Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- you will use the copy only for the purposes of research or private study
- you will recognise the author's right to be identified as the author of the thesis and due acknowledgement will be made to the author where appropriate
- you will obtain the author's permission before publishing any material from the thesis.
A METHODOLOGY FOR AGGREGATING INDUSTRIES OF INPUT-OUTPUT MODELS, WITH APPLICATION TO NEW ZEALAND INTERINDUSTRY DATA.

A thesis submitted in fulfilment of the requirements for the Degree of Master of Applied Science in the University of Canterbury

by

J.R. Rodgers

Lincoln College
March 1978

A METHODOLOGY FOR AGGREGATING INDUSTRIES OF INPUT-OUTPUT MODELS, WITH APPLICATION TO NEW ZEALAND INTERINDUSTRY DATA.

by J.R. Rodgers

Aggregation, as it applies to input-output analysis, is the process of grouping industries of an interindustry model into sectors so as to produce an intersectoral model, which is more manageable in that its dimensions are smaller than those of the original model. A problem which arises as a result of aggregation is that, in general, forecasts of sector gross outputs, obtained from the intersectoral model, differ from those obtained by aggregating forecasts of industry gross outputs produced by the original interindustry model. This phenomenon is known as aggregation bias.

When there is a need to condense an input-output model into one of smaller dimensions, it is desirable that the resulting intersectoral model is subject to the smallest possible degree of aggregation bias. The researcher may also wish to place constraints upon the intersectoral model to ensure that certain industries are, or are not, aggregated into the same sector. Although there is a substantial body of theory which specifies the conditions under which the
intersectoral model is not subject to aggregation bias, the extent to which various industry groupings satisfy these conditions is difficult to assess subjectively. Therefore, a formalised procedure, which groups industries into sectors so that aggregation bias is minimized, and which allows the imposition of constraints upon the intersectoral model, is a useful aid to the input-output analyst. Such a procedure is developed in this thesis and is applied to the most recent interindustry model of the New Zealand economy.
## CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>INTRODUCTION</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>INTRODUCTION</td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>The Aims of the Study</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>The Aggregation Problem in Economic Analysis</td>
<td>2</td>
</tr>
<tr>
<td>1.3</td>
<td>The Aggregation Problem in Input-Output Analysis</td>
<td>8</td>
</tr>
<tr>
<td>1.4</td>
<td>The Need for Aggregation in Input-Output Analysis</td>
<td>14</td>
</tr>
<tr>
<td>1.5</td>
<td>Summary</td>
<td>20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter</th>
<th>AGGREGATION IN STATIC, OPEN, INPUT-OUTPUT ANALYSIS: A REVIEW OF THE LITERATURE</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>AGGREGATION IN STATIC, OPEN, INPUT-OUTPUT ANALYSIS: A REVIEW OF THE LITERATURE</td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td>A Mathematical Statement of Aggregation Bias</td>
<td>22</td>
</tr>
<tr>
<td>2.2</td>
<td>A Mathematical Statement of First-Order Aggregation Bias</td>
<td>24</td>
</tr>
<tr>
<td>2.3</td>
<td>The Conditions under which Aggregation Bias Vanishes</td>
<td>25</td>
</tr>
<tr>
<td>2.4</td>
<td>The Conditions under which First-Order Aggregation Bias Vanishes</td>
<td>37</td>
</tr>
<tr>
<td>2.5</td>
<td>Aggregation Bias in Output and Income Multipliers</td>
<td>42</td>
</tr>
<tr>
<td>2.6</td>
<td>Current Aggregation Procedures</td>
<td>44</td>
</tr>
<tr>
<td>2.6.1</td>
<td>Non Clustering Methods of Aggregation</td>
<td>45</td>
</tr>
<tr>
<td>2.6.2</td>
<td>Cluster Analysis</td>
<td>52</td>
</tr>
<tr>
<td>2.6.3</td>
<td>Aggregation using Cluster Analysis</td>
<td>55</td>
</tr>
<tr>
<td>2.7</td>
<td>Summary</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>Appendix to Chapter 2</td>
<td>65</td>
</tr>
<tr>
<td>Chapter</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>3</td>
<td>A METHODOLOGY FOR AGGREGATION</td>
<td>67</td>
</tr>
<tr>
<td>3.1</td>
<td>The Capabilities Required of a Methodology for Aggregation</td>
<td>67</td>
</tr>
<tr>
<td>3.1.1</td>
<td>The Aggregation Criterion</td>
<td>67</td>
</tr>
<tr>
<td>3.1.2</td>
<td>Aggregation Constraints</td>
<td>68</td>
</tr>
<tr>
<td>3.1.3</td>
<td>Other Requirements</td>
<td>70</td>
</tr>
<tr>
<td>3.2</td>
<td>A Clustering Algorithm for Aggregating Industries of an Input-Output Model</td>
<td>70</td>
</tr>
<tr>
<td>3.3</td>
<td>Implementation of the Clustering Algorithm</td>
<td>85</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Construction of the Initial Similarity Matrix</td>
<td>85</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Updating the Similarity Matrix</td>
<td>86</td>
</tr>
<tr>
<td>3.3.3</td>
<td>The Computer Routine</td>
<td>98</td>
</tr>
<tr>
<td>3.4</td>
<td>Summary</td>
<td>101</td>
</tr>
<tr>
<td>4</td>
<td>APPLICATION OF THE METHODOLOGY FOR AGGREGATION</td>
<td>103</td>
</tr>
<tr>
<td>4.1</td>
<td>Results of Aggregation Based on Similarity of Partially Aggregated Input Coefficients</td>
<td>103</td>
</tr>
<tr>
<td>4.2</td>
<td>Results of Constrained Aggregation, Based on Similarity of Partially Aggregated Input Coefficients</td>
<td>121</td>
</tr>
<tr>
<td>4.3</td>
<td>Results of Aggregation Based on Similarity of Input Coefficients</td>
<td>134</td>
</tr>
<tr>
<td>4.4</td>
<td>An Assessment of Aggregation Bias in Forecasts of Gross Output</td>
<td>149</td>
</tr>
<tr>
<td>4.5</td>
<td>Summary</td>
<td>155</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

1.1 The Aims of the Study

The aims of this study are fourfold: (1) to analyse the problems associated with aggregating industries of a static, open, input-output model, (2) to examine the theoretical conditions under which the problems are minimized, (3) to develop an aggregation procedure which conforms as closely as possible with the theoretical conditions, and (4) to apply the aggregation procedure to the most recent input-output model of the New Zealand economy and evaluate its performance.

Aggregation problems are not restricted to input-output models. Indeed, the aggregation problem is common to many areas of economic analysis. Consequently, the next section of this introductory chapter presents a brief account of the difficulties caused by the necessity to aggregate in building economic models which are not of the input-output type. This is followed, in Section 1.3, by a discussion of the undesirable consequences of aggregating industries of an input-output model. Despite the problems associated with aggregation, frequently the input-output analyst is forced to aggregate for a number of reasons which are presented in Section 1.4. Finally, in Section 1.5, the main points of this
chapter are summarised and the contents of the remainder of this thesis are previewed.

1.2 The Aggregation Problem in Economic Analysis

Aggregation is a process whereby some of the information, which is available for the solution of a problem, is sacrificed in order to make the problem more manageable. For example, a set of prices of individual commodities might be replaced by a single price index number, or a set of incomes of individual consumers might be replaced by their total, or average, income. The indiscriminate use of aggregated variables in economic models leads to various complications when these models are used for explanatory, or forecasting, purposes.

Economic models are often set up on the assumption that simple aggregates, or averages, are directly related. However, the economic theory underlying the model is usually stated in micro terms, based on decisions taken by individual consumers or firms. Consequently, economic models, which purport to explain the aggregate behaviour of groups of consumers or firms, can be viewed as derived relations, the transition from the level of the individual economic unit to the level of the market, the industry, the economy etc., being made through aggregation. Admittedly, economic theory recognises instances

1. For detailed expositions on this topic see, for example, Theil (1954), Allen [(1959), chapter 20], Green (1964) and Fisher (1969).
when the whole does not equal the sum of the component parts and to this extent the micro theory might be deficient in a way which is relevant to the problem of aggregation. For example, if each household's consumption expenditure depends, not only upon its own income, but also upon total income, then a micro theory, which omits total income, contains a specification error. However, if no specification error exists, then the variables, which appear in a model of aggregate economic behaviour, are derived by aggregating individual relationships. This procedure frequently results in an economic model, the variables of which are not simple aggregates or averages. The following two examples, taken from the areas of statistical demand analysis and production function analysis, demonstrate this phenomenon.

The classical theory of consumer behaviour states that an individual's demand for a good is determined by the relative prices of all commodities and his real income. Although there is no a priori basis for assuming that individual demand functions are linear or nonlinear, for simplicity, linearity is assumed:

\[ q_i = a_{0i} + a_{1i} \frac{p_1}{p} + \ldots + a_{ki} \frac{p_k}{p} + \beta_i \frac{y_i}{p} \]

where \( q_i \) is the \( i \)th individual's demand for the good, \( p_j \) is the price of good \( j \) (\( j=1,2,\ldots,k \)) and the same prices are assumed to be paid by all consumers, \( y_i \) is the income of consumer \( i \), \( p \) is a weighted average of all prices, and \( a_{0i}, a_{1i}, \ldots, a_{ki}, \beta_i \) are the parameters of the \( i \)th individual's demand function.
The market demand function is derived by aggregating the individual demand functions of all N consumers who comprise the market:

\[ \sum_{i=1}^{N} q_i = \sum_{i=1}^{N} \alpha_{0i} + \sum_{i=1}^{N} \alpha_{1i} p_i / p + \ldots + \sum_{i=1}^{N} \alpha_{ki} p_k / p + \sum_{i=1}^{N} \beta_i y_i / p \ldots (2) \]

By defining \[ \sum_{i=1}^{N} q_i = Q, \sum_{i=1}^{N} \alpha_{0i} = \alpha_0', \sum_{i=1}^{N} \alpha_{1i} = \alpha_1', \ldots, \sum_{i=1}^{N} \alpha_{ki} = \alpha_k', \]

\[ \sum_{i=1}^{N} \beta_i = \bar{\beta} \text{ and } \frac{\sum_{i=1}^{N} \beta_i y_i}{N} = \bar{y} \] the market demand function becomes:

\[ Q = \alpha_0 + \frac{\alpha_1 p_1}{p} + \ldots + \frac{\alpha_k p_k}{p} + \bar{\beta} \bar{y} \ldots (3) \]

This market demand function is simple to interpret. Its intercept is the sum of the individual micro intercepts, the change in the market demand for the good per unit change in the price of good \( j \) \((j = 1, 2, \ldots, k)\) is, *ceteris paribus*, the sum of the individual changes in quantity demanded, and the marginal propensity to consume in the aggregate, \( \bar{\beta} \), is the mean of the individual marginal propensities to consume, \( \beta_i \).

Note, however, that the linear market demand function does not express a relationship between simple aggregates; aggregate real income equals the market size multiplied by a weighted average of individual real incomes, the weights being the individual marginal propensities to consume, \( \beta_i \). Consequently, statistical estimation of a market demand function using time series observations of simple aggregates of quantity
demanded and real income, as well as relative prices, will, in general, result in biased estimators of the aggregate parameters. Predictions of aggregate quantity demanded, derived from this estimated model, are also biased.

