD cannot tell the direction of errors (Butterfield and Mules, 1980). The formula for MAD is as follows:

$$MAD_j = \frac{1}{n}\sum_{i=n}^{n}\left|t_{ij}^* - t_{ij}\right| \tag{4.6}$$

The third measure of distance test is the mean similarity index. The measure was first developed by Isard and Romanoff (1968, cited in Morrison and Smith, 1974). It takes the form:

$$S_{ij} = 1 - (t_{ij}^{*} - t_{ij})/(t_{ij}^{*} + t_{ij}) \tag{4.7}$$

The S_{ij} is equal to unity when $t_{ij}^{*} = t_{ij}$ and is equal to zero when one is zero and the other is non-zero. The closer the mean of this measure is to unity, the closer the tables fit each other.

The fourth stage of the testing routine is Theil's inequality coefficient U. U takes the form:

$$U = \sqrt{\frac{\sum\left(t_{ij}^{*} - t_{ij}\right)^{2}}{\sum {t_{ij}}^{2}}} \tag{4.8}$$

U can assume a value of 0 if and only if the predicted values and the actual values are identical. When its value is unity it means that the estimated values are no better than naïve forecasts of no-change basis. Further, it should be noted that there is no upper bound on the value, which Theil states "is tantamount to saying that it is possible to do considerably worse than by extrapolating on a no-change basis" (Theil, 1967: 28). Theil's U is comprised of three inequality proportions: 1) the bias proportion, U^{M}, 2) the variance proportion, U^{S}, and 3) the covariance proportion, U^{C}. Together these sum to unity, that is, $U^{M} + U^{S} + U^{M} = 1$.

$$U^M = \frac{\left(\overline{t_{ij}^*} - \overline{t_{ij}}\right)^2}{\frac{1}{n}\sum\left(t_{ij}^* - t_{ij}\right)^2} \tag{4.9}$$

$$U^S = \frac{(S_{t*} - S_t)^2}{\frac{1}{n}\sum\left(t_{ij}^* - t_{ij}\right)^2} \tag{4.10}$$

$$U^C = \frac{2(1-r)S_{t*}S_t}{\frac{1}{n}\sum\left(t_{ij}^* - t_{ij}\right)^2} \tag{4.11}$$

where: $\overline{t_{ij}^*}$ = the mean of the t_{ij}^* = $1/n \sum t_{ij}^*$

$\overline{t_{ij}}$ = the mean of the t_{ij} = $1/n \sum t_{ij}$

S_{t*} = standard deviation of t_{ij}^* = $\sqrt{1/n \sum\left(t_{ij}^* - \overline{t_{ij}^*}\right)^2}$

S_t = standard deviation of t_{ij} = $\sqrt{1/n \sum\left(t_{ij} - \overline{t_{ij}}\right)^2}$

r = the correlation coefficient = $\frac{\frac{1}{n}\sum\left(t_{ij}^* - \overline{t_{ij}^*}\right)\left(t_{ij} - \overline{t_{ij}}\right)}{S_{t*}S_t}$

These components can provide more information than U. To see this, consider the implications of each. Because they sum to unity, each must take a value between zero and one. If the bias proportion, U^M, is large this means there is substantial differences between the estimated and survey-based coefficients, obviously the most critical issue under examination. The variance proportion, U^S, measures how much of the inequality is accounted for by higher or lower variances. Finally, the covariance proportion, U^C, is in effect a residual and Theil labels efforts to minimize it as "hopeless" as such endeavors are outside of the capabilities of the researcher (Theil, 1967). Given that the desired outcome on the other two proportions are low numbers then a good

estimate would have a larger value for the covariance proportion relative to the other proportions. Finally, caution must be exercised:

> It must be emphasized that the values of the components of Theil's U in isolation do not indicate good or bad estimates. These values must only be evaluated in conjunction with Theil's U, because the components merely decompose Theil's U and within that context alone provide some insights into the variations due to bias of estimation and the variance and covariance of U (Jalili, 2000: 231).

The fifth and final stage of the testing routine borrows from information theory. The information index provides the same information as the Chi-square test, but does not require independence between the variables. This technique cannot accommodate situations in which t_{ij} is zero and t_{ij}^* is non-zero. There exists the possibility of omitting such elements (Morrison and Smith, 1974), a step which was employed in this study. This equation takes the form:

$$I(\mathbf{T}:\mathbf{T}^*) = \sum_i \sum_j \left| t_{ij}^* \log_2 \frac{t_{ij}^*}{t_{ij}} \right| \qquad (4.12)^4$$

where: **T** = the column vector of benchmark input coefficient

T* = the column vector of estimated input coefficients

The closer t_{ij} is to t_{ij}^* for each i and j, the lower the information value and thus, the more accurate the estimate is to the survey value. If and only if both coefficients are equal for all i and j then $\log_2 t_{ij}^*/t_{ij}$ equals 0; conversely, high values of I are symptoms of

[4] Here, I indicates a functional relationship and the colon operator (as in the original text) between **T** and **T*** indicates A* given A or as is more commonly noted **A***|**A**.

extensive differences (Hewings, 1971). As such, lower values of I indicate a better goodness of fit.

Multiplier Testing

Once the comparison of coefficients has been completed and the shocks run, the multipliers can be compared. Unfortunately, a review of the literature reveals that the most prevalent method of comparison is eyeball comparisons lacking in statistical rigor (e.g. Brucker, Hastings, and Latham III, 1990; Taylor and Fletcher, 1993; Oosterhaven, van der Knijff, and Eding, 2003; Bonn and Harrington, 2008). Much like the comparison of the coefficients, no single test can gauge the closeness of the multipliers. As such, a number of treatments will be employed. The multipliers to be tested are the output, employment, and total value-added multipliers. Further, because the direct impact only appears in one sector, a comprehensive test of the individual sector's multipliers is not possible as it would require division by zero. As such, we will be testing the total effects as a proxy for examining the individual sectoral multipliers.

The first test is a paired t test to see if the non-survey multipliers differ significantly from the survey-based multipliers. The second test is the Spearman rank correlation coefficient, which measures statistical correlation between the two scenarios. The final test of the multipliers is the mean absolute percentage error which will measure the overall size of the errors between the scenarios. Like the evaluation of the input coefficients, the scenarios will be evaluated based on the overall performance

of the scenarios' multipliers for all of the tests with no single test being given greater weight. Again, we will set an *a priori* decision criteria, this time if any scenario performs best on two of the tests, it will be determined to have performed best relative to the survey-based scenario.

