Regional Research Institute
West Virginia University

Technical Document Series

Disaggregating Input-Output Models

Xueting Zhao

RRI TechDoc 2014-03

Disaggregating Input-Output Models

Abstract

This Technical Document describes the steps to disaggregate an industry sector in an I-O table using Python. A disaggregation method that based on weight factors is used in this document. The calculation has three parts. The first part is to calculate the disaggregated input matrix DIM, which represents all the input from common sectors into the new sectors. The second part is to calculate the disaggregated output matrix DOM, which represents all possible output weights of new sectors into common sectors. The last part is to calculate the intra matrix DINM, which represents the allocation of intra-industry sales in the disaggregated sectors. This document states the problem and introduces the calculation method of the paper by Wolsky (1984), then follows with an example based on the data of the paper by Linder et al. (2013).

1   Introduction

Aggregation bias has long been recognized in the context of input-output modeling (Miller and Blair 2009). Jackson and Comer (1993) and Comer and Jackson (1997), for example, have discussed methods for constructing disaggregated national IO accounts using older disaggregated benchmark accounts, and Lahr and Stevens (2002) have discussed aggregation issues in the context of national table regionalization. In the context of environmental input-output modeling, the aggregation takes on additional importance because of the potentially dramatic emissions coefficient differences among the subsectors comprising an aggregate sector. This Technical Document reviews and provides python code for implementing Wolsky's 1984 IO table disaggregation method using the data on which Lindner's (2013) founded his work.

1.1   Problem Statement

Figure 1 shows a schematic example of the original I-O table, which includes the common sectors from S_1 to S_{N+1}. Suppose S_{N+1} is the sector that will be disaggregated.

Three matrix partitions could be defined from Figure 1: *Input matrix IM, Output matrix OM and Intra matrix INM*. Let A be the technical coefficient matrix of the original $(N+1) \times (N+1)$ economy. a_{ij} is the element of the A matrix, for $i, j = 1$ to N.

		COMMON SECTORS						
		S_1	S_2	...	S_N	S_{N+1}	FD	X
COMMON SECTORS	S_1 S_2 ... S_N					IM		
	S_{N+1}		OM			INM	f_{N+1}	x_{N+1}

Figure 1: Schematic example of an original I-O Table

Figure 2 shows a schematic example of the disaggregation I-O table from Figure 1. The sector S_{N+1} is disaggregated into n new sectors from S^*_{N+1} to S^*_{N+n}. The three corresponding disaggregated matrices could be defined from Figure 2: *Disaggregated input matrix DIM, Disaggregated output matrix DOM and Disaggregated intra matrix DINM*. A^* is the technical coefficient matrix that describes the same economy as A with the only difference being that the last sector (sector $N+1$) has been disaggreaged into n distinct sub-sectors. Matrix A^* is of size $(N+n) \times (N+n)$. a^*_{ij} is the element of the A^* matrix, for $i, j = 1$ to $N+n$.

		COMMON SECTORS				NEW SECTORS				
		S_1	S_2	...	S_N	S^*_{N+1}	...	S^*_{N+n}	FD	X
COMMON SECTORS	S_1 S_2 ... S_N						DIM			
NEW SECTORS	S^*_{N+1} ... S^*_{N+n}		DOM				DINM		f^*_{N+1} ... f^*_{N+n}	x^*_{N+1} ... x^*_{N+n}

Figure 2: Schematic example of the disaggregation I-O Table

1.2 Disaggregation Calculation

Following Wolsky (1984), let σ_k be the weight ratio of the total output of the kth new sector to the total output of the disaggregated sector: $\sigma_k = x^*_{N+k} / x_{N+1}$, for $k = 1$ to n. The total output produced by the sector to be disaggregated must be conserved, so $\sum \sigma_k = 1$. The disaggregation estimation can be made for the unknown technical coefficients by assuming that the new sectors have identical technologies and they supply the other sectors proportionally to their output weights. The relationship between the technical coefficient matrices A and A^* could be described as follows. Because the common sectors in matrix A^* are unchanged, the first N sectors of both models are identical. It could be described as

$$a^*_{ij} = a_{ij}, \quad \text{for} \quad i, j = 1, \ldots, N \tag{1}$$

Since new sectors are assumed to have identical technologies, the technical coefficients of the new sectors are equal to the sector to be disaggregated. This could be described as

$$a^*_{i,N+k} = a_{i,N+1}, \quad \text{for} \quad i = 1, \ldots, N \quad \text{and} \quad k = 1, \ldots, n \tag{2}$$

Since the supply from the new sectors to the other sectors are assumed to be proportionally to their output weights, it could be described as

$$a^*_{N+k,i} = \sigma_k a_{N+1,i}, \quad \text{for} \quad i = 1, \ldots, N \quad \text{and} \quad k = 1, \ldots, n \tag{3}$$

The technical coefficients of the intra matrix could be described as

$$a^*_{N+k,N+k'} = \sigma_k a_{N+1,N+1}, \quad \text{for} \quad k = 1, \ldots, n \tag{4}$$

2 Example

2.1 Case Description

This section uses the data of the paper by Linder et al. (2013) as an example, to disaggregate sector 23 (electricity production, heat and water distribution and supply)

of the Chinese I-O table into nine new sectors. The original 42×42 sectors matrix will be disaggregated into a 50×50 sectors matrix finally.