Further complications arise if prices vary between consumers, if aggregation takes place over commodities as well as over consumers, or if individual demand functions are nonlinear.

In production function analysis, the aggregation problem is to show how a production function for an industry,

2. The "true" aggregate econometric model can be written in matrix form as $$Q = X\gamma + u,$$ but the model being estimated is misspecified as $$Q = \tilde{X}\gamma + u.$$ Assuming prices and income are exogenous, estimators of $$\gamma$$ are given by:

$$\hat{\gamma} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'Q$$

$$= (\tilde{X}'\tilde{X})^{-1}\tilde{X}'(X\gamma + u)$$

$$= (\tilde{X}'\tilde{X})^{-1}\tilde{X}'X\gamma + (\tilde{X}'\tilde{X})^{-1}\tilde{X}'u$$

Hence $$E(\hat{\gamma}) = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'X\gamma$$

$$\neq \gamma$$

Estimators are unbiased only if all individuals have the same marginal propensity to consume the particular good, or if the distribution of incomes is stable through time [Malinvaud (1970), p. 133].
or an entire economy, is related to those of single firms.²

At the level of the individual firm, economic theory specifies a relationship between value added, labour and capital. For simplicity, a Cobb-Douglas relationship is assumed:

\[ v_i = \alpha_i \ell_i^\beta_i k_i^\gamma_i \quad \ldots (4) \]

where \( v_i \) is value added for the \( i \)th firm,

\( \ell_i \) is labour employed by the \( i \)th firm,

\( k_i \) is capital utilised by the \( i \)th firm, and

\( \alpha_i, \beta_i \) and \( \gamma_i \) are the parameters of the \( i \)th firm's production function.

For estimation purposes, the model is transformed into a form which is linear in logarithms:

\[ \log(v_i) = \log(\alpha_i) + \beta_i \log(\ell_i) + \gamma_i \log(k_i) \quad \ldots (5) \]

Aggregating over \( N \) firms within an industry, an aggregate production function of the following form is obtained:

\[ \sum_{i=1}^{N} \log(v_i) = \sum_{i=1}^{N} \log(\alpha_i) + \sum_{i=1}^{N} \beta_i \log(\ell_i) + \sum_{i=1}^{N} \gamma_i \log(k_i) \quad \ldots (6) \]

By defining \( \log(V) = \frac{1}{N} \sum_{i=1}^{N} \log(v_i), \ \log(\alpha) = \frac{1}{N} \sum_{i=1}^{N} \log(\alpha_i), \)

3. This should not be confused with the problem relating to aggregation of inputs within an individual firm. For a discussion of the aggregation problem in production function analysis, see Bridge [(1971), pp. 348-352].
\[
\beta = \frac{\sum_{i=1}^{N} \beta_i \log(L_i)}{\sum_{i=1}^{N} \log(L_i)}, \quad \gamma = \frac{\sum_{i=1}^{N} \gamma_i \log(k_i)}{\sum_{i=1}^{N} \log(k_i)}, \quad \log(L) = \frac{1}{N} \sum_{i=1}^{N} \log(L_i)
\]

and \( \log(K) = \frac{1}{N} \sum_{i=1}^{N} \log(k_i) \) the production function for the average firm becomes:

\[
\log(V) = \log(a) + \beta \log(L) + \gamma \log(K) \quad \ldots(7)
\]

or

\[
V = aL^\beta K^\gamma \quad \ldots(8)
\]

The aggregate production function is also Cobb-Douglas in form and is readily interpretable. The parameters \( \beta \) and \( \gamma \) are weighted averages of the parameters \( \beta_i \) and \( \gamma_i \) respectively in the micro production functions. The parameter \( a \), however, is the geometric mean of the parameters \( a_i \) in the production functions for the individual firms. The variables in the aggregate production function are also geometric means of value added, labour and capital, respectively. If simple averages of these variables are used, instead of geometric means, in estimating the parameters \( a, \beta \) and \( \gamma \) then the estimators so derived are biased and predictions based upon the estimated model are also biased.

The unfortunate fact about aggregation is that, although the theoretically correct aggregate model can be derived from economic theory, data are seldom available in the form required for its statistical estimation.
1.3 The Aggregation Problem in Input-Output Analysis

Aggregation bias is a potential danger in many economic models but in this thesis attention is confined to problems of aggregation in input-output models. The example given below demonstrates how the aggregation of industries of an input-output model can lead to unreliable forecasting.

Consider a static, open, input-output model, consisting of four industries, each of which produces one and only one commodity, using a single method of production in which inputs are nonsubstitutable. The amount of each input required to produce one unit of output by a given industry is assumed to be independent of that industry's total output. For a given set of final demands, the model's solution states the gross output which each industry must produce in order to satisfy both intermediate and final demand. For example, let the interindustry flows be those in Table 1.3.1.

<table>
<thead>
<tr>
<th>Table 1.3.1 Hypothetical Interindustry Transactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Origin</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>Industry 1</td>
</tr>
<tr>
<td>Industry 2</td>
</tr>
<tr>
<td>Industry 3</td>
</tr>
<tr>
<td>Industry 4</td>
</tr>
</tbody>
</table>
The technical coefficients of the system are given in Table 1.3.2. The elements of column \( i (i=1,2,3,4) \) represent the amount of input from each industry, which is required by industry \( i \) in order to produce one unit of output. The elements of Table 1.3.2, and the tables which follow, have been expressed as fractions to avoid rounding errors.

**Table 1.3.2**

<table>
<thead>
<tr>
<th>Origin</th>
<th>Industry 1</th>
<th>Industry 2</th>
<th>Industry 3</th>
<th>Industry 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industry 1</td>
<td>10/100</td>
<td>10/100</td>
<td>20/100</td>
<td>10/100</td>
</tr>
<tr>
<td>Industry 2</td>
<td>20/100</td>
<td>30/100</td>
<td>0/100</td>
<td>20/100</td>
</tr>
<tr>
<td>Industry 3</td>
<td>15/100</td>
<td>0/100</td>
<td>40/100</td>
<td>0/100</td>
</tr>
<tr>
<td>Industry 4</td>
<td>10/100</td>
<td>10/100</td>
<td>20/100</td>
<td>25/100</td>
</tr>
</tbody>
</table>

The interdependence coefficients (or direct and indirect requirements) of the system are given in Table 1.3.3. The elements of column \( i (i=1,2,3,4) \) represent the change in gross output of each industry, which is required to meet a unit change in final demand for the product of industry \( i \).

**Table 1.3.3**

<table>
<thead>
<tr>
<th>Origin</th>
<th>Industry 1</th>
<th>Industry 2</th>
<th>Industry 3</th>
<th>Industry 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industry 1</td>
<td>30300/23925</td>
<td>5100/23925</td>
<td>11900/23925</td>
<td>5400/23925</td>
</tr>
<tr>
<td>Industry 2</td>
<td>10800/23925</td>
<td>37350/23925</td>
<td>7400/23925</td>
<td>11400/23925</td>
</tr>
<tr>
<td>Industry 3</td>
<td>7575/23925</td>
<td>1275/23925</td>
<td>42850/23925</td>
<td>1350/23925</td>
</tr>
<tr>
<td>Industry 4</td>
<td>7500/23925</td>
<td>6000/23925</td>
<td>14000/23925</td>
<td>34500/23925</td>
</tr>
</tbody>
</table>
Given a new set of final demands equal to 700 (industry 1), 500 (industry 2), 1000 (industry 3) and 1500 (industry 4), gross outputs, required to satisfy intermediate plus final demand, are calculated as follows:

\[
\frac{(30300)(700)}{23925} + \frac{(5100)(500)}{23925} + \frac{(11900)(1000)}{23925} + \frac{(5400)(1500)}{23925} = 1829.0 \text{ (industry 1)}
\]

\[
\frac{(10800)(700)}{23925} + \frac{(37350)(500)}{23925} + \frac{(7400)(1000)}{23925} + \frac{(11400)(1500)}{23925} = 2120.6 \text{ (industry 2)}
\]

\[
\frac{(7575)(700)}{23925} + \frac{(1275)(500)}{23925} + \frac{(42850)(1000)}{23925} + \frac{(1350)(1500)}{23925} = 2123.9 \text{ (industry 3)}
\]

\[
\frac{(7500)(700)}{23925} + \frac{(6000)(500)}{23925} + \frac{(14000)(1000)}{23925} + \frac{(34500)(1500)}{23925} = 3093.0 \text{ (industry 4)}
\]

Now suppose a new model is constructed in which the original four industries are grouped to form two composite industries. For convenience, each composite industry will be referred to as a sector. The solution to this aggregated model specifies the gross output of each sector, which is required to meet its intermediate and final demand. For example, if industries 1 and 2, in Table 1.3.1, are grouped to form sector 1, and if industries 3 and 4 are combined to form sector 2, then intersector transactions flows are those given in Table 1.3.4.
Table 1.3.4

Hypothetical Intersector Transactions

<table>
<thead>
<tr>
<th>Origin</th>
<th>Sector 1</th>
<th>Sector 2</th>
<th>Final Demand</th>
<th>Gross Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sector 1</td>
<td>1500</td>
<td>700</td>
<td>1800</td>
<td>4000</td>
</tr>
<tr>
<td>Sector 2</td>
<td>550</td>
<td>800</td>
<td>1150</td>
<td>2500</td>
</tr>
</tbody>
</table>

The technical coefficients of the aggregated system are given in Table 1.3.5. The elements of column i (i=1,2) represent the amount of input from each sector, which is required by sector i in order to produce one unit of output.

Table 1.3.5

Aggregated Technical Coefficients

<table>
<thead>
<tr>
<th>Origin</th>
<th>Sector 1</th>
<th>Sector 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sector 1</td>
<td>30/80</td>
<td>7/25</td>
</tr>
<tr>
<td>Sector 2</td>
<td>11/80</td>
<td>8/25</td>
</tr>
</tbody>
</table>

The interdependence coefficients are given in Table 1.3.6. The elements of column i (i=1,2) represent the change in gross output of each sector, which is required to meet a unit change in final demand for the product of sector i.

Table 1.3.6

Aggregated Interdependence Coefficients

<table>
<thead>
<tr>
<th>Origin</th>
<th>Sector 1</th>
<th>Sector 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sector 1</td>
<td>1360/773</td>
<td>560/773</td>
</tr>
<tr>
<td>Sector 2</td>
<td>275/773</td>
<td>1250/773</td>
</tr>
</tbody>
</table>
The previous set of final demands (that is, 700 for industry 1, 500 for industry 2, 1000 for industry 3 and 1500 for industry 4) are aggregated in the same way to give final demands of 1200 for sector 1 and 2500 for sector 2. Gross outputs, required to satisfy both intermediate and final demand, are calculated as follows:

\[
\begin{align*}
\frac{(1360)(1200)}{773} + \frac{(560)(2500)}{773} &= 3922.4 \text{ (sector 1)} \\
\frac{(275)(1200)}{773} + \frac{(1250)(2500)}{773} &= 4469.6 \text{ (sector 2)}
\end{align*}
\]

Sector outputs can be obtained from the intersectoral model, as demonstrated above, or by aggregating the outputs of the interindustry model, but, in general, the two sets of results differ. In the above example, the combined output of industries 1 and 2 is 3949.6 compared with 3922.4 for sector 1, and the combined output of industries 3 and 4 is 5216.9 compared with 4469.6 for sector 2. The discrepancies between the two sets of results are known collectively, and individually, as aggregation bias.