The first such treatment will be a paired t test. The paired t test has been used to compare multipliers in Rickman and Schwer (1995a; 1995b). This test will allow us to compare the survey-based multipliers with the estimated multipliers and compare them with the student's t statistic for various confidence intervals. The null hypothesis is that the difference between the two sets of multipliers is zero, with the alternative hypothesis being otherwise. The formula for this test is:

$$t = \frac{\bar{D}}{S/\sqrt{n}} \tag{4.13}$$

where: $\bar{D} = \frac{1}{n}\sum_{i=1}^{n} D_i$

$D_i = X_i^* - X_i$

X_i^* = the estimated ith total effect

X_i = the survey-based ith total effect

$$S^2 = \frac{\sum_{i=1}^{n}(D_i - \bar{D})^2}{n-1}$$

The second test of multipliers is the Spearman rank correlation coefficient. The Spearman rank correlation coefficient measures statistical correlation between two

variables. It has been used to compare multipliers in Rickman and Schwer (1995a; 1995b) and Willis (1997). The operational form is:

$$\rho = 1 - \frac{6\sum d^2}{n(n^2-1)} \tag{4.14}$$

where: d = the difference in the rank of the ith multiplier

The interpretation of the Spearman rank correlation coefficient is as follows: if ρ equals zero than there is no linear correlation. If ρ is close to one, it indicates positive correlation. ρ values close to negative one indicate negative correlation between the survey-based and non-survey multipliers.

The third test of the multipliers is the mean absolute percentage error (MAPE). The MAPE divides the error (or difference between the two multipliers) by the survey-based multiplier and is then averaged over all multipliers. It holds two chief advantages: 1) it does not penalize large errors if the two multipliers differ in level; and 2) errors cannot offset each other. Unfortunately, the direction of errors is not revealed. It carries an implicit assumption that larger multipliers can absorb more error. The reason why this is so is that the same size error represents a smaller percentage of the total multiplier as the size of the multiplier increases. The formula is:

$$MAPE = \frac{\left|\sum_{i=1}^{n}(X_i - X_i^*)/X_i\right|}{n} \tag{4.15}$$

The MAPE does not have any fixed interpretation other than the value a researcher decides is an acceptable level of error. Because of the lack of an objective benchmark, the relative performance of each scenario will be judged on its closeness to zero.

V. RESULTS

Scenario Results

Before we proceed with a discussion of the results a discussion of the impacts and multipliers of the scenarios is in order. As can be seen in table 5.1, the Modified Wet Corn Milling scenario most closely matched the survey effects both in multipliers and total effects. The two scenarios based on other basic organic chemical manufacturing, the Base and No Agriculture scenarios both overestimated the indirect effects. The Base and No Agriculture scenarios overestimated this by a magnitude of 4 times, whereas Modified Wet Corn Milling overestimated it by a magnitude of 2.5 times. While the employment multipliers of all of the scenarios are significantly higher than those reported in most of the studies presented in chapter 3, it should be noted that the current scenarios are for state-wide impacts and not county- or multi-county-level impacts.

Table 5.1: Scenario Employment Effects and Multipliers

Scenario	Direct	Indirect	Induced	Total	Multiplier
Survey	31.2	57.2	168.4	256.9	8.2
Base	49.0	225.1	172.1	446.2	9.1
No Agriculture	49.0	222.9	171.7	443.6	9.1
Modified Wet Corn Milling	36.0	117.2	143.5	296.6	8.2

Examining the output multipliers, we note that both the Base and No Agriculture scenarios had the highest output multipliers, 1.640 and 1.637, respectively. The

Modified Wet Corn Milling scenario's larger output multiplier compared to Swenson's previous studies (2006, 2007, 2008) is likely the result of these scenarios being run for state-wide impacts as opposed to county-level impacts in his previous studies. While the scenarios all produced relatively similar results in their induced effects, there were large differences in the indirect column. The Base and No Agriculture scenarios overestimated the indirect effects by over 3 times while the Modified Wet Corn Milling scenario was just 3 times. This variation is likely the result of the large differences in the overall sum of the input coefficients. These results are summarized below in table 5.2.

Table 5.2: Scenario Output Effects and Multipliers

Scenario	Direct	Indirect	Induced	Total	Multiplier
Survey	$100,000,000	$12,241,117	$18,474,305	$130,715,422	1.307
Base	$100,000,000	$45,364,704	$18,630,506	$163,995,210	1.640
No Agriculture	$100,000,000	$45,101,534	$18,581,305	$163,682,839	1.637
Modified Wet Corn Milling	$100,000,000	$33,030,265	$15,517,675	$148,547,940	1.485

The results of the total value-added multiplier are reported below in Table 5.3. Total value-added represents the increase in gross domestic (or regional) product in a region from the operation of the biofuels facility. Once again, the Base and No Agriculture scenarios over-estimated the multiplier, more than doubling the survey-derived multiplier. The Modified Wet Corn Milling scenario's multiplier was 1.5 times larger than the survey-based multiplier (and higher than Swenson's previous studies, which were county-based, (2006, 2007, 2008).

Table 5.3: Scenario Total Value-Added Effects and Multipliers

Scenario	Direct	Indirect	Induced	Total	Multiplier
Survey	$15,302,548	$4,934,600	$10,483,067	$30,720,215	2.008
Base	$8,885,007	$19,453,380	$10,554,535	$38,892,922	4.377
No Agriculture	$8,885,007	$19,340,266	$10,526,664	$38,751,937	4.361
Modified Wet Corn Milling	$9,063,231	$14,973,132	$8,790,707	$32,827,070	3.622

Overall, the largest differences between the non-survey scenarios and the survey scenario were in the indirect effects. The inclusion of larger total input coefficients in the non-survey scenarios led to larger amounts of inputs purchased by suppliers in the region. Both of the other basic organic chemical manufacturing-based scenarios consistently reported higher multipliers than the survey-based scenario. The Modified Wet Corn Milling scenario reported higher multipliers for output and total value-added than the survey, but its employment multiplier was in line with the survey.

Input Coefficient Testing Results[5]

Wilcoxon Signed-Rank Test

As stated in the previous chapter, the first test of the input coefficients is a Wilcoxon signed-rank test to see if there is consistent over- or under-estimation of the input coefficients. The calculated Z-scores for the test were as follows: 7.044 and 6.88 for the Base and the No Agriculture scenarios, respectively. The calculated Z-scores indicate consistent overestimation of the input coefficients at all significance levels for

[5] For the purposes of testing, all instances where both the survey-based input coefficient and the non-survey-based input coefficient were zero were excluded from analysis.

both of these scenarios. Thus, these scenarios are significantly different from the survey. The Modified Wet Corn Milling scenario's Z-score was 1.16, which indicates there is neither consistent over- nor underestimation at the 1% significance level and thus we fail to reject to reject the null hypothesis stated in chapter 4 that the survey-based and the non-survey-based trade coefficient are not significantly different. For convenience, table 5.4 below provides the level of significance for a two-tailed z-test for the 10%, 5%, and 1% levels of significance.

Table 5.4: Level of Significance Table under a Two-Tailed Test

Z	% Significant
1.64	10%
1.96	5%
2.58	1%

Simple Regression

The second stage in the testing regiment is a simple regression. Here a β equal to unity and an α equal to zero indicate that the benchmark coefficients and the estimated coefficients are not statistically different. Thus, an intercept not statistically different from zero and a slope coefficient not statistically different from zero are desirable. The first step in our regression analysis was to drop any instances where both the survey and non-survey coefficients were equal to zero. This step was done to alleviate any concern such as was raised by Round (1983) that zero coefficients could have a significant impact on the weighting.