Two sets of weight factors were derived in the paper. One set comprises the input weight factors $\rho_k = x^*_{N+k} \big/ x_{N+1}$, which split all inputs into the new electricity production sectors from the common sectors, by definition $\sum_{k=1}^{n} \rho_{i,k}{=}1$. The input weight factors are derived by taking the weighted sum of the power plants Operation and Maintenance costs and annual electricity generation output.

The other set comprises the output weight factors w_k, which split the output of the new electricity production sectors into all the other sectors, according to the regionally weighted industry consumption mixes, by defintion $\sum_{k=1}^{n} w_{k,i}{=}1$.

2.2   Calculation

With the familiar definition of the I-O models, the amount of good i that sector j consumes from sector i is noted as z_{ij}, and we have $a_{ij} = z_{ij}/x_j$. In this case, for the first 41 sectors, the models are identical. That is, $z^*_{ij} = z_{ij}$, for i, $j = 1$ to 41. We will calculate the disaggregated input matrix DIM, the disaggregated output matrix DOM and the disaggregated intra matrix DINM, respectively.

Disaggregated input matrix DIM, represents all the input from common sectors into the new sectors. It could be calculated by following the method introduced above.

$$a^*_{i,N+k} = a_{i,N+1}, \quad \text{for} \quad i = 1, \ldots, N \quad \text{and} \quad k = 1, \ldots, n \tag{2}$$

$$\frac{z^*_{i,N+k}}{x^*_{N+k}} = \frac{z_{i,N+1}}{x_{N+1}} \tag{5}$$

$$z^*_{i,N+k} = \frac{x^*_{N+k}}{x_{N+1}} z_{i,N+1} \tag{6}$$

$$z^*_{i,N+k} = \rho_{i,k} z_{i,N+1} \tag{7}$$

$$\sum_{k=1}^{n} z^*_{i,N+k} = z_{i,N+1} \tag{8}$$

Disaggregated output matrix DOM, represents all possible output weights of new sectors into common sectors. It could be calculated by

$$a^*_{N+k,i} = w_{k,i} * a_{N+1,i}, \quad \text{for} \quad i = 1, \ldots, N \quad \text{and} \quad k = 1, \ldots, n \tag{3}$$

$$\frac{z^*_{N+k,i}}{x^*_i} = w_{k,i} \frac{z_{N+1,i}}{x_i} \tag{9}$$

$$z^*_{N+k,i} = w_{k,i} z_{N+1,i} \tag{10}$$

$$\sum_{k=1}^{n} z^*_{N+k,i} = z_{N+1,i} \tag{11}$$

Intra matrix DINM, represents the allocation of intra-industry sales in the new sectors. It could be calculated by

$$a^*_{N+k,N+k'} = \rho_k a_{N+1,N+1}, \quad \text{for} \quad k = 1, \ldots, n \tag{4}$$

$$a^*_{N+k,N+k'} x^*_{N+k'} x_N + 1 = \rho_k a_{N+1,N+1} x^*_{N+k'} x_N + 1 \tag{12}$$

$$a^*_{N+k,N+k'} x^*_{N+k'} = \rho_k a_{N+1,N+1} x_N + 1 \frac{x^*_{N+k'}}{x_{N+1}} \tag{13}$$

$$z^*_{N+k,N+k'} = \rho_k z_{N+1,N+1} \rho_{k'} \tag{14}$$

$$\sum_{k=1}^{n} \sum_{k'=1}^{n} z^*_{N+k,N+k'} = z_{N+1,N+1} \tag{15}$$

For the example in this paper, we have $N = 41$, $k = 9$. The dimensions of the matrices IM, OM and INM are 41*1, 1*41 and 1*1, respectively; the outcome of the disaggregated matrices DIM, DOM and DINM have the dimensions of 41*9, 9*41 and 9*9, respectively.

The *input weight factor* ρ is a 1 by 9 matrix, which is given as

$$\rho = \begin{bmatrix} \rho_1 & \rho_2 & \cdots & \rho_9 \end{bmatrix} \tag{1}$$

Since the production efficiency of the new sectors will be affected by some of the common sectors, the supply from these common sectors to the new sectors needs to be manually allocated.

The *output weight factor* w is the 9*41 matrix, which is given as

$$w = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1,41} \\ w_{21} & w_{22} & \cdots & w_{2,41} \\ \vdots & & & \\ w_{91} & w_{92} & \cdots & w_{9,41} \end{bmatrix} \tag{2}$$

Reorganize the I-O table by moving the sector 23 to the last row and column of the intermediate use part, and put the weight factors in the corresponding positions as showing in Figure 3. The format of Figure 3 will be used to read the data from the Excel file to Python.