The cause of aggregation bias is the existence of "product mixes" in the aggregated model, which violates one of the basic assumptions of input-output analysis, namely, that each industry produces a single commodity. The preceding example demonstrates that conventional grouping of industries 1 and 2 and industries 3 and 4 into separate sectors results in aggregate technical coefficients of the form given in
Table 1.3.7.

<table>
<thead>
<tr>
<th>Origin</th>
<th>Sector 1</th>
<th>Sector 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sector 1</td>
<td>[\frac{(a_{11}+a_{21})X_1+(a_{12}+a_{22})X_2}{X_1+X_2}]</td>
<td>[\frac{(a_{13}+a_{23})X_3+(a_{14}+a_{24})X_4}{X_3+X_4}]</td>
</tr>
<tr>
<td>Sector 2</td>
<td>[\frac{(a_{31}+a_{41})X_1+(a_{32}+a_{42})X_2}{X_1+X_2}]</td>
<td>[\frac{(a_{33}+a_{43})X_3+(a_{34}+a_{44})X_4}{X_3+X_4}]</td>
</tr>
</tbody>
</table>

The elements \(a_{ij}\) are the technical coefficients of the original interindustry model, and \(X_i\) \((i=1,2,3,4)\) is the gross output of industry \(i\) in the period in which interindustry flows are collected, that is, the "base period".

Sector 1 produces the two commodities originally produced by industries 1 and 2, while sector 2 produces the two commodities originally produced by industries 3 and 4. A change in one or more of the ratios \(X_1/(X_1+X_2)\), \(X_2/(X_1+X_2)\), \(X_3/(X_3+X_4)\) or \(X_4/(X_3+X_4)\) will cause the technical coefficients of the aggregated model to change, even if those of the original model remain stable. It is this instability in the technical coefficients, caused by changes in the product mix, which leads to aggregation bias.

When forecasts, which are subject to aggregation bias, are used for economic planning, the result is an inconsistent plan. For example, suppose two industries, one of which produces cars and the other of which produces trucks, are
aggregated to form the motor vehicle sector. Consider an increase in final demand for motor vehicles, which is attributable entirely to an increase in demand for cars. If planners calculate input requirements for the motor vehicle sector using predicted gross sector outputs and aggregate technical coefficients of the form given in Table 1.3.7, then, in general, the car industry will find itself short of some inputs but it will have a surplus of others [Ellman (1969), p. 72].

1.4 The Need for Aggregation in Input-Output Analysis

It is necessary to distinguish between aggregation, which is introduced during the construction of an interindustry table, and aggregation as used by the researcher to reduce the size of an existing interindustry table.

Aggregation inevitably takes place at the model construction stage, when industries are defined and data on interindustry flows are collected. In many countries, including New Zealand, the statistical unit used in data collection is the establishment, and establishments are classified into industries according to their main line of production. However, most establishments produce secondary products, either independently of the primary product, or as a joint product resulting from the same production process as the primary
product and sharing a common input structure with it. Unless secondary products are transferred to separate homogeneous industries, product mixes invariably result. In practice, there are a number of reasons why it is impossible to eliminate all product mixes in this way. Firstly, the resulting number of industries would be extremely large and the interindustry table would be too detailed to be used in analysing the behaviour of the economy as a whole. Secondly, interindustry flow data, involving such artificially constructed industries, would be of dubious quality, and thirdly, the cost and effort involved in compiling the table would be prohibitively high. Furthermore, as the number of industries increases, the possibility of technical substitution between the products of different industries increases. Thus, a more detailed model is more likely to approximate the assumption of homogeneity of outputs, but it is also more likely to violate the assumption of nonsubstitutability of inputs. Obviously, a compromise is required when industries are defined. Consequently, all input-output models contain some degree of aggregation, which is introduced at the time when interindustry flows are estimated.

Faced with the inevitability of aggregation, model builders should be aware of the problem of aggregation bias and of the conditions under which it is likely to be more, or less, serious. Such knowledge will be of use in defining industries and in deciding whether or not to transfer specific secondary products from the industry into which the parent
establishment has been classified.

Once the basic data have been collected, further aggregation may take place before the interindustry table is published. If an industry is so small that its interindustry transactions are not significant when data are rounded to a common unit of measurement, it may be grouped with another industry for publication purposes.

The most recent interindustry model of the New Zealand economy pertains to the year 1965-66, and it has been aggregated for both the reasons mentioned above. Most of its 109 industries contain product mixes which were introduced when industries were defined. The only secondary product to receive special attention is slate wool which has been transferred from the freezing industry to the farming industry. Establishments were originally classified into 142 industries and those which did not have significant interindustry flows, when monetary values were rounded to the nearest $100,000, were aggregated with other industries, reducing the final number in the published tables to 109 [New Zealand Department of Statistics (1974), Part 1, p. 14]. A separate table has also been published in which the 109 industries have been aggregated further into 44 composite industries.

Once the interindustry table is in its final published form, further aggregation may be necessary, or at least desirable.
Firstly, aggregation of industries in an input-output model may be used to aid analysis of the economic system which the model represents. After all, a model, by definition, is a simplified representation of a more complex system, and it is a useful analytic tool if it reveals the essential structure of the underlying system, while ignoring the less important details. This is the same reasoning which leads to the use of averages, index numbers and various other aggregates in economic analysis. Originally, input-output models were seen as an alternative to the use of models based on highly aggregated data [Leontief (1951), p. 210] but the emergence of models containing hundreds of industries raises the question of whether simplicity has been sacrificed for the sake of detail. Of course, if one is interested in detail then no type of aggregation will be satisfactory. Frequently, however, detailed analysis of just a few industries is all that is required so if the remainder can be grouped into just a few sectors, without introducing an unacceptable amount of aggregation bias, then the result is a more tractable model.

Secondly, changes in the technical coefficients of an input-output model may result from technological change as well as from changes in the product mix. If a significant proportion of coefficient instability in an aggregated model is due to changing product mix, then a more detailed model will become obsolete less rapidly than the aggregated model. However, if technological change is the primary factor
causing changes in the coefficients, then an aggregated model, which can be updated frequently, may produce more accurate forecasts than a detailed model which can be updated infrequently owing to cost constraints. Empirical studies, carried out to date, offer conflicting evidence on this matter. The most comprehensive study of this type is that of Bezdek and Dunham (1976). The authors point out that a change in technology is, itself, likely to result in a change in the product mix so it is impossible to isolate changes in the input-output coefficients which are due to technological change, from those which are due to changes in the product mix. Consequently, their empirical investigations were designed merely to measure the degree to which coefficient change is associated with changing product mix. An index of product mix change between 1958 and 1963 was compiled for each of 53 sectors in the 85 sector input-output model of the United States economy. Each of the 53 sectors was chosen because it is an aggregate of two or more industries which appear in the more detailed interindustry models of the U.S. economy for 1958 and 1963. The indices of product mix change and the changes in the sectors' technical coefficients, over the same period, were compared but no statistically significant relationship was found to exist. The authors then constructed an 85 order hybrid matrix by aggregating the 1963 detailed coefficients using 1958 output weights. Theoretically, if changes in the product mix are important in causing changes in the sector coefficients, one
would expect the elements of the hybrid matrix to be close to those of the 1958 matrix. Alternatively, if factors other than changes in the product mix are important in causing changes in the sector coefficients, the elements of the hybrid matrix would be expected to resemble those of the 1963 matrix. For each of the 53 sectors under consideration, the elements of the hybrid matrix more closely approximated the 1963 coefficients than the 1958 coefficients. On the basis of these results, Bezdek and Dunham conclude that changes in the product mix are not significantly related to intertemporal instability in the technical coefficients, although in doing so they disagree with earlier findings which are cited and briefly discussed in their paper. The implication arising from this research is that an aggregated model may be preferable to a detailed model if the coefficients of the former can be updated frequently using published, official statistics.

Finally, there is little doubt that high speed computers have eliminated much of the need to aggregate interindustry tables in order to avoid the computational burden of solving the model. However, with the linkage of several national input-output models to form international input-output models, the number of industries could conceivably run into the thousands. Since the time required to invert a matrix with known methods increases with the cube of the number of rows (or columns), aggregation of industries in models of this kind is likely to be necessary,
both to reduce computation and to aid comprehension.

1.5 Summary

In previous sections of this chapter, the aggregation problem has been discussed in relation to economic model building in general and input-output analysis in particular. Although aggregation enables the construction of a more manageable model, it is likely to result in biased estimators of parameters of econometric models and biased forecasts of gross output obtained from input-output models. Nevertheless, aggregation is unavoidable in input-output analysis. It takes place when the model is constructed and further aggregation may be necessary to simplify analysis, to enable frequent updating of the technical coefficients or to reduce the computational burden of solving extremely large systems of equations.

Aggregation in input-output models has received its fair share of attention in the economic literature and there is now a substantial body of theory which specifies the conditions under which industries of an existing interindustry table may be aggregated, without introducing bias into the results. Unfortunately, most of these conditions are so stringent that they are unlikely to be fulfilled in practice. Relatively few attempts have been made to translate the results of theoretical research into a practical procedure.
which can be used to aggregate large-scale input-output models, without introducing an unacceptable level of aggregation bias. The primary objective of this study is to develop such a procedure.

The remainder of this thesis is organised as follows. Chapter 2 contains a review of the literature dealing with aggregation bias in static, open, input-output models. The conditions under which aggregation bias disappears, entirely and approximately, are derived and previous contributions to the development of a practical aggregation procedure are discussed. Chapter 3 develops a method of aggregation which is demonstrated in Chapter 4. The 109 industry input-output model of the New Zealand economy, published by the Department of Statistics is aggregated into 44 sectors and the results are compared to the 44 industry model, published by the same department. Finally, the main conclusions of the study are presented in Chapter 5.
CHAPTER 2

AGGREGATION IN STATIC, OPEN, INPUT-OUTPUT ANALYSIS: A REVIEW OF THE LITERATURE

2.1 A Mathematical Statement of Aggregation Bias

Consider an economy consisting of \( n \) industries whose interrelationships are represented by an input-output model of the form:

\[
AX + F = X
\]

where \( A \) is an \( nxn \) matrix of technical coefficients, \( X \) is an \( nx1 \) vector of gross outputs, and \( F \) is an \( nx1 \) vector of final demands.