As indicated in Table 5.5, the alphas of the three scenarios are all quite low, which are the desired results. However, only the Modified Wet Corn Milling scenario's alpha is not statistically different from zero. Looking at the betas of these three scenarios, both the Base and the No Agriculture scenarios are not statistically different from unity at the 95% significance level. However, the null hypothesis that the beta is not statistically different from unity can be rejected for the Modified Wet Corn Milling scenario at the 95% significance level (with a calculated t value of 1.99 just narrowly missing the critical value of 1.96). Finally, it should be noted that the coefficient of determination, R^2, for all three scenarios was quite low indicating poor goodness of fit. For convenience, the critical values of the t-distribution for a two-tailed test at several significance levels is provided below in table 5.6

Overall, the Modified Wet Corn Milling scenario performed the best of these three scenarios. Its intercept was not statistically different from zero at the 95% significance level and its slope coefficient, while not different from unity at the 95% significance level was not statistically different from unity at the 99% significance level. The Base and No Agriculture scenarios both performed relatively equally. While they both had slopes that were not statistically different from unity at the 95% level, neither had intercepts that were statistically different from zero.

Table 5.5: Regression Results

Non-Survey Scenario	α	Std Error	t Value	β	Std Error	t Value	R^2	n
Base	0.001404	0.0004	3.13**	0.779000	0.1553	1.42	0.1198	187
No Agriculture	0.001401	0.0005	3.11**	0.779154	0.1557	1.42	0.1198	186
Modified Wet Corn Milling	0.001359	0.0007	1.88	0.563869	0.2195	1.99*	0.0444	144

*Statistically different at the 5% significance level
**Statistically different at the 1% significance level

Table 5.6: Critical Values of the t-Distribution for a two-tailed test

T	Level of Significance
1.96	5%
2.576	1%

Measures of Distance from the Mean

The third testing routine measures the distance from the mean. Table 5.7 summarizes these results. The standardized mean absolute difference and the mean absolute difference both measure the degree of skew in the errors. The first test of this routine is the standardized mean absolute difference. With this test, smaller values are desirable. Additionally, it should be noted that sectors in which the non-survey input coefficient equaled zero were excluded from analysis. The Modified Wet Corn Milling scenario outperformed the other two scenarios with a standardized mean absolute difference nearly one third that of the other scenarios.

The mean absolute deviation is the second test of the measure of distance. Again, smaller values indicate less skewness. It should be noted that the reported mean MAD values represent less than 0.3 percent of the sum of the input coefficients for each of our three estimates. When this value is compared to the survey-based input coefficients, it is just greater than 1% of the sum of the input coefficients for the Modified Wet Corn Milling scenario and less than 1% for the two Base and No Agriculture scenarios. As indicated by the reported MAD values, the Modified Wet Corn Milling scenario had the most skewness, a result inconsistent with the standardized mean absolute difference. A possible explanation for this is that errors of the same magnitude in the Modified Wet Corn Milling scenario were more heavily penalized due to differences in size of the coefficient. The Base and No Agriculture scenario held similar values, with the Base scenario slightly outperforming the No Agriculture scenario.

Finally, the mean similarity index results indicate that the Modified Wet Corn Milling scenario's approach to estimating input coefficients most closely matched the survey-based coefficients. As discussed above, the closer the value of the mean similarity index is to unity, the closer the two tables fit each other. Here, the No Agriculture scenario outperformed the Base scenario.

In sum, for two of our three measures of distance, the Modified Wet Corn Milling scenario's coefficients outperformed the two other scenarios. The Modified Wet Corn

Milling scenario's SMAD was nearly one third of that calculated using the other scenarios. The Modified Wet Corn Milling scenario's MAD was larger than the MAD of the Base and No Agriculture scenarios. In the MAD, the Base scenario performed closest to the survey-based input coefficients. Finally, the mean similarity index for the Modified Wet Corn Milling scenario indicated that it was the closest to the benchmark input coefficients. In order to rank the performance of each of the scenarios in this test, the evenly-weighted average of the three tests was used as the decision criteria. Despite performing poorest in the mean absolute difference, the Modified Wet Corn Milling scenario outperformed the other scenarios.

Table 5.7: Measures of Distance from the Median

Treatment	**Base**	**No Agriculture**	**Modified Wet Corn Milling**
Standardized Mean Absolute Difference	3555.56	3555.56	1326.33
Mean Absolute Difference	0.00166	0.00168	0.00189
Mean Similarity Index	0.5083	0.5164	0.8557

Theil's Inequality Coefficient

The fourth test of the input coefficients was Theil's inequality coefficient, U. U measures the difference between actual and predicted results. Based on the results reported in Table 5.8, none of our non-survey techniques did a good job of predicting the input coefficients. In fact, all three of the scenarios did worse than a naïve estimate

could have done. The Modified Wet Corn Milling scenario performed the worst, with a U value of 2.6259. The No Agriculture scenario performed the best slightly outperforming the Base scenario with U values of 2.1420 and 2.1424, respectively.

Breaking Theil's U into proportions, we see that the bias proportion for all three scenarios is quite low with the Modified Wet Corn Milling scenario most closely mimicking the survey-based coefficients with a reported value less than one quarter that reported by the other scenarios. As indicated by Theil (1967), this is the most crucial element and positive values indicate errors of central tendency. The Modified Wet Corn Milling scenario holds the highest value for the variance proportion, indicating greater dispersion around the mean relative to the other two non-survey approaches. Lower values in this element indicate closeness of standard deviations. The No Agriculture scenario outperformed the other two scenarios for this proportion. Finally, the third element, the covariance proportion represents both a residual and an error due to incomplete covariation. As Theil argues, from a quadratic loss viewpoint, higher values in this proportion are the least controllable.

Overall, none of the three scenarios performed particularly well. The high values of U indicate high dissimilarity between the benchmark input coefficients and the estimated input coefficients. The lower bias proportion seen in the Modified Wet Corn Milling scenario is desirable, but is offset by the higher variance proportion relative to the other scenarios. The Base and No Agriculture scenarios both demonstrated higher

degrees of systematic error as evidenced by the bias proportions, with the No Agriculture scenario slightly outperforming the Base scenario. Likewise, the No Agriculture scenario performed the best on the variance proportion. Finally, the No Agriculture scenario reported the highest covariance proportion, which of the three proportions is the one that a unity value is preferred (Jalili, 2000). However, it must be stressed that the components only tell us the source of the error and that the overall U is the only basis of comparison (and likewise the element used in our decision criteria).