Figure 3: Schematic example of Reorganized I-O Table

2.3 Supporting Python Code

```python
################################################
## Import Site-Packages that will be used
import xlrd
# xlrd: extract data from Microsoft Excel spreadsheet files
import xlwt
# xlwt: create spreadsheet files compatible with Microsoft Excel files
import numpy as np
# Numpy: array processing for numbers, strings, records, and objects


################################################
## Read data from Excel file
file_path = 'J:\\Technical Documents\\Disaggregating the Electricity Sector
\\Chinese_National_IO_Table_2007.xls'
workbook = xlrd.open_workbook(file_path)
table = workbook.sheet_by_name('sheet1') # read the data from 'sheet1'

nrows = table.nrows # get the number of rows
ncols = table.ncols # get the number of columns

################################################
## Calculate Disaggregate Input Matrix DIM and save it into the new Excel file
    # Read data of matrix IM
IM = table.col_values((ncols-10),0,(nrows-10))
# get the data of Input matrix at col(ncols-10) from row(0) to row(nrows-10)
arrayIM = np.array(IM)
arrayIM.shape = (41,1)

    # Read data of rho
rho = table.row_values(0,(ncols-9),ncols)
# get the data of rho at row(0) from col(ncols-9) to col(ncols)
arrayrho = np.array(rho)
arrayrho.shape = (1,9)

    # Manually allocate rho of some of the common sectors
i_array = np.ones((41,1))  # array with all elements as 1
newrho = np.dot(i_array, arrayrho) # with shape 41*9
```

```python
a = [1,2,10,22,23,25]
# the rows of the common sectors that needed to be manually allocated
for i in a:
    MArho = table.row_values(i,(ncols-9),ncols)
    # get the value of the manually allocated rho
    newrho[i] = MArho
    # replace the value of the manually allocated rho to the newrho

    # Manipulate matrix DIM
DIM = arrayIM * newrho

    # Export the result into a new sheet
Result = xlwt.Workbook()
sheet1 = Result.add_sheet('Disaggregated_Input_Matrix')
# Exprot the result to the sheet of 'Disaggregated_Input_Matrix'

for row, array in enumerate(DIM):
    for col, value in enumerate(array):
        sheet1.write(row, col, value)

## Calculate Disaggregate Output Matrix DOM and save it into the new Excel file
    # Read data of matrix OM
OM = table.row_values((nrows-10),0,41)
# get the data of Output matrix at row(nrows-10) from col(0) to col(41)

    # Read data of omega
omega_list = []
for i in range((nrows-9),nrows):  # set the range of omega
    omega_data = table.row_values(i,0,41) # get data at row i from col(0) to col(41)
    omega_list.append(omega_data) # combine data together
omega = np.array(omega_list) # make omega as a matrix

    # Manipulate matrix DOM
arrayOM = np.array(OM) # make OM as a matrix
diaOM = np.diag(arrayOM) # diagonalize matrix OM
DOM = np.dot(omega,diaOM) # Calculate matrix DOM

    # Export the result into a new sheet
sheet2 = Result.add_sheet('Disaggregated_Output_Matrix')
# Exprot the result to the sheet of 'Disaggregated_Output_Matrix'
```

```python
for row, array in enumerate(DOM):
    for col, value in enumerate(array):
        sheet2.write(row, col, value)

## Calculate Intra Matrix DINM and save it into the new Excel file
    # Read data of matrix INM
INM = table.cell((nrows-10),(ncols-10)).value
# get the data of Intra matrix at the cell (nrows-10),(ncols-10)

    # Manipulate matrix DINM
arrayrhotrans = np.transpose(arrayrho)
DINM =np.array(INM*np.dot(arrayrhotrans, arrayrho))

    # Export the result into a new sheet
sheet3 = Result.add_sheet('Intra_Matrix')
# Exprot the result to the sheet of 'Intra_Matrix'

for row, array in enumerate(DINM):
    for col, value in enumerate(array):
        sheet3.write(row, col, value)

name = 'J:\\Technical Documents\\Disaggregating the Electricity Sector
\\RESULTS.xls'
# export file path.
Result.save(name)
```

References

Comer, J. C., R. W. Jackson (1997) A note on adjusting national input-output data for regional table construction. *Journal of regional science*, 37, 145-153.

Jackson, R. W., J. C. Comer (1993) An alternative to aggregated base tables in input-output table regionalization. *Growth and Change*, 24, 191-205.

Lahr, M. L., B. H. Stevens (2002) A study of the role of regionalization in the generation of aggregation error in regional input-output models. *Journal of regional science*, 42, 477-507.

Lindner, S., J. Legault, and D. Guan (2013) Disaggregating the Electricity Sector of China's Input-Output Table for Improved Environmental Life-Cycle Assessment. *Economic Systems Research*, 25, 300-320.

Miller, R. E., P. D. Blair. 2009. Input-Output Analysis: Foundations and Extensions. Cambridge, UK: Cambridge University Press.

Lindner, S., J. Legault, and D. Guan (2012) Disaggregating Input-Output Models with Incomplete Information. *Economic Systems Research*, 24, 329-347.

Wolsky, A. (1984) Disaggregating Input-Output Models. *The Review of Economics and Statistics*, 66, 283-291.