The usual assumptions concerning the model are made: each industry produces one and only one commodity, using a single production process in which inputs are strictly complementary and are used in fixed proportions. Given any final demand vector, \( F \), the gross outputs of the \( n \) industries, which are required to satisfy both intermediate and final demand, are given by:

\[
X = (I-A)^{-1}F
\]

4. Throughout this chapter it will be assumed that interindustry transactions are expressed in value terms. Two studies, which discuss aggregation of industries when interindustry flows are expressed in physical units are McManus (1956a) and Morimoto (1971). A treatment of aggregation bias in lagged and dynamic input-output models can be found in McManus (1956a), Arata (1959), Morimoto (1970) and Ven (1974). A review of aggregation in input-output models is given by Kym and Norsworthy (1976).
The industry outputs, given by equation (10), can be consolidated into the outputs of \( m \) sectors \((m<n)\) by premultiplying \( X \) by an \( mxn \) aggregation operator of the form:

\[
T = \begin{bmatrix}
1 & 1 & 0 & \ldots & 0 & \ldots & 0 & 0 \\
0 & 0 & 1 & \ldots & 1 & \ldots & 0 & 0 \\
& & & & & & & \\
& & & & & & & \\
& & & & & & & \\
0 & 0 & 0 & \ldots & 0 & \ldots & 1 & 1
\end{bmatrix}
\]

...(11)

Each row of \( T \) contains at least one unit and the remaining elements are zero, signifying that each sector contains at least one industry. Each column of \( T \) contains exactly one unit and the remaining elements are zero, specifying that each industry is allocated to one, and only one, sector. Hence, the \( m \) sector outputs are given by:

\[
TX = T(I-A)^{-1}F
\]

...(12)

Alternatively, the original input-output model can be aggregated to form a hybrid model of the form:

\[
\bar{A}X + \bar{F} = \bar{X}
\]

...(13)

where \( \bar{A} \) is an \( mxm \) matrix of aggregated technical coefficients, obtained by aggregating base period transaction flows between industries of the same sector, and dividing by aggregated base period outputs of industries within sectors. (This procedure was illustrated in Section 1.3.) \( \bar{X} \) is an \( mx1 \) vector of sector gross outputs, and \( \bar{F} \) is an \( mx1 \) vector of aggregated final demands, equal to \( TF \).
The gross outputs of the m sectors, which are required to satisfy intermediate plus final demand, are given by:

\[ \bar{X} = (I-\bar{A})^{-1}TF \]  

...(14)

Hence, the gross outputs of the m sectors can be determined by equation (12) or by equation (14). The discrepancies between the two sets of results are the elements of the vector:

\[ \bar{X} - TX = (I-\bar{A})^{-1}TF - T(I-A)^{-1}F \]  

...(15)

which is known as aggregation bias. When the right hand side of equation (15) is a zero vector, aggregation bias is said to vanish, or, alternatively, aggregation is said to be perfectly consistent.

2.2 A Mathematical Statement of First-Order Aggregation Bias

Expanding \((I-A)^{-1}\) and \((I-\bar{A})^{-1}\) as power series [Waugh (1950)] of the form:

\[ (I-A)^{-1} = I + A + A^2 + A^3 + \ldots \]  

...(16)

\[ (I-\bar{A})^{-1} = I + \bar{A} + \bar{A}^2 + \bar{A}^2 + \ldots \]  

...(17)

and substituting equations (16) and (17) into equation (15) gives:

\[ \bar{X} - TX = (\bar{A}T-TA)F + (\bar{A}^2T-\bar{T}A^2)F + \ldots \]  

...(18)

successive terms of which converge to zero. The approximation

5. The term "aggregation bias" was first used by Theil [(1957), p. 116] although the concept had been discussed earlier by Leontief (1951), Hatanaka (1952), Holzman (1953), Barna (1954), Malinvaud (1954), Balderston and Whiten (1954), Morgenstern and Whiten (1955), McManus (1956a) and (1956b) and Fei (1956).
obtained by truncating second and higher order terms:
\[ \bar{X} - TX = (\bar{A}T - TA)F \] ... (19)
is known as first-order aggregation bias [Theil (1957), p. 117].
A grouping of industries such that first-order aggregation bias vanishes is satisfactory for most purposes, even if total aggregation bias is not zero, since second and higher order terms of equation (18) are usually small.

2.3 The Conditions under which Aggregation Bias Vanishes

The expression for aggregation bias, given in equation (15), can be expressed as:
\[ \bar{X} - TX = (I - \bar{A})^{-1}TF - T(I - A)^{-1}F \]
\[ = (I - \bar{A})^{-1}(T(I - A) - (I - \bar{A})T)(I - A)^{-1}F \]
\[ = (I - \bar{A})^{-1}(\bar{A}T - TA)(I - A)^{-1}F \] ... (20)
Hence, aggregation bias vanishes if:
\[ \bar{A}T = TA \] ... (21)

Equation (21) provides us with a necessary and sufficient condition for perfectly consistent aggregation; that industries grouped into the same sector require equal, aggregate inputs, from industries within sectors, in order to produce one unit.

---

6. This condition was originally derived by Hatanaka [(1952), p. 302] for an input-output model with no intraindustry transactions, that is, X is a vector of net outputs and A has a leading diagonal of zeros. Later McManus [(1956b), p. 484] proved that equation (21) is valid only for a model in which intraindustry transactions are included and X is a vector of gross outputs.
of output. In other words, the following equalities must hold

\[ \sum_{i} a_{ij} = \sum_{i} a_{ik} = a_{1l} \] for all \( j, k \) in sector 1; \( j \neq k \)
for all \( i \) in sector 1

\[ \sum_{i} a_{ij} = \sum_{i} a_{ik} = a_{12} \] for all \( j, k \) in sector 2; \( j \neq k \)
for all \( i \) in sector 1

\[ \vdots \]

\[ \sum_{i} a_{ij} = \sum_{i} a_{ik} = a_{lm} \] for all \( j, k \) in sector \( m \); \( j \neq k \)
for all \( i \) in sector 1

\[ \sum_{i} a_{ij} = \sum_{i} a_{ik} = a_{21} \] for all \( j, k \) in sector 1; \( j \neq k \)
for all \( i \) in sector 2

\[ \sum_{i} a_{ij} = \sum_{i} a_{ik} = a_{22} \] for all \( j, k \) in sector 2; \( j \neq k \)
for all \( i \) in sector 2

\[ \vdots \]

\[ \sum_{i} a_{ij} = \sum_{i} a_{ik} = a_{2m} \] for all \( j, k \) in sector \( m \); \( j \neq k \)
for all \( i \) in sector 2

\[ \vdots \]

\[ \sum_{i} a_{ij} = \sum_{i} a_{ik} = a_{ml} \] for all \( j, k \) in sector \( m \); \( j \neq k \)
for all \( i \) in sector \( m \)

\[ \sum_{i} a_{ij} = \sum_{i} a_{ik} = a_{m2} \] for all \( j, k \) in sector 2; \( j \neq k \)
for all \( i \) in sector \( m \)

\[ \vdots \]

\[ \sum_{i} a_{ij} = \sum_{i} a_{ik} = a_{mm} \] for all \( j, k \) in sector \( m \); \( j \neq k \)
for all \( i \) in sector \( m \)

where \( A = \{a_{ij}\} \), \( \tilde{A} = \{\tilde{a}_{ij}\} \) and \( \sum_{i} a_{ij} \) and \( \sum_{i} a_{ik} \) are known as the partially aggregated input coefficients
When the partially aggregated input coefficients are equal, the technical coefficients of the aggregated model are not affected by changes in the product mix. For example, suppose \( A \) is a 4x4 matrix of technical coefficients of the form:

\[
A = \begin{pmatrix}
a_{11} & a_{12} & a_{13} & a_{14} \\
a_{21} & a_{22} & a_{23} & a_{24} \\
a_{31} & a_{32} & a_{33} & a_{34} \\
a_{41} & a_{42} & a_{43} & a_{44} \\
\end{pmatrix}
\]

and industries 1 and 2 are grouped to form sector 1, while industries 3 and 4 are grouped to form sector 2, using the aggregation operator:

\[
T = \begin{pmatrix}
1 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 \\
\end{pmatrix}
\]

Therefore,

\[
TA = \begin{pmatrix}
a_{11} + a_{21} & a_{12} + a_{22} & a_{13} + a_{23} & a_{14} + a_{24} \\
a_{31} + a_{41} & a_{32} + a_{42} & a_{33} + a_{43} & a_{34} + a_{44} \\
\end{pmatrix}
\]

The aggregated matrix of technical coefficients is of the form:

\[
\tilde{A} = \begin{pmatrix}
\tilde{a}_{11} & \tilde{a}_{12} \\
\tilde{a}_{21} & \tilde{a}_{22} \\
\end{pmatrix}
\]

where

\[
\tilde{a}_{11} = \frac{(a_{11} + a_{21})X_1 + (a_{12} + a_{22})X_2}{X_1 + X_2}
\]

\[
\tilde{a}_{12} = \frac{(a_{13} + a_{23})X_3 + (a_{14} + a_{24})X_4}{X_3 + X_4}
\]

\[
\tilde{a}_{21} = \frac{(a_{31} + a_{41})X_1 + (a_{32} + a_{42})X_2}{X_1 + X_2}
\]

\[
\tilde{a}_{22} = \frac{(a_{33} + a_{43})X_3 + (a_{34} + a_{44})X_4}{X_3 + X_4}
\]
Therefore,

\[
\tilde{A}T = \begin{pmatrix}
\tilde{a}_{11} & \tilde{a}_{11} & \tilde{a}_{12} & \tilde{a}_{12} \\
\tilde{a}_{21} & \tilde{a}_{21} & \tilde{a}_{22} & \tilde{a}_{22}
\end{pmatrix}
\]

Consequently, \( \tilde{A}T = TA \) if the partially aggregated input coefficients are equal, that is, if:

\[
\begin{align*}
a_{11} + a_{21} &= \bar{a}_{11} + \bar{a}_{22} \\
a_{13} + a_{23} &= a_{14} + a_{24} \\
a_{31} + a_{41} &= a_{32} + a_{42} \\
a_{33} + a_{43} &= a_{34} + a_{44}
\end{align*}
\]

in which case the aggregate technical coefficients are not affected by changes in the product mix.

The equalities of equation (22) are satisfied when industries, which are consolidated into a single sector, have identical input coefficients, but this condition, though sufficient, is not necessary for perfectly consistent aggregation. 7

Perfectly consistent aggregation is also achieved by aggregating two industries, providing the output of one industry is wholly absorbed as an input into the second industry, [Holzman (1953), p. 330 and Barna (1954), p. 180]. For example, an industry mining iron ore might be aggregated with an industry producing iron and steel, provided that the entire output of the iron mining industry is an input into the iron and steel producing industry. Under this condition the coefficients of the aggregated model are not affected by changes in the product mix. For example, suppose industry 1 sells its entire output to industry 2. The matrix of technical coefficients is given by:

\[
\begin{pmatrix}
0 & a_{12} & 0 & \ldots & 0 \\
a_{21} & a_{22} & a_{23} & \ldots & a_{2n} \\
a_{31} & a_{32} & a_{33} & \ldots & a_{3n} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
a_{n1} & a_{n2} & a_{n3} & \ldots & a_{nn}
\end{pmatrix}
\]

and the matrix of aggregate technical coefficients is given by:

\[
\bar{A} = \begin{pmatrix}
a_{21} \frac{x_1 + (a_{12} + a_{22})x_2}{X_1 + X_2} & a_{23} & \ldots & a_{2n} \\
a_{31} \frac{x_1 + a_{32}x_2}{X_1 + X_2} & a_{33} & \ldots & a_{3n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{n1} \frac{x_1 + a_{n2}x_2}{X_1 + X_2} & a_{n3} & \ldots & a_{nn}
\end{pmatrix}
\]

where \(X_1\) is the output of industry 1, and \(X_2\) is the output of industry 2, in the base period.
Now, if industry 1 sells all its output to industry 2, then \( a_{12}X_2 = X_1 \). Consequently, the matrix of aggregate technical coefficients takes the form:

\[
\bar{A} = \begin{pmatrix}
\frac{a_{21}a_{12} + a_{12} + a_{22}}{a_{12} + 1} & a_{23} & \cdots & a_{2n} \\
\frac{a_{31}a_{12} + a_{32}}{a_{12} + 1} & a_{33} & \cdots & a_{3n} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{a_{n1}a_{12} + a_{n2}}{a_{12} + 1} & a_{n3} & \cdots & a_{nn}
\end{pmatrix}
\]

the elements of which are not affected by the product mix of industries 1 and 2 and therefore remain stable, provided the elements of \( A \) are stable.