Table 5.8: Theil's Inequality Coefficient, U

Treatment	**Base**	**No Agriculture**	**Modified Wet Corn Milling**
Inequality Coefficient, U+	2.142416	2.142036	2.625939
Bias Proportion	0.043705	0.043736	0.015953
Variance Proportion	0.332141	0.331973	0.392958
Covariance Proportion	0.624154	0.624291	0.589868

+ Values of zero indicate least overall error; values greater than unity indicate a naïve forecast could perform better than the scenario estimate

Information Index

The fifth and final test of our input coefficients is the goodness of fit utilizing the information index approach. As seen in Table 5.9, both of the Base and No Agriculture scenarios outperform the Modified Wet Corn Milling scenario for goodness of fit. As indicated in the previous chapter, lower values of I indicates a better goodness of fit (with a value of zero indicating a perfect fit). The poor performance of the three

scenarios in our goodness of fit test should not come as a surprise, given the low R^2 as revealed by the regression tests.

Table 5.9: Goodness of Fit

Treatment	Base	No Agriculture	Modified Wet Corn Milling
Information Index	1.50235	1.50235	1.71145

As we can see from our different tests, none of the scenarios do a satisfactory job of estimating the input coefficients found in our survey scenario. Overall, the Modified Wet Corn Milling scenario performed best, ranking first in three of the five tests employed. The Wet Corn Milling scenario performed the best in the Wilcoxon, the simple regressions, and two of the three measures of distance from the mean. However, it performed quite poorly in Theil's U and the goodness of fit measure. The second best performing scenario was the No Organics scenario which outperformed the Base scenario on nearly every test. In summation, the Modified Wet Corn Milling scenario met the *a priori* decision criteria of performing best in three of the measures. These rankings of the scenarios' performance are provided below in table 5.10.

Table 5.10: Composite Scenario Rankings

Treatment	Base	No Agriculture	Modified Wet Corn Milling
Wilcoxon Signed-Rank Test	3	2	**1**
Simple Regression	2	3	**1**
Measures of Distance from the Mean	2.5*	2.5*	**1**
Theil's U	2	**1**	3
Information Theory	**1.5***	**1.5***	3

*Tied Rank

However, as noted by Jensen (1980), the partitive accuracy of our scenario is not the chief concern. Ultimately, policymakers and many researchers are more concerned with the results these scenarios generate. As such, despite differences between the estimated and survey-based coefficients, we must also investigate how this affects the multipliers.

Multiplier Testing Results[6]

However, it should be noted that in order to correct for the large differences in the direct effects between the wet corn milling sectors and the all other basic organic chemical manufacturing sectors an additional step was needed. This is because when a shock is applied to a scenario, it is directly applied to a given sector. For the survey-based scenario and the other basic organic chemical manufacturing-based scenarios (Base and No Agriculture), this was applied to the all other basic organic chemical manufacturing sector, while for the Modified Wet Corn Milling scenario, this direct

[6] Instances where both the survey-based and the non-survey based total effects were equal to zero were dropped from the analysis.

effect was applied to the wet corn milling sector. As such, to correct for this difference, the direct effects that would have been applied to the wet corn milling sector for the Modified Wet Corn Milling scenario were instead added to the other basic organic chemicals sector so that the comparisons could be done on even footing. This modification was applied for all of the multiplier tests.

Paired t Test

The first test of the multipliers is a paired t test. As a reminder, in lieu of using the actual multiplier, the sector-by-sector total effects were used as a proxy due to issues with division by zero. As can be seen, both the Base and No Agriculture scenarios lead us to conclude that we can safely reject the null hypothesis that the total effects are equal for the employment and the output effects. However, we must fail to reject the null hypothesis at the 95% significance level that the total value-added effects are the same. The Modified Wet Corn Milling scenario's calculated t-score leads us to fail to reject the null hypothesis at a 95% significance level for the employment, output, and total value-added effects, that is, the effects for these three are not statistically different. The calculated t-scores for these scenarios are available below in Table 5.11.

Table 5.11: Paired t Test

Non-Survey Approach	Employment	Output	Total Value-Added
Base	4.04**	4.21**	1.08
No Agriculture	3.99**	4.18**	1.07
Modified Wet Corn Milling	1.51	1.78	0.27

**Statistically different at the 1% significance level

Spearman's Rank Correlation Coefficient

The second test of the multipliers is Spearman's rank correlation coefficient. The results of this test are summarized below in Table 5.12. All of the effects demonstrated strong positive correlation. The strongest positive correlation in the employment effects was in the Modified Wet Corn Milling scenario. The second strongest positive correlation for employment effects was found in the No Agriculture scenario. The weakest positive correlation was found in the Base scenario, but this was not a large difference.

When looking at the output effects, the Modified Wet Corn Milling scenario showing the strongest positive correlation, seconded only by the Base scenario. The No Agriculture scenario the weakest positive correlation, but again, it was not a substantial difference.

Finally, examining, the total value-added effects, the strongest positive correlation was found in the Modified Wet Corn Milling scenario. Like the output effects, the Base scenario demonstrated the second strongest positive correlation. The weakest positive correlation was again found in the No Agriculture scenario.

However, all of the correlation coefficients are so close it is extremely difficult to definitively state that one scenario outperformed another scenario. A possible explanation for the similarity between the scenarios may be that the induced effects

resulting from the employee spending all follow similar patterns and this may be causing the multipliers to be moving with one another.

Table 5.12: Spearman's Rank Correlation Coefficient

Non-Survey Approach	Employment	Output	Total Value-Added
Base	0.959332	0.953889	0.959015
No Agriculture	0.959374	0.953692	0.958781
Modified Wet Corn Milling	0.961265	0.954112	0.959285

Mean Absolute Percentage Error

The third and final test of the multipliers is the mean absolute percentage error. Here lower values are preferable to higher values. As can be seen in table 5.13 below, the smallest errors occurred in the Modified Wet Corn Milling scenario with values less than half those found in the other scenarios. The Base scenario and the No Agriculture scenarios performed similarly, with the No Agriculture scenario slightly outperforming the Base scenario.

Table 5.13: Mean Absolute Percentage Error

Non-Survey Approach	Employment	Output	Total Value-Added
Base	2.7327	2.7247	2.7236
No Agriculture	2.6404	2.6326	2.6316
Modified Wet Corn Milling	0.9939	1.2058	1.1875

Using our *a priori* benchmark laid out in the previous chapter, the Modified Wet Corn Milling scenario is judged to have performed best having finished with the best performance in two of the three tests relative to the other scenarios. As can be seen by an examination of the various statistical techniques employed on the total effects, the Modified Wet Corn Milling scenario consistently outperformed the Base scenario and the No Agriculture scenario. The Modified Wet Corn Millings scenario's employment and output effects were not statistically significantly different from the survey scenario at the 95% level based on the paired t test, while neither of the other scenarios was statistically close. All of the scenarios performed strongly in the Spearman rank correlation coefficient test, but the Modified Wet Corn Milling scenario performed best (albeit relatively indistinguishable in magnitude). Finally, the Modified Wet Corn Milling scenario performed best using the mean absolute percentage error test. These results are summarized in table 5.14.