Aggregation bias was given in equation (20) as:

\[
\bar{x}^{-}TX = (I-\bar{A})^{-1}(\bar{A}T - TA)(I-A)^{-1}F
\]

and this simplifies to:

\[
\bar{x}^{-}TX = (I-\bar{A})^{-1}(\bar{A}T - TA)X
\]

If industry 2 absorbs the entire output of industry 1, \( X \) is a column vector, the first two elements of which are \( a_{12}X_2 \) and \( X_2 \), respectively. Under these conditions, \( \bar{A}TX \) equals \( TAX \) and so aggregation bias vanishes for any final demand vector, the first element of which is equal to zero.

This result can be extended to more than two industries. For example, industries 1, 2 and 3 can be aggregated, without introducing aggregation bias, if industry 1 sells all its output to industry 2, and if industry 2 sells all its output to industry 3. In general, if industries 1, 2, \ldots, \( k \) are such
that industry \(i\) sells all its output to industry \(i+1\) 
\((i=1,2,\ldots,k-1)\), then aggregation bias vanishes for any final 
demand vector, the first \(k-1\) elements of which are equal to 
zero.

If final demands are proportional to those of the base 
period, then industries may be aggregated in any manner at all 
without producing biased results [Balderston and Whiten (1954), 
p. 108; Theil (1957), p. 120; Neudecker (1970), p. 922]. With 
fixed technical coefficients, the only source of change in the 
product mix is gross output changes. However, if new final 
demands are proportional to base period final demands, then 
the corresponding gross outputs are proportional to gross 
outputs in the base period, and so the product mix remains 
stable.

The formal proof of this condition begins with a 
derivation of an expression for the matrix of aggregate 
technical coefficients. Conventional aggregation of \(n\) 
industries into \(m\) sectors leads to:

\[
\bar{Y}_o = TY_o T' \tag{23}
\]

\[
\bar{X}_o^D = TX_o^D T' \tag{24}
\]

\[
\bar{Y}_o = \bar{A}X_o^D \tag{25}
\]

where \(Y_o\) is an \(nxn\) matrix of interindustry transaction flows 
in the base period and 
\(Y_o = AX_o^D \tag{26}\)

8. The "doubting Thomas" will find empirical verification of 
this condition in an article by Hewings [(1972), p. 17].
\( \bar{Y}_o \) is an \( m \times m \) matrix of intersector transaction flows in the base period.

\( X_o^D \) is an \( n \times n \) diagonal matrix, the elements of which are gross outputs of industries in the base period.

\( \bar{X}_o^D \) is an \( m \times m \) diagonal matrix, the elements of which are sector gross outputs in the base period.

Equation (25) implies that:

\[
\bar{A} = \bar{Y}_o (\bar{X}_o^D)^{-1} \quad \ldots (27)
\]

Substituting equations (23) and (24) into equation (27) gives:

\[
\bar{A} = TY_o T' (TX_o^D T')^{-1} \quad \ldots (28)
\]

Finally, substituting equation (26) into equation (28) gives:

\[
\bar{A} = TAX_o^D T' (TX_o^D T')^{-1} \quad \ldots (29)
\]

or

\[
\bar{A} = TAG_o \quad \ldots (30)
\]

where

\[
G_o = X_o^D T' (TX_o^D T')^{-1} \quad \ldots (31)
\]

and

\[
T G_o = I \quad \ldots (32)
\]

Equation (29) states that input coefficients to each sector are weighted averages of the partially aggregated input coefficients of industries within sectors. The weights are base period outputs of the constituent industries, expressed as proportions of the total base period output of the sector, as was demonstrated in Section 1.3 above.

The expression for \( \bar{A} \), given in equation (30), is substituted into equation (20) and aggregation bias becomes:

\[
\bar{X} - TX = (I - \bar{A})^{-1} (TAG_o T - TA)(I - A)^{-1} F
\]

\[
= (I - \bar{A})^{-1} TA (G_o T - I)(I - A)^{-1} F \quad \ldots (33)
\]

Now, if final demands are proportional to those of the base period (that is, if \( F = \phi F_o \) where \( \phi \) is a scalar) then
aggregation bias is:
\[ \tilde{X} - TX = \phi(I-A)^{-1}TA(G_0T - I)(I-A)^{-1}F \]
\[ = \phi(I-A)^{-1}TA(G_0T - I)X_o \quad \text{[by equation (10)]} \]
\[ = \phi(I-A)^{-1}TA(G_0TX_o - X_o) \quad \ldots(34) \]
But, by equation (31):
\[ G_0TX_o = X_o D T' (TX_oD T')^{-1}TX_o \quad \ldots(35) \]
where \((TX_oD T')^{-1}TX_o = i_m\) (an mx1 vector of units). Also, \(T'i_m = i_n\)
and \(X_o D i_n = X_o\) so:
\[ G_0TX_o = X_o \quad \ldots(36) \]
Finally, substituting equation (36) into equation (34), it is evident that aggregation bias vanishes when final demands all change by the same proportion. Note that when \(\phi=1\) final demands equal those of the base period. Consequently, there is no aggregation bias in the base period.

Perfectly consistent aggregation also results when industries, which are combined to form a single sector, are strictly complementary in the sense that their outputs are always used in fixed proportions, both by other industries and by final consumers. This condition is stated mathematically by the two equations:
\[ G_0 \bar{F} = F \quad \ldots(37) \]
and \[ G_0TA = A \quad \ldots(38) \]

9. Note that equation (37) does not contradict the aggregation rule, \(TF=\bar{F}\), since \(TF = TG_0 \bar{F}\) which equals \(\bar{F}\) since \(TG_0 = I\).
When equations (37) and (38) hold, the product mix in each sector is perfectly stable and so are the aggregate technical coefficients.\(^{10}\) Aggregation bias vanishes since the substitution of \(A = G_0TA\) and \(F = G_0\vec{F}\) into \(TX = T(I-A)^{-1}F\) gives:

\[
TX = T(I - G_0TA)^{-1}G_0\vec{F}
\]

\[
= T[I + (G_0TA) + (G_0TA)^2 + ... ]G_0\vec{F}
\]

\[
= \{TG_0 + T(G_0TA)G_0 + T(G_0TA)^2G_0 + ... \}\vec{F}
\]

and since \(\vec{A} = TAG_0\) [by equation (30)] and \(TG_0 = I\) [by equation (32)]

\[
TX = (I + \vec{A} + \vec{A}^2 + ... )\vec{F}
\]

\[
= (I - \vec{A})^{-1}\vec{F}
\]

\[
= \vec{x}
\]

Furthermore, when industry outputs are used in fixed proportions, they can be calculated from sector outputs, using \(X = G_0\vec{X}\) since:

\[
G_0\vec{X} = G_0(I - \vec{A})^{-1}\vec{F}
\]

\[
= G_0[I + \vec{A} + \vec{A}^2 + ... ]\vec{F}
\]

\[
= G_0[I + (TAG_0) + (TAG_0)^2 + ... ]\vec{F}
\]

\[
= \{I + (G_0TA)G_0 + (G_0TA)^2G_0 + ... \}G_0\vec{F}
\]

\[
= (I + A + A^2 + ... )F \quad \text{[from equations (37) & (38)]}
\]

\[
= (I - A)^{-1}F
\]

\[
= x
\]

\(^{10}\) This condition was stated by Holzman [(1953), p. 327], Dorfman, Samuelson and Solow [(1958), p. 243] and Chenery and Clark [(1959), pp. 35-36]. It was proved by Malinvaud [(1954), p. 199] and Stone [(1961), p. 103] although their proofs differ from that given here.
Finally, Theil [(1957), pp. 120-121] put forward a procedure which he called "perfect aggregation". It entails replacing the matrix $G_o$ in equation (30) by the matrix:

$$G = x^D T' (T x^D T')^{-1} \quad \ldots (39)$$

where $x^D$ is an nxn diagonal matrix, the elements of which are industry gross outputs in the current period.

However, current outputs are unknown at the time aggregation takes place (indeed, if they are known there is no need to forecast sector outputs) so this is no solution to the aggregation problem. Stone [(1961), p. 104], however, when discussing perfect aggregation, points out that in some cases it may be possible to update $G_o$, in which case aggregation bias might be reduced.

In summary, consistent aggregation of $n$ industries into $m$ sectors is possible under the following conditions:

1. When industries, which are grouped into a single sector, have equal, aggregate, per unit input requirements from industries within sectors (that is, equal partially aggregated input coefficients).

2. When industries, which are aggregated into the same sector, have identical input coefficients.

3. When $k$ industries, which are aggregated into the same sector, are such that the entire output of industry $i$ is absorbed by industry $i+1$ ($i=1, 2, \ldots, k-1$).

4. When final demands are equal, or proportional, to base period final demands.
(5) When the products of industries, which are grouped into the same sector, are always used in fixed proportions by other industries and by final consumers.

Each of these conditions is extremely stringent and it is unlikely that any of them will be fulfilled exactly in practice. However, they do suggest strategies, which might be followed when aggregation is necessary.

Condition (2) is more binding than condition (1) but both imply that industries with similar input structures are potential candidates for aggregation, for example, cars and trucks, or mutton and wool. However, it is quite possible that more unlikely combinations of industries may have similar input structures. In a large input-output model, it is extremely difficult to determine which industries best approximate condition (1) (particularly) or condition (2), simply by scanning its technical coefficients.

Condition (3) is useful in determining which industries can be aggregated safely for the input-output analyst is likely to have a priori knowledge of industries, which are likely to satisfy this condition. Verification is easily performed using the interindustry table.

Final demands must be specified for each industry before condition (4) can be tested. If it is satisfied, then any aggregation pattern is admissible. However, it is more likely that the intersector model will need to be determined before final demands are known. In such cases, there is no
way of knowing whether or not condition (4) is satisfied when the grouping of industries needs to be undertaken.

Condition (5) suggests that industries producing complementary goods, such as nuts and bolts, or skis and bindings, might be aggregated. However, less obvious combinations, which satisfy this requirement approximately, will be difficult to find in a large input-output model, without a systematic search procedure. Utilisation of this condition also requires some knowledge of final demands when the intersector model is constructed.

2.4 The Conditions under which First-Order Aggregation Bias Vanishes

In Section 2.2, first-order aggregation bias was defined as:

\[(\bar{A}T - TA)F = (I + \bar{A})TF - T(I + A)F \quad \ldots(40)\]

There are a number of conditions under which expression (40) is a zero vector.

Firstly, it is evident that first-order aggregation bias vanishes if \(\bar{A}T = TA\). This condition has been discussed already in Section 2.3.

Secondly, if final demands are proportional to base period outputs, that is, if \(F = \phi X_0\) where \(\phi\) is a scalar, then first-order aggregation bias vanishes since:
$$\begin{align*}
(\widetilde{AT} - TA)F &= \phi(\widetilde{AT} - TA)X_O \\
&= \phi(TAG_oT - TA)X_O \quad \text{[from equation (30)]} \\
&= \phi TA(G_oTX_o - X_o) \\
&= 0 \quad \text{[from equation (36)]}
\end{align*}$$

However, since final demands are usually unknown at the time when aggregation is required, this condition is of limited use.

Theil has derived two conditions, under which first-order aggregation bias vanishes, and which can best be seen by examining the elements of the vector $(\widetilde{AT}-TA)F$, as given in Table 2.4.1 below.

**Table 2.4.1**

<table>
<thead>
<tr>
<th>First-Order Aggregation Bias as a Vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_1 (\bar{a}<em>{11} - \Sigma a</em>{ij})F_j + \Sigma (\bar{a}<em>{12} - \Sigma a</em>{ij})F_j + \ldots + \Sigma (\bar{a}<em>{1m} - \Sigma a</em>{ij})F_j$</td>
</tr>
<tr>
<td>$\pi_2 (\bar{a}<em>{21} - \Sigma a</em>{ij})F_j + \Sigma (\bar{a}<em>{22} - \Sigma a</em>{ij})F_j + \ldots + \Sigma (\bar{a}<em>{2m} - \Sigma a</em>{ij})F_j$</td>
</tr>
<tr>
<td>$\ldots$</td>
</tr>
<tr>
<td>$\Sigma (\bar{a}<em>{m1} - \Sigma a</em>{ij})F_j + \Sigma (\bar{a}<em>{m2} - \Sigma a</em>{ij})F_j + \ldots + \Sigma (\bar{a}<em>{mm} - \Sigma a</em>{ij})F_j$</td>
</tr>
</tbody>
</table>

where $\pi_1, \pi_2, \ldots, \pi_m$ represent sectors 1, 2, ... m. The notations $\Sigma$ and $\Sigma$ represent summation over all industries in sector $\pi$. $j$ and $i$

If each of the $m^2$ terms, which form this vector, are zero then first-order aggregation bias vanishes [Theil (1957), p. 119]. Alternatively, if these terms are zero in the base period and
industries within sectors have final demands which are proportional to those of the base period, that is, if:

\[ F_j = \phi_1 F_{j0} \quad \text{for all } j \text{ in sector 1} \]
\[ F_j = \phi_2 F_{j0} \quad \text{for all } j \text{ in sector 2} \]
\[ \vdots \]
\[ F_j = \phi_m F_{j0} \quad \text{for all } j \text{ in sector } m \]

(where \( \phi_1, \phi_2, \ldots, \phi_m \) are scalars)

then first-order aggregation bias vanishes [Theil (1957), p. 120]. Both these conditions require a knowledge of final demands at the time when aggregation takes place, which limits their usefulness. Even if final demands are known, these conditions do little to suggest an appropriate pattern of aggregation.

First-order aggregation bias vanishes if the outputs of industries grouped into the same sector are always used in fixed proportions, since:

\[ T(I + A)F = T(I + G_O TA)G_O F \quad \text{[from equations (37) & (38)]} \]
\[ = (TG_O + TG_O TAG_O)F \]
\[ = (I + TAG_O)F \quad \text{[from equation (32)]} \]
\[ = (I + A)F \quad \text{[from equation (30)]} \]
\[ = (I + A)TF \]

This condition has been discussed already in Section 2.3.

Morimoto [(1970), pp. 121-122] has proved that first-order aggregation bias vanishes if the structure of final demands within each sector corresponds to that of outputs in
the base period, that is, if \( F = G_o \bar{F} \), since:

\[
T(I + A)F = T(I + A)G_o \bar{F} \\
= (TG_o + TAG_o)\bar{F} \\
= (I + \bar{A})\bar{F} \quad \text{[from equations (32) & (30)]} \\
= (I + \bar{A})TF
\]

This condition is weaker than the condition of fixed proportional usage but it still requires that final demands are known when industries are grouped into sectors.

A second theorem by Morimoto [(1970), p. 122] states that if some industries are not aggregated and the changes in final demand occur only in the unaggregated industries, then first-order aggregation bias vanishes, regardless of the way in which the other industries are aggregated. The final demand vector takes the form:

\[
F = \begin{pmatrix}
F_1 \\
F_2 \\
\vdots \\
F_n
\end{pmatrix}
\]

where \( F_i = F_{i0} \) in aggregated industries, and \( F_i = F_{i0} + \Delta F_i \) in industries which are not aggregated.

Hence, the vector \( \Delta F = F - F_o \) contains zero elements except for those corresponding to industries which are not aggregated. Morimoto proceeds to prove that \((\bar{A}T - TA)\Delta F\) equals zero by

11. An empirical test of this theorem can be found in Hewings [(1972), p. 19]. Hewings aggregated a 51 industry model in successive stages to form a four sector model, three of which were left unaggregated. The amount of aggregation bias in the solutions was found to be very small.
observing that the columns of $\bar{A}_T$ and $T_A$, which correspond to the unaggregated industries, are equal and hence the same columns of $\bar{A}_T - T_A$ contain only zero elements. The only nonzero elements in $\Delta F$ are those corresponding to the unaggregated industries, so $(\bar{A}_T - T_A)\Delta F = 0$.

Since $\Delta F = F - F_0$ and $(\bar{A}_T - T_A)\Delta F = 0$,

$$(\bar{A}_T - T_A)F = (\bar{A}_T - T_A)F_0 \quad \ldots (41)$$

But total aggregation bias vanishes in the base period, so, using equation (18),

$$(\bar{A}_T - T_A)F_0 + (\bar{A}^2T - T_A^2)F_0 + (\bar{A}^3T - T_A^3)F_0 + \ldots = 0$$

Therefore,

$$(\bar{A}_T - T_A)F_0 = -(\bar{A}^2T - T_A^2)F_0 - (\bar{A}^3T - T_A^3)F_0 - \ldots \quad \ldots (42)$$

Total aggregation bias in any period is given by:

$$\bar{x}_{TX} = (\bar{A}_T - T_A)F + (\bar{A}^2T - T_A^2)F + (\bar{A}^3T - T_A^3)F + \ldots$$

which, by equation (41), equals:

$$\bar{x}_{TX} = (\bar{A}_T - T_A)F_0 + (\bar{A}^2T - T_A^2)F + (\bar{A}^3T - T_A^3)F + \ldots$$

and, by equation (42), equals:

$$\bar{x}_{TX} = (\bar{A}^2T - T_A^2)(F - F_0) + (\bar{A}^3T - T_A^3)(F_0 - F_0) + \ldots$$

so aggregation bias is of the second order.

This is a useful result for if one wishes to assess the effect of changes in final demands of just a few industries from their base period values, the remaining industries may be aggregated in any manner without introducing a large amount of bias into the solution.
2.5 Aggregation Bias in Output and Income Multipliers

The bulk of the literature dealing with aggregation bias in input-output analysis has been concerned with inconsistencies in forecasts of gross output, derived from an original and aggregated model. It would appear that only two studies have considered the effects of aggregation on output and income multipliers. The first is that of Doeksen and Little (1968) who, using four empirical input-output models, observed that the output and income multipliers of three industries, which were left unaggregated, changed very little as the remaining industries were aggregated, two at a time, until they formed one composite sector.

In the case of the output multipliers, the result is not surprising [see Rodgers (1977), pp. 154-155]. Output multipliers from an original n industry model are given by:

\[ m' = i_n'(I-A)^{-1} \]

where \( i_n' \) is a row vector of \( n \) units.

Output multipliers from an aggregated \( m \) sector model, the first \( k \) sectors of which contain a single industry, are given by:

\[ m_1' = i_m'(I-\bar{A})^{-1} \]

where \( i_m' \) is a row vector of \( m \) units.

Since \( i_n' = i_m'T \) and \( (I-\bar{A})^{-1}T \) is \( (I-\bar{A})^{-1} \) with the last column repeated \( n-k \) times, the difference between output multipliers of the \( k \) single-industry sectors, as derived from
the original and aggregated model, form the first $k$ elements of the vector:

$$\bar{m}' - m' = i'_m (I - \bar{A})^{-1} T - i'_m T (I - A)^{-1} \quad \ldots \ (45)$$

Expression (45) can be expanded as:

$$\bar{m}' - m' = i'_m \{(\bar{A}T - TA) + (\bar{A}^2 T - TA^2) + \ldots \} \quad \ldots \ (46)$$

Morimoto [(1970), p. 122] has pointed out that the columns of $\bar{A}T$ and $TA$, which correspond to the unaggregated industries, are equal, and therefore these same columns of $(\bar{A}T - TA)$ are zero vectors. Therefore, the difference between output multipliers for these industries, obtained from the original and aggregated models, is of the second order only.

The second result is more difficult to explain. Income multipliers obtained from the original model are given by:

$$m'_Y = w'_I (I - A)^{-1} \quad \ldots \ (47)$$

where $w'_I$ is a $1 \times n$ row vector, the elements of which are income arising per unit of output for each industry.

Income multipliers, obtained from the aggregated model, are given by:

$$\bar{m}'_Y = \bar{w}' (I - A)^{-1} \quad \ldots \ (48)$$

where $\bar{w}'$ is a $1 \times m$ row vector, the elements of which are income arising per unit of output for each sector, and

$$\bar{w}' = w'G_o \quad \ldots \ (49)$$

Since postmultiplication of equation (48) by $T$ simply results in the last column being repeated $n-k$ times, aggregation bias in the income multipliers of the $k$ unaggregated industries forms the first $k$ elements of the vector:
\[
\begin{align*}
\tilde{m}_Y - m_Y &= \tilde{w}'(I-A)^{-1}T - w'(I-A)^{-1} \\
&= w'G_o(I+\tilde{A}+\tilde{A}^2+\ldots)T - w'(I+\tilde{A}+\tilde{A}^2+\ldots) \\
&= w'\{(G_oT-I)+(G_o\tilde{A}T-A)+(G_o\tilde{A}^2T-A^2)+\ldots\} \quad (50)
\end{align*}
\]

Although the columns of \(G_oT\) and \(I\) which correspond to the unaggregated industries are equal, the same columns of the second term of equation (50) are equal only under the following conditions:

(a) If the technical coefficients of income arising in the aggregated industries are all equal.

(b) If the products of industries, which are aggregated into the same sector, are used in fixed proportions by all other industries. In this case \(G_oTA=A\) and

\[
G_o\tilde{A}T - A = G_oTA(G_oT-I)
\]

and the columns of \(G_oT\) and \(I\), which correspond to the unaggregated industries, are equal.

So, in general, aggregation bias in income multipliers is of the first order. It is therefore surprising that Doekszen and Little observed such a small difference between income multipliers of the unaggregated industries in the original and aggregated models.

2.6 Current Aggregation Procedures

The first author to suggest a formal procedure for grouping industries into sectors was Fisher (1958b). His approach was extended by Neudecker (1970). Their contributions
are reviewed in Section 2.6.1 below, under the heading of "non clustering" methods of aggregation, to distinguish them from the methodology of Mukherjee (1970), Kossov (1972) and Blin and Cohen (1977), all of whom employ "clustering" algorithms to aggregate industries. A review of these three papers is preceded by a brief description of the technique known as "cluster analysis".

2.6.1 Non Clustering Methods of Aggregation

Fisher [(1958b), p. 251] distinguishes between "special purpose" aggregation, where the objective is to predict the output of a single key industry for a given set of final demands, and "general purpose" aggregation, where output predictions for all industries are required. In either case, Fisher assumes that final demands are unknown at the time when aggregation is required, and consequently, they may be regarded as random variables. Since the aggregation bias associated with each sector is a function of the final demands, it too is a random variable.