Table 5.14: Composite Scenario Rankings

Treatment	**IMPLAN – Base**	**IMPLAN - No Organics**	**Wet Corn Milling**
Paired t Test	3	2	**1**
Spearman Rank Correlation Coefficient	**2***	**2***	**2***
Mean Absolute Percentage Error	3	2	**1**

* Tied Rank

VI. CONCLUSIONS

A preponderance of the evidence reveals that the two other basic organic chemical manufacturing scenarios, both the dropping of the corn and soy and the dropping of all agricultural and forestry inputs, produce higher multipliers than the survey-based scenario. This increase was most noticeable in the indirect effects, which is nearly eight times larger in the employment and four times larger in the output in both scenarios relative to the survey scenario. The Modified Wet Corn Milling scenario, while it performs better, still exaggerates the indirect effects, nearly doubling the projection for employment and nearly tripling the projection for output compared with the survey scenario. All of the non-survey scenarios overstate the output indirect effects relative to the survey scenario, with the Base and No Agriculture scenarios nearly quintupling them and the Modified Wet Corn Milling scenario inflating them by three-and-a-half times.

The indirect effects of the scenarios represent the purchases of the firms and their suppliers. The consistent overestimation of coefficients as indicated by the Wilcoxon signed-rank test would predict this overestimation of the indirect effects as the amount of first-round purchases by the firm would be too large. Based on the Wilcoxon signed-rank test, only the Modified Wet Corn Milling scenario is able to closely represent the survey-based coefficients. Not surprisingly, the indirect effects of the Modified Wet Corn Milling scenario are the closest to those of the survey.

The Modified Wet Corn Milling scenario performs best in the regression with the alpha not being statistically different from zero at the 95% level and the beta not being statistically different from unity at the 99% level. The other scenarios perform poorly in our regression test. All three scenarios demonstrate poor goodness of fit as evident by the R^2 as well as the information index test. The Modified Wet Corn Milling approach most closely models the survey-based coefficients based on the three measures of difference. These two tests along with the measures of difference from the mean, which test for skewness, determine if the direct purchases in the non-survey scenarios will match the appropriate sectors in to the survey scenario and their full effects will be felt through the individual sectoral multiplier effect.

When we examine our input coefficients using Theil's inequality index, U and its three component measures, we see the severity of the differences of the input coefficients. The calculated U scores all greater than unity, indicate that all three of the non-survey techniques performed more poorly than a naïve estimate extrapolated on a no-change basis could have. While there are not large errors due to central tendency, there are large errors in the variances and standard deviations.

As has been consistently stressed throughout this paper, ultimately policymakers are most concerned about model performance and not accuracy in the coefficients matrices. As previously described, the paired t-test measures how close the non-survey multipliers are to the survey-based multipliers. Both the Base and No Agriculture

methods fail the paired t-test for employment and output effects, but are not statistically different from the survey scenario multipliers for total value-added effects. On the other hand, the Modified Wet Corn Milling approach is not statistically different than the survey-based multipliers for each of the three effects.

All three scenarios show strong positive correlation with the survey-based scenario as evidenced by the high value calculated in the Spearman rank correlation coefficient. The importance of the Spearman rank correlation coefficient is that it tells if large values in individual sectors in the survey-based scenario are associated with large values in the sectors of the non-survey-based scenarios. This is important for researchers trying to present information to policymakers on which sectors will be most greatly affected by the operations of a biofuels plant. The strong positive correlation found in each scenario tells us that estimates of which sectors will be most greatly impacted will be reasonably accurate.

Looking at our mean absolute percentage error, the Modified Wet Corn Milling scenario performs closest to the survey scenario. Again, the MAPE measures overall percentage of accuracy. The results indicate that the Modified Wet Corn Milling scenario will be approximately one percent different than the survey-based scenario on average. The Base and No Agriculture scenarios will have absolute errors just greater than 2.5%.

The reason why the Modified Wet Corn Milling scenario scenario performs closest to the survey likely results from a number of factors. First, Swenson's (2006) approach is based on trying to mimic the actual operations of an ethanol plant, while the other two approaches are the average of multiple industries. Second, and perhaps most importantly, the percentage of purchases of corn and soy in the Wet Corn Milling approach most closely match the percentage of the survey-based scenario. In both of the other basic organic chemical manufacturing scenarios, corn, soy, and even all of the agricultural purchases only represent a very tiny portion of the firm's total purchases and this affects the input coefficients.

There are several reasons why the Modified Wet Corn Milling scenario based on Swenson's (2006) methodology may have differed from the survey-based scenario: 1) Swenson's scenario only tries to approximate an ethanol facility, while the survey-based scenario tries to approximate biofuels plants, which include both corn ethanol and soy biodiesel; 2) the adjustments to decrease the multipliers of the declining cost industries, which may be correct, were not reflected in the survey-based scenario; and 3) there may exist some differences in the exact methods carried out by Swenson that may have been unintentionally neglected by this researcher. Overall, the Modified Wet Corn Milling approach advocated by Swenson (2006,) is the most promising alternative to direct survey data.

These results suggest that in the face of technological change, non-survey techniques are not accurate. As such, researchers concerned with the accuracy of their prognostications may want to consider direct surveys. These surveys can provide detailed information on the purchases of the firm, which will result in a greater degree of accuracy in the researcher's forecasts. A higher degree of accuracy in these forecasts can help ensure that the decisions made by policymakers are made with the best information possible.

Limitations

This current study is not without its limitations. First, we have no way of knowing any potential errors in the survey data. Second, firms do not keep data in a format that fits input-output analysis. These first two problems are data collection issues that are always uncertain when dealing with a survey. Third, the lack of a generally accepted testing routine provides the uncertainty that there may be some information our tests failed to capture. Finally, our survey targeted new biofuels plants in Missouri. As such, there may be differences in the production process of older plants or plants that have recently expanded.

The study showed that there are substantial differences between pre-packaged data and direct survey data in the biofuels industry due to technological change (as evidenced by the economies of scale) and the time lag that is omnipresent between when data is collected and packaged for researcher use. Unfortunately, this study does

not address the issue of whether packaged data are sufficient in the absence of technological change due to the scope of the survey.

Further, this study only examined the issue of technological change in one industry and as such, these results may not be generalizable for other industries.

Implications

This study highlights differences between the non-survey techniques discussed above both in terms of the methodologies and the resulting values which could in turn, be reported to policymakers. It draws attention to potential inaccuracies in reported values relative to the survey data and which elements of a forecast are most likely to be overstated. This study provides guidance as to the margin of error (as reported by the MAPE) that a researcher can expect between his or her prognostications and reality.

The most critical implication is that any attempts to estimate the local impacts (be it at a state-wide or at a county-level) of new biofuels plant only be done through surveys or by speaking with industry experts. Detailed cost information (or even estimates) about these plants can reveal a great deal of information. Unfortunately, the default options available in IMPLAN do not allow reasonable approximations of the actual purchases of a biofuels plant due to the differences in primary data and standard secondary data.