With special purpose aggregation, Fisher's method entails grouping industries, other than the key industry, into m-1 sectors, in such a way that the expected squared

12. Fisher's methodology was converted from algebraic to matrix notation by Neudecker [(1970), p. 921-922]. Both notations will be used in this section, depending upon which is the more convenient and illuminating form in which to work.
aggregation error, associated with the key industry, is
minimized. In other words, the problem is to find an
aggregation operator, $T$, such that
\[
\lambda_i = E\{(\bar{x}-TX)(\bar{x}-TX)\} \quad \ldots (51)
\]
is minimized. The subscript $i$ refers to the $i^{th}$ element in
the leading diagonal of the matrix on the right hand side of
equation (51), the $i^{th}$ industry being the key industry. The
matrix, $T$, is as defined in equation (11), except that row $i$
contains a single nonzero element, equal to unity, in column $i$.
If the matrix, $T$, which minimizes $\lambda_i$, results in large
aggregation errors for the other sectors, this is irrelevant.

Morimoto's second theorem, stated in Section 2.4, is
pertinent to special purpose aggregation, since the key
industry is left unaggregated. Consequently, if the only
final demand which changes is that of the key industry, then
first-order aggregation bias vanishes, and total aggregation
bias is small, regardless of the way in which the other
industries are grouped. However, if some of the other final
demands change, then aggregation bias again becomes a problem.

With general purpose aggregation, Fisher's method
involves grouping the $n$ industries into $m$ sectors such that
the expected sum of squared aggregation errors, for all $m$
sectors, is minimized. The objective, in this case, is to find
an aggregation operator, $T$, which minimizes
\[
\lambda = \text{tr}\{E\{(\bar{x}-TX)(\bar{x}-TX)\}'\} \quad \ldots (52)
\]
Here, aggregation errors in each of the sectors are considered
to be equally important.
Minimization of $\lambda_1$ or $\lambda$ requires some specific assumptions about the distribution of final demands. Fisher makes two assumptions:

(a) that the expected value of each industry's final demand equals its value in the base period, that is:

$$E(F) = F_o$$ ...(53)

(b) that final demands are uncorrelated and have variances which are proportional to their base period values, that is:

$$E\{(F-F_o)(F-F_o)\}' = \phi F_o^D$$ ...(54)

where $\phi$ is a scalar and $F_o^D$ is a diagonal matrix, the elements of which are base period final demands.

Therefore,

$$E(FF') = \phi F_o^D + F_o F_o'$$ ...(55)

Substituting equation (15) and (55) into equation (52) gives:

$$\lambda = \text{tr}\{(I-\bar{A})^{-1}_T - T(I-A)^{-1}\}E(FF')\{(I-\bar{A})^{-1}_T - T(I-A)^{-1}\}'$$

$$= \text{tr}\{(I-\bar{A})^{-1}_T - T(I-A)^{-1}\}(\phi F_o^D + F_o F_o')\{(I-\bar{A})^{-1}_T - T(I-A)^{-1}\}'$$

$$= \phi \text{tr}\{(I-\bar{A})^{-1}_T - T(I-A)^{-1}\}F_o^D\{(I-\bar{A})^{-1}_T - T(I-A)^{-1}\}'$$

$$= \phi \{(I-\bar{A})^{-1}_T - T(I-A)^{-1}\}F_o^D\{(I-\bar{A})^{-1}_T - T(I-A)^{-1}\}$$

since aggregation bias vanishes in the base period.\[13\]

---

13. \{(I-\bar{A})^{-1}_T - T(I-A)^{-1}\}(\phi F_o^D + F_o F_o')\{(I-\bar{A})^{-1}_T - T(I-A)^{-1}\}'

$$= \phi \{(I-\bar{A})^{-1}_T - T(I-A)^{-1}\}F_o^D\{(I-\bar{A})^{-1}_T - T(I-A)^{-1}\}' +$$

$$\{(I-\bar{A})^{-1}_T - T(I-A)^{-1}\}F_o^D\{(I-\bar{A})^{-1}_T - T(I-A)^{-1}\}'$$

$$= \phi \{(I-\bar{A})^{-1}_T - T(I-A)^{-1}\}F_o^D\{(I-\bar{A})^{-1}_T - T(I-A)^{-1}\}'$$

since in Section 2.3 it was proved that:

$$\{(I-\bar{A})^{-1}_T - T(I-A)^{-1}\}F_o = 0$$
Similarly, substituting equations (15) and (55) into equation (51) gives:

\[
\lambda_i = \{(I-\bar{A})^{-1}T - T(I-A)^{-1}\}E(FF')\{(I-\bar{A})^{-1}T - T(I-A)^{-1}\}'\}
i
\[
= \phi \{(I-\bar{A})^{-1}T - T(I-A)^{-1}\}P_D\{(I-\bar{A})^{-1}T - T(I-A)^{-1}\}'\}
i
\[
= \phi \sum_{J=1}^{m} \sum_{j}^{J} (b_{iJ} - b_{ij})^2 F_{0j} \ldots (57)
\]

where \(b_{iJ}\) is the \(J^{th}\) element in row \(i\) of the matrix \((I-\bar{A})^{-1}\),

\(b_{ij}\) is the element corresponding to the \(j^{th}\) industry
of sector \(J\) in row \(i\) of the matrix \((I-A)^{-1}\).

To avoid as much computation as possible,\(^14\) Fisher

proposed the use of two approximate criteria for both \(\lambda_i\) and \(\lambda\). The first approximation is derived by replacing the

interdependence coefficients of the aggregated model with

weighted averages of the partially aggregated interdependence

coefficients of the original model. The weights are base period

final demands of industries within sectors, expressed as

proportions of the total base period final demand for the

sector. Hence, the matrix \((I-\bar{A})^{-1}\) in equation (56) is replaced

by:

\[
(I-\bar{A})^{-1} \approx T(I-A)^{-1}P_D T' (T F_D T')^{-1} \ldots (58)
\]

and equivalently, the coefficients \(b_{iJ}\) in equation (57) are

replaced by:

\[
b_{iJ} \approx \bar{b}_{iJ} = \frac{\sum_{J}^{m} F_{0j} b_{ij}}{\sum_{J}^{m} F_{0j}} \ldots (59)
\]

\(^{14}\) The need to avoid computation was, of course, far more

important at the time that Fisher was writing than it is today.
Substituting equation (58) into equation (56) gives:

$$\lambda \leftrightarrow \lambda' = \phi \text{tr}\{\{T(I-A)^{-1}F_o^D\}(I - T'(T_P^D T')^{-1}F_P^D\)}\{T(I-A)^{-1}\}'\} \ldots (60)$$

and substituting equation (59) into equation (57) gives:

$$\lambda_i \leftrightarrow \lambda_i' = \phi \sum_{j=1}^{m} \sum_{j} (b_{ij} - b_{ij})^2 F_{oj} \ldots (61)$$

both of which avoid the calculation of $\bar{A}$ and $(I-\bar{A})^{-1}$.

In Fisher's second approximation, the matrices $(I-A)^{-1}$ and $(I-\bar{A})^{-1}$ are replaced by $^{15}$ $(I+A)$ and $(I+\bar{A})$, the technical coefficients of the aggregated model are approximated by unweighted averages of the partially aggregated technical coefficients of the original model, that is,

$$\bar{A} \leftrightarrow TAT'(TT')^{-1} \ldots (62)$$

and $F_o^D$ is replaced by $I$, implying homoscedastic final demands. Substituting these approximations into equation (56) gives:

$$\lambda \leftrightarrow \lambda'' = \phi \text{tr}\{TA(I - T'(TT')^{-1}I)A'T'\} \ldots (63)$$

Analogously, for special purpose aggregation, $b_{ij}$ in equation (57) is replaced by:

$$\bar{a}_{ij} = \frac{\sum_{j} a_{ij}}{n_j} \ldots (64)$$

where $n_j$ is the number of industries in sector $J$.

---

15. These approximations are justified by the expansion of the inverse matrices as power series:

$$(I-A)^{-1} = I + A + A^2 + \ldots$$

$$(I-\bar{A})^{-1} = I + \bar{A} + \bar{A}^2 + \ldots$$
Hence, equation (57) is approximated by:

\[
\lambda_i \approx \lambda''_i = \phi \sum_{J=1}^{m} \sum_{j} (\tilde{a}_{ij} - a_{ij})^2 f_{oj} \quad \ldots (65)
\]

Equations (63) and (65) avoid the calculation of \((I-A)^{-1}\) as well as \(\overline{A}\) and \((I-\overline{A})^{-1}\).

Fisher [(1958b), p. 255] mentions a procedure, which he developed for finding the "optimal" aggregation scheme for special purpose aggregation. It involves partitioning the \(n\) numbers \(b_{ij}\) in equation (61) or \(a_{ij}\) in equation (65) into \(m\) groups such that the weighted sum of squared deviations from each group mean is as small as possible, the weights being base period final demands. A method of solving such a problem is discussed in another article by Fisher (1958a) and involves a recursive partitioning of the set of numbers into two groups until the desired number of groups is obtained.

In the absence of a similar procedure for use with general purpose aggregation, Fisher performs a partitioning of the diagonal elements of the matrices \((I-A)^{-1}\) and \(A\) into \(m\) groups in order to derive the final sectors.\(^{16}\)

Neudecker [(1970), pp. 923-926] extended Fisher's methodology by treating the matrix \(\overline{A}\) as a variable, rather than being fixed, for a given pattern of aggregation, equal to \(TAG_o\) as in equation (30). He found that the expression for \(\overline{A}\),

\(^{16}\) It was noted that the diagonal elements of these matrices were large compared to the off-diagonal elements.
which minimizes the expected sum of squares of aggregation errors, \( \lambda \), as given in equation (56), is equal to:

\[
\tilde{\lambda} = TA(I-A)^{-1}T_O^D(T(I-A)^{-1}T_O^D)^{-1}
\]

The minimum expected sum of squares of aggregation errors for a given aggregation operator, \( T \), is equal to:

\[
\lambda = tr(T(I-A)^{-1}T_O^D(T^D(I-A)^{-1}T_O^D)^{-1}T^D)\}
\]

Neudecker seems to suggest [(1970), p. 923] that a suitable aggregation scheme can be found by experimenting with various forms of the matrix, \( T \), and choosing the one which minimizes \( \lambda \) in equation (67). The matrix of aggregated technical coefficients is then calculated from equation (66).

In practice, this is an impossible task, for the number of possible ways, in which \( n \) industries can be grouped into \( m \) sectors, is a Stirling number of the second kind, equal to:

\[
\sum_{k=0}^{m} \frac{(-1)^{m-k} C_k}{m!} k^n
\]

For example, the number of possible ways of aggregating 25 industries into five sectors is 2,436,684,974,110,751.

Consequently, it is impossible to investigate more than a small subset of all possible aggregation patterns. No indication is given as to how one might decide which aggregations are potentially promising. Furthermore, the expression, used by Neudecker for the matrix of aggregated technical coefficients, is difficult to interpret, compared to the conventional form

17. Since equations (66) and (67) are a little tedious to derive, their derivation (which is not given by Neudecker) appears in the appendix at the end of this chapter.
given in equation (30).

Finally, the methods of both Fisher and Neudecker depend upon assumptions concerning the behaviour of final demands, which are difficult to test, and which can be questioned on \textit{a priori} grounds.