Further, biofuels production is characterized by economies-of-scale and therefore as the productive capacity of plants continues to increase, the estimates derived from non-survey approaches will have the potential to become less accurate and the predictions offered by researchers can become increasingly distant from reality. As such, despite the cost or the lack of willingness of would-be or current operators to provide information, there is evidence that in an industry rife with technological change, prepackaged datasets can provide misleading results and researchers may want to conduct surveys in order to best model the impacts.

.

APPENDIX A: THE SURVEY

Description of the Business

1. Name and description of operation:

2. What are the major products/services that will be/are offered by this operation? (add more rows if needed)

1)	
2)	
3)	
4)	
5)	

3. In what year a)will/did operations begin and b) will/did the plant begin operating at full capacity?

a)	
b)	

4. What new operations that buy your product(s) have opened or expanded in the vicinity of this operation? Please describe.

1)	
2)	
3)	

5. Will/did your operation require any special investments or services by local or state government (access road construction, extra road maintenance, new water or wastewater connections, unusual amounts of solid or hazardous waste, special fire or hazardous materials services, security, etc.)?

1)	
2)	
3)	
4)	

6. Will/did your operation make any special investments that provide benefits to the public (gas lines that other businesses can/will use, public road improvments, etc.)?

1)	
2)	
3)	
4)	

7. What percentage of your owners/members reside in Missouri?

8. What was the initial (start-up) cost for this project?

Of this initial cost how much was

1)	**Borrowed:**	
2)	**Invested by Owners/Members:**	
3)	**Grants:**	

9. Full-time employment.

How many people does your firm employ full-time?

10. Part-time and seasonal employment.

(You may choose to report this as numbers in a) and b) or as full-time equivalents in c))

a) How many people do you employ part-time and/or seasonally?

b) Please estimate the total annual hours of your part-time and seasonal employees.

c) What are the full-time equivalents of your part-time and seasonal employees?

11). On the next page, we ask you to report your firm's income and expenses for 2007. In your judgement, was there anything unusual about 2007 (for example, it was a start-up year, there was a disaster, etc.)?

Please click on the tab labeled "Annual Revenue & Expenses" below to continue.

Please report the following information for 2007. Add additional rows if necessary.
Rows 25, 55, 63, 103, 105, 107 and 113 contain formulas.

Operating Income and Expenses for 2007	Dollar Amount	Estimated Percent in Missouri
Sales By Product (or product category) (List each major product that is more than 5% of total sales)		
1)		
2)		
3)		
4)		
5)		
6)		
7)		
8)		
9)		
10) Miscellaneous (sum all others)		
Other Income (retail, custom work, etc.)		
1)		
2)		
3)		
Total Revenue	$0	
MATERIALS EXPENSE (COSTS OF GOODS SOLD)		
Commodities/Raw Materials Examples: corn, cattle, timber, etc. (Specify if more than 5% of total costs)		
1)		
2)		
3)		
4)		
5)		
6)		
7)		

8)			
9)			
10)	Misc (add all the remaining)		

Supplies and Other Expenses

Examples: chemicals, packaging materials, etc. (Specify if more than 5% of total costs)

1)			
2)			
3)			
4)			
5)			
6)			
7)			
8)			
9)			
10)	Purchases for re-sale		
11)	Miscellaneous (add all the remaining)		

Total Materials Expenses	$0

OPERATING EXPENSES

Employment Expenses

Salaries & wages (Including withholding)		
Employee benefits (health care, retirement, etc.)		
Subtotal of wages, salaries and benefits paid	$0	
State personal income tax withholding for employees		
Federal personal income tax withholding for employees		0%
Employee insurance (Worker's compensation)		

General and Administrative Expenses

Communication Expenses (Telephone, Internet, IT, DSL, cable modem, etc)		

	Utilities (fuel, natural gas, electricity, water, sewer, garbage/trash)		
	Custodial/cleaning services		
	Repairs and maintanence		
	Plant and office supplies (including postage, etc.)		
	Business and professional services (legal, management, accounting, payroll, brokerage fees and commissions, membership fees, and other hired services)		
	Machine hire		
	Meetings and travel (include travel to trade shows)		
	Business insurance (comprehensive)		
	Depreciation		
	Bad debt		
	Interest paid (inventory, mortgage, bank and credit card fees, other)		
	Auto and truck expense		
	Freight expense		
	Storage		
	Property taxes		
	State (franchise, corporate income) tax		
	Social Security and Medicare taxes (employer portion)		
	Federal corporate income tax		
	Other taxes and licenses		
	Other Expenses (please specify - add rows if necessary)		
1)			
2)			
3)			
4)			
5)			
6)			
7)			
8)			
9)	Miscellaneous (add all the remaining)		

Total Operating Expenses	$0
Total Expenses	$0

Net Profits	$0	
Payments to investors		
Retained earnings		
Other (please describe below):		
Check (should approximately equal zero)	$0	

APPENDIX B: IMPUTATIONS

This appendix shows the survey categories that were aggregated for the convenience of the respondents and the corresponding IMPLAN categories into which they were disaggregated.

Survey Category: Catalysts, methanol, enzymes, chemicals, denaturants, etc.

Sector Number	Sector Name
125	All other basic inorganic chemical manufacturing
126	Other basic organic chemical manufacturing

Survey Category: Utilities

Sector Number	Sector Name
31	Electric power generation, transmission, and distribution
32	Natural gas distribution
33	Water, sewage and other systems
390	Waste management and remediation services

Survey Category: Freight

Sector Number	Sector Name
332	Air transportation
333	Rail transportation
334	Water transportation
335	Truck transportation
339	Couriers and messengers*

Survey Category: Auto and Truck Expense

Sector Number	Sector Name
115	Petroleum refineries
150	Tire manufacturing
228	Material handling equipment manufacturing
276	Automobile manufacturing*
277	Light truck and utility vehicle manufacturing*
278	Heavy duty truck manufacturing*
279	Motor vehicle body manufacturing*
280	Truck trailer manufacturing*
283	Motor vehicle parts manufacturing
414	Automotive repair and maintenance, except car washes

Survey Category: Plant and Office Supplies

Sector Number	Sector Name
110	Stationery product manufacturing
113	Printing
210	Vending, commercial, industrial, and office machinery manufacturing*
234	Electronic computer manufacturing*
235	Computer storage device manufacturing*
236	Computer terminals and other computer peripheral equipment manufacturing*
237	Telephone apparatus manufacturing*
300	Wood television, radio, and sewing machine cabinet manufacturing*
301	Office furniture and custom architectural woodwork and millwork manufacturing*
302	Showcase, partition, shelving, and locker manufacturing*
313	Office supplies (except paper) manufacturing
314	Sign manufacturing*
427	Postal service

Note: Sectors not found in the technical coefficients of the other basic organic chemicals sector are denoted with an asterisk. As such, zero of survey firms' spending was allocated to these particular sectors.

REFERENCES

Altman, I. "The Economic Impacts of Doubling the Capacity of the Macon County Ethanol Plant." Community Policy Analysis Center, S-2002-03, University of Missouri – Columbia, October 2002.

Bates, J. and M. Bacharach. Input-Output Relationships 1954:1966 (A Programme For Growth, 3). London: Chapman and Hall, 1963.

Bonn, M. A. and J. Harrington. "A Comparison of Three Economic Impact Models for Applied Hospitality and Tourism Research." Tourism Economics 14 (2008): 769 – 789.

Braschler, C. H. and G. T. Devino. "Non-survey Approach to I/O Modeling." Microcomputer-Based Input-Output Modeling. D.M. Otto and T.G. Johnson ed., pp. 70-79. Boulder CO: Westview Press, 1993.

Brown, D. M. and F. Giarrantani. "Input-Output as a Simple Econometric Model: A Comment." Review of Economics and Statistics 61 (1979): 621 – 623.

Brucker, S. M., S. E. Hastings, and W. R. Latham III. "The Variation of Estimated Impacts from Five Regional Input-Output Models." International Regional Science Review 13 (1990): 119 – 139.

Butterfield, M. and T. Mules. "A Testing Routine for Evaluating Cell by Cell Accuracy in Short-Cut Regional Input-Output Tables." Journal of Regional Science 20 (1980): 293 – 310.

Crihfield, J.B. and H.S. Campbell. "Evaluating Alternative Regional Planning Models." Growth and Change 22 (1991): 1 – 16.

Czamanski, S. and E. E. Malizia. "Applicability and Limitations in the Use of National Input-Output Tables for Regional Studies." Papers and Proceedings of the Regional Science Association 23 (1969): 65 – 78.

De La Torre Ugarte, D. G., B. C. English, and K. Jensen. "Sixty Billion Gallons by 2030: Economic and Agricultural Impacts of Ethanol and Biodiesel Expansion." American Journal of Agricultural Economics 89 (2007): 1290 – 1295.

Eathington, L. and D.A. Swenson. "Dude, Where's My Corn? Constraints on the Location of Ethanol Production in the Corn Belt." Paper presented at Southern Regional Science Association Annual Meeting, Charleston, SC, 29 – 31 March 2007.

Gerking, S. D. "Reconciling "Rows Only" and "Columns Only" Coefficients in an Input-Output Model." International Regional Science Review 1 (1976): 30 – 46.

Harrigan, F. J., J. W. McGilvray, and I. H. McNicoll. "Simulating the Structure of a Regional Economy." Environment and Planning A 12 (1980): 927 – 936.

Hewings, G. J. D. "Regional Input-Output Models in the U.K.: Some Problems and Prospects for the Use of Nonsurvey Techniques." Regional Studies 5 (1971): 11 – 22.

Isard, W. "Some Empirical Results and Problems of Regional Input-Output Analysis." Studies in the Structure of the American Economy. W. Leontief ed., pp. 116-181. New York: Oxford University Press, 1953.

Isard, W. and Romanoff, E. "The Printing and Publishing Industries of Boston SMSA, 1963: And Comparison with the Corresponding Philadelphia Industries." Regional Science Research Institute Technical Paper No. 7, Regional Science Research Institute, Cambridge MA, 1968. As cited in: Morrison, W. I. and P. Smith. "Nonsurvey Input-Output Techniques at the Small Area Level: An Evaluation." Journal of Regional Science 14 (1974): 1 – 14.

Jalili, A. R. "Evaluating Relative Performances of Four Non-survey Techniques of Updating Input-Output Coefficients." Economics of Planning 33 (2000): 221 - 237

Jensen, R. C. "The Concept of Accuracy in Regional Input-Output Models." International Regional Science Review 5 (1980): 139 – 154.

Jensen, R. C. and D. McGaurr. "Reconciliation of Purchases and Sales Estimates in an Input-Output Table." Urban Studies 13 (1976): 59 – 65.

Johnson, T. G. "The Potential Rural Development Impacts of Biodiesel on Virginia's Eastern Shore." Rural Policy Research Institute, Working Paper Series, University of Missouri – Columbia, October 1995.

Jones, L. L. "Input-Output Modeling and Resource Use Projection." Faculty Paper Series 97-10, Dept. of Agr. Econ., Texas A&M University, 1997. <http://purl.umn.edu/23993> Retrieved 4 June 2010.

Jones, L. L., T.L. Sporleder, and G. Mustafa. "Estimation Bias in Regional Input-Output Models Using Secondary Data." Canadian Journal of Agricultural Economics 20 (1972): 10 – 17.

Jones, L. L., T.L. Sporleder, and G. Mustafa. "A Source of Bias in Regional Input-Output Models Estimated from National Coefficients." Annals of Regional Science 7 (1973): 67 – 74.

Kimura, H. "On the Errors of Outputs Due to Errors of Technical Coefficients in Leontief's Open Input-Output Models." Annals of the Institute of Statistical Methods 7 (1955): 205 – 213.

Lahiri, S. "On Reconciling Purchases and Sales Estimates of a Regional Input-Output Table." Socio-Economic Planning Sciences 18 (1984): 337 – 342.

Lazarus, W. F., D. E. Platas, and G. W. Morse. "IMPLAN's Weakest Link: Production Functions or Regional Purchase Coefficients." Journal of Regional Analysis and Policy 32 (2002): 33 – 48.

Leontief, W. "Quantitative Input and Output Relations in the Economic System of the United States." Review of Economics and Statistics 18 (1936): 105 – 125.

Leontief, W. Input-Output Economics, 2nd ed. New York and Oxford: Oxford University Press. 1986.

Lindall, S. D., D. Olson, and G. Alward. "Deriving Multi-Regional Models Using the IMPLAN National Trade Flows Model." Journal of Regional Analysis and Policy 36 (2006): 76 – 83.

Low, S. A. and A. M. Isserman. "Ethanol and the Local Economy: Industry Trends, Location Factors, Economic Impacts, and Risks." Economic Development Quarterly 23 (2009): 71 – 88.

Ma, J. C., J. K. Scott, and T. G. Johnson. "Soy Diesel Processing in Buchanan County, Missouri: Potential Impacts." Paper presented at Annual Meetings of the Liquid Fuels and Industrial Uses for Agriculture, Nashville, TN, September 1996.

Matuszewski, T. I., P. R. Pitts, and J. A. Sawyer. "Linear Programming Estimates of Changes in Input Coefficients." The Canadian Journal of Economics and Political Science 30 (1964): 203 – 210.

McMenamin, D. G. and J. E. Haring. "An Appraisal of Nonsurvey Techniques for Estimating Regional Input-Output Models." Journal of Regional Science 14 (1974): 191 – 205.

Miernyk, W. H. Elements of Input-Output Analysis. New York: Random House, 1967.

Miernyk, W. H. "Comments on Czamanski and Malizia." Papers and Proceedings of the Regional Science Association 23 (1969): 81-82.

Miernyk, W. H. "Comments on Recent Developments in Regional Input-Output Analysis." International Regional Science Review 1 (1976): 47 – 55.

Miller, R. E. and P. D. Blair. Input-Output Analysis: Foundations and Extensions. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1985.

Minnesota IMPLAN Group. IMPLAN Pro Version 2.0 User's Guide. Stillwater, MN: Minnesota IMPLAN Group, Inc., 2000.

Minnesota IMPLAN Group. "Using IMPLAN V3: The IMPLAN V3 Modeling System." Presented at the 2010 IMPLAN Users Conference, St. Louis, MO, 2 – 3 June 2010.

Moore, F. T. and J. W. Petersen. "Regional Analysis: An Interindustry Model of Utah." The Review of Economics and Statistics 37 (1955): 368 – 383.

Morrison, W. I. and P. Smith. "Nonsurvey Input-Output Techniques at the Small Area Level: An Evaluation." Journal of Regional Science 14 (1974): 1 – 14.

Oosterhaven, J., E. C. van der Knijff, and G. J. Eding. "Estimating Interregional Economic Impacts: An Evaluation of Nonsurvey, Semisurvey, and Full-Survey Methods." Environment and Planning A 35 (2003): 5 – 18.

Park, S., M. Mohtadi, and A. A. Kubursi. "Errors in Regional Nonsurvey Input-Output Models: Analytical and Simulation Results." Journal of Regional Science 21 (1981): 321 – 339.

Petersan, D. N. "Estimated Economic Effects for The Nordic Biofuels Ethanol Plant in Revenna, Nebraska." Economic Development Department, Nebraska Public Power District. June 2002. <http://www.ci.ravenna.ne.us/impact.pdf> Retrieved 5 June 2010.

Pierce, V., J. Horner, and R. Milhollin. "Employment and Economic Benefits of Ethanol Production in Missouri." Commercial Agriculture Program, University of Missouri Extension,February 2007. <http://agebb.missouri.edu/commag/ethanolreport2007.pdf> Retrieved 4 April 2010.

Ralston, S.N., S. E. Hastings, and S. M. Brucker. "Improving Regional I-O Models: Evidence against Uniform Regional Purchase Coefficients across Rows." Annals of Regional Science 20 (1986): 65 – 80.

Rickman, D. S. and R. K. Schwer. "A Comparison of the Multipliers of IMPLAN, REMI, and RIMS II: Benchmarking Ready-Made Models for Comparison." Annals of Regional Science 29 (1995a): 363 – 374.

Rickman, D. S. and R. K. Schwer. "Multiplier Comparisons of the IMPLAN and REMI Models Across Versions: Illuminating Black Boxes." Environment and Planning A 27 (1995b): 143 – 151.

Rose, A. "Technological Change and Input-Output Analysis: An Appraisal." Regional Dynamics 2 (1996): 99 – 112.

Round, J. I. "Nonsurvey Techniques: A Critical Review of the Theory and the Evidence." International Regional Science Review 8 (1983): 189 – 212.

Schaeffer, W. A. and K. Chu. "Nonsurvey Techniques for Constructing Regional Interindustry Models." Papers and Proceedings of the Regional Science Association 23 (1969): 83 – 101.

Schaeffer, W. A. "Regional Impact Models." Regional Science Research Institute, West Virginia University, 1999. <http://www.rri.wvu.edu/WebBook/Schaeffer/regionalGT.pdf> Retrieved 14 May 2010.

Smith, P. and W. J. Morrison. Simulating the Urban Economy. London: Pioneer Press. 1974.

Stevens, B. H. and G. A. Trainer. "Error Generation in Regional Input-Output Analysis and its Implications for Non-Survey Models." RSRI Discussion Paper Series, No. 108, Regional Science Research Institute, Philadelphia PA, December 1978.

Stevens, B. H., G. I. Treyz, D. J. Ehrlich, and J. R. Bower. "A New Technique for the Construction of Non-Survey Regional Input-Output Models." International Regional Science Review 8 (1983): 271 – 286.

Stevens, B. H., G. I. Treyz, and M. L. Lahr. "On the Comparative Accuracy of RPC Estimating Techniques." Frontiers of Input-Output Analysis. R. E. Miller, K. R. Polenske, and A. Z. Rose ed., pp. 245 – 257, New York and Oxford: Oxford University Press, 1989.

Su, T. T. "A Note on Regional Input-Output Models." Southern Economic Journal 36 (1970): 325 – 327.

Swanson, M. J., G. W. Morse, and K. I. Westeren. "Regional Purchase Coefficients Estimates from Value-Added Tax Data." Journal of Regional Analysis and Policy 29 (1999): 31 – 50.

Swenson, D. A. "Input-Outrageous: The Economic Impacts of Modern Biofuels Production." Paper presented at the Mid-continent Regional Science Association and the Biennial IMPLAN National Users Conference, Indianapolis, IN, 8 – 10 June 2006.

Swenson, D. A. "Understanding Biofuels Economic Impact Claims." Working Paper, Dept. of Econ., Iowa State University, April 2007. <http://www2.econ.iastate.edu/research/webpapers/paper_12790.pdf> Retrieved 1 July 2010.

Swenson, D. A. "The Economic Impact of Ethanol Production in Iowa." Working Paper, Dept. of Econ., Iowa State University, January 2008. <http://www.econ.iastate.edu/sites/default/files/publications/papers/p11222-2008-01-01.pdf> Retrieved 1 July 2010.

Swenson, D. A. and L. Eathington. "Determining the Regional Economic Values of Ethanol Production in Iowa Considering Different Levels of Local Investment." Working Paper, Dept. of Econ., Iowa State University, July 2006. <http://www.econ.iastate.edu/sites/default/files/publications/papers/paper_12687.pdf > Retrieved 1 July 2010.

Swenson, D. A. "Re: Request for Copy of Spreadsheet used in Input-Outrageous." E-mail to J. D. Rossi. August 17, 2010.

Taylor, D. T. and R. R. Fletcher. "Three Comparisons of Regional Purchase Coefficients Used in Estimating the Economic Impacts of Tourism and Outdoor Recreation." Regional Science Perspectives 23 (1993): 18 – 32.

Theil, H. Applied Economic Forecasting Amsterdam: North-Holland Publishing Company 1967.

Thomassin, P. J. and L. Baker. "Macroeconomic Impact of Establishing a Large-scale Fuel Ethanol Plant on the Canadian Economy." Canadian Journal of Agricultural Economics 48 (2000): 67 – 85.

Tiffany, D. G. and V. R. Eidman. "Factors Associated with Success of Fuel Ethanol Producers." Dept. of App. Econ. Staff Paper Series P03-7, University of Minnesota, August 2003. <http://www.apec.umn.edu/staff/dtiffany/staffpaperp03-7.pdf> Retrieved 26 July 2010.

U.S. Census Bureau. *County-to-County Worker Flow Files*. Washington, DC: United States, Internet Release Date 6 March 2003. <http://www.census.gov/population/www/cen2000/commuting/index.html> Retrieved 25 August 2010.

Willis, K. G. "Spatially Disaggregated Input-Output Tables: An Evaluation and Comparison of Survey and Nonsurvey Results." Environment and Planning A 19 (1997): 107 – 116.