\textbf{2.6.2 Cluster Analysis}

Cluster analysis encompasses a number of techniques which can be used to sort a set of data units, such as persons, objects etc., into a number of mutually exclusive categories or "clusters". Each data unit is described by an observation on each of a number of variables and all variables are taken into account when data units are sorted into groups. Data units, which are allocated to the same cluster, have a high degree of similarity, measured over all variables, while data units, which are allocated to different clusters, are relatively dissimilar.

Cluster analysis can also be used to sort the variables into groups. Each variable is described by a set of observations over all data units, and variables which are allocated to the same cluster have a high degree of similarity, compared to variables which are allocated to different clusters.

Hence, the aim of cluster analysis is to sort entities (that is, data units or variables) into groups, such that there is a high degree of "natural association" between members of the same group, and a low degree of "natural association"
between members of different groups \citep{Anderberg1973}, p. 3.

Whether the aim is to group data units or variables, a measure of similarity is required for each pair of entities. \footnote{For a full discussion of similarity measures, see Anderberg \citep{Anderberg1973}, chapters 4 and 5 and Duran and Odell \citep{DuranOdell1974}, chapter 1, pp. 3-18.}

When variables are being clustered, the most common measure of similarity is the product moment correlation coefficient:

\[ r_{jk} = \frac{\sum_{i=1}^{m} (x_{ji} - \bar{x}_j)(x_{ki} - \bar{x}_k)}{\sqrt{\sum_{i=1}^{m} (x_{ji} - \bar{x}_j)^2} \sqrt{\sum_{i=1}^{m} (x_{ki} - \bar{x}_k)^2} } \]

where \( x_{ji} \) is the \( i \)th observation on the \( j \)th variable, \( x_{ki} \) is the \( i \)th observation of the \( k \)th variable, \( \bar{x}_j \) is the mean of the \( j \)th variable, \( \bar{x}_k \) is the mean of the \( k \)th variable, and \( m \) is the number of data units.

When data units are being clustered, similarity is measured by a distance function, such as the squared Euclidean distance:

\[ d_{jk} = \sum_{i=1}^{n} (x_{ij} - x_{ik})^2 \]

where \( x_{ij} \) is the \( i \)th observation on the \( j \)th data unit, \( x_{ik} \) is the \( i \)th observation of the \( k \)th data unit, and \( n \) is the number of variables.
Measures of similarity are used to construct a triangular similarity matrix, depicting the strength of association between each possible pair of entities. For example, if there are n entities and $s_{jk}$ represents the similarity measure between the $j^{th}$ and $k^{th}$ entity, then the similarity matrix consists of $\binom{n}{2}$ elements, arranged as follows:

$$
\begin{pmatrix}
  s_{21} \\
  s_{31} & s_{32} \\
  s_{41} & s_{42} & s_{43} \\
  \vdots & \vdots & \vdots \\
  s_{n1} & s_{n2} & s_{n3} & \cdots & s_{n,n-1}
\end{pmatrix}
$$

Beginning with $n$ clusters, each containing one entity, the similarity matrix is searched to find the most similar pair of clusters, (say) $p$ and $q$ ($p > q$). If similarity is defined by a correlation measure, $s_{pq}$ is a maximum. If similarity is defined by a distance measure, $s_{pq}$ is a minimum. Clusters $p$ and $q$ are merged to form a new cluster, which assumes the label $q$, the number of clusters is reduced by 1, and entries in the similarity matrix are updated to reflect the degree of similarity between the new cluster, $q$, and all remaining clusters. Similarity measures involving cluster $p$ are deleted from the similarity matrix. The search of the similarity matrix resumes and the next most similar pair of clusters is identified, merged, and the similarity matrix is updated again. The procedure continues until the required
number of clusters is obtained, or until all entities have been merged into a single group.

Various clustering algorithms differ in terms of the way in which similarity is defined and the way in which the similarity matrix is updated.

The groups formed during this process are nested, in the sense that once two entities are merged, they are joined together permanently and cannot be reallocated to different clusters at a later stage in the procedure. As a result, the number of possible groupings, which need to be examined at each stage (except the first), is considerably less than what would be required under a system of complete enumeration. On the other hand, cluster analysis does not guarantee an optimal grouping of entities past the first stage.

2.6.3 Aggregation using Cluster Analysis

Mukherjee [(1970), pp. 661-670] used a clustering algorithm to group the 36 industry input-output model of the Indian economy into nine sectors. The criterion used to combine industries is similarity of input coefficients, as measured by the distance function:

$$d_{jk} = \sum_{i=1}^{36} |a_{ij} - a_{ik}|$$  \hspace{1cm} (69)

where $d_{jk}$ is a measure of the distance between industries $j$ and $k$ in 36 dimensional space spanned by the input coefficients.
The smaller the value of $d_{jk}$, the more similar are the input structures of industries $j$ and $k$. With 36 industries there are 630 distance measures, which Mukherjee ranked in ascending order, and the ranks were arranged in the form of a similarity matrix, that is,

$$s_{jk} = \text{rank}(d_{jk})$$

Beginning with 36 sectors, each consisting of one industry, the two sectors, (say) $p$ and $q$, with the smallest rank are merged to form a new sector, which assumes the label $q$. Measures of similarity are updated, using the complete linkage method [see Anderberg (1973), pp. 138-139]:

$$s_{qr} = \max(s_{pr}, s_{qr})$$ \hspace{1cm} \ldots(71)$$

$$s_{pr} = 0$$ \hspace{1cm} \ldots(72)$$

where $r$ is any sector other than $p$ or $q$.

With 35 sectors, one consisting of two industries, and the other 34 of one industry each, the next two most similar sectors are merged. Similarity measures between the new sector and all others are updated, using equations (71) and (72), and the process continues until the original 36 sectors have been grouped into the nine required.

Mukherjee found that the procedure described above grouped certain industries which, on a priori grounds, appeared unrelated. He therefore imposed constraints to ensure that aggregation took place only within the three major groups of industries in the Indian interindustry table, namely, "agriculture and allied activities", "mining and manufacturing" and "all other activities". This is achieved simply by setting
the initial similarity measures between industries in different groups, at artificially high levels. The nine sectors so produced were more acceptable on a priori grounds.

Although there is no guarantee that Mukherjee's methodology will produce an aggregation pattern for the nine sectors which minimizes aggregation bias, the criterion used to group industries is theoretically sound. Furthermore, it is superior to the method of Neudecker in that it does not require the evaluation of numerous aggregation patterns, chosen in a subjective manner by the analyst.

Kossov's method [(1972), pp. 241-248] is similar to that of Mukherjee in that it systematically combines industries into sectors, using similarity of input coefficients as the criterion for aggregation. However, unlike Mukherjee's procedure, similarity of input structures is measured by the product moment correlation coefficient and the final number of sectors in the aggregated model is determined by the algorithm, itself.

The procedure begins by normalising input coefficients such that:

\[ \sum_{i=1}^{n} \tilde{a}_{ij} = 0 \quad j=1,2,\ldots,n \quad \ldots(73) \]

\[ \sum_{i=1}^{n} \tilde{a}_{ij}^2 = 1 \quad j=1,2,\ldots,n \quad \ldots(74) \]

which eliminates the influence of value added. The coefficient of correlation between the input coefficients of industries
and $k$ is calculated as:

$$r_{jk} = \frac{\sum_{i=1}^{n} \tilde{a}_{ij} \tilde{a}_{ik}}{\left( \sum_{i=1}^{n} \tilde{a}_{ij}^2 \right)^{1/2} \left( \sum_{i=1}^{n} \tilde{a}_{ik}^2 \right)^{1/2}}$$

...(75)

where the symbol "$\sim$" denotes normalised input coefficients.

The larger the value of $r_{jk}$, the more acceptable is the aggregation of the two industries.

Extending the correlation measure from two industries to a sector, $\pi$, containing $N$ industries, the average correlation between their input structures is:

$$\alpha = \frac{2S}{N(N-1)}$$

...(76)

where $S = \sum_{k>j} r_{kj}$ for all $k,j$ in sector $\pi$

...(77)

The average correlation between the input structures of all industries in sector $\pi$ and those of all other industries is:

$$\beta = \frac{\sum_{i,j=1}^{n} r_{ij} - 2S}{N(N-N)}$$

for all $i$ in sector $\pi$, $i \neq j$

...(78)

and the aggregation of these $N$ industries into a single sector is considered acceptable if $\gamma = \alpha/\beta > 1$.

The algorithm used by Kossov can now be outlined as follows. Beginning with $n$ sectors, each consisting of one industry, the correlation coefficient between the input structures of all pairs of industries is calculated, using equation (75), and $N$ is set equal to two for each of the $\binom{n}{2}$ potential two-industry sectors. The "within sector" correlation is calculated, using equations (76) and (77), the "between sector" correlation is calculated, using equation (78), and their ratio, $\gamma$, is determined for each potential two-industry
sector. The values of $\gamma$ form the elements of the similarity matrix.

The two sectors with the largest value of $\gamma$ are merged, provided $\gamma > 1$, $N$ is increased to $N + 1$, for the newly formed sector and entries in the similarity matrix are updated, using equations (76), (77) and (78). The procedure continues until no two sectors have a value of $\gamma$ which is greater than unity.

Kossov's algorithm is an attractive one but it has one major disadvantage; the use of the correlation coefficient as a measure of similarity, even though input coefficients have been normalised, leads to the grouping of industries whose original input coefficients are either equal or proportional, in approximate terms. For example, if one industry has input coefficients \{0.10, 0.20, 0.05, 0.05\} and another industry has input coefficients \{0.20, 0.40, 0.10, 0.10\} then both industries display the same normalised input structure of \{0.000, 0.816, -0.408, -0.408\}. Theoretical results, on the other hand, suggest that industries should be aggregated if their input coefficients are equal, but not if their input coefficients are proportional. Hence, the application of a distance function to the original input coefficients, as used in the method suggested by Mukherjee, is a more suitable procedure for aggregating industries of an input-output model. Alternatively, the correlation coefficient may be used if input coefficients, including value added, are normalised.

Blin and Cohen (1977) use two clustering algorithms to
aggregate the 1967 input-output model of the United States economy from 83 industries into one sector, in 82 successive stages. The criterion used to aggregate industries is technological similarity, as measured by:

(a) either the correlation coefficient between normalised input coefficients, as given in equation (75), that is,

\[ s_{jk} = \frac{1}{n} \sum_{i=1}^{n} \tilde{a}_{ij} \tilde{a}_{ik} \]  

(b) or the squared Euclidean distance between input coefficients, including value added, that is,

\[ s_{jk} = \sum_{i=1}^{n+1} (a_{ij} - a_{ik})^2 \]  

where \( a_{n+1,j} \) and \( a_{n+1,k} \) are per unit value added for industries \( j \) and \( k \), respectively.

The first algorithm is a centroid method\(^{19}\) which was developed by Sokal and Michener (1958). The method merges, at each stage, sectors with the most similar mean vectors (of input coefficients) or centroids. At the beginning of the procedure, each sector consists of only one industry so its centroid is its set of input coefficients and the initial similarity matrix is constructed using either equation (75) or equation (79).

Following the merger of the two most similar sectors, the centroid of the new sector is calculated as a weighted average of the centroids of the merged sectors. The correlation coefficients, or squared Euclidean distances, between the

\(^{19}\) Centroid methods are discussed by Anderberg [(1973), pp. 140-142].
centroid of the new sector and all remaining sectors are calculated and the similarity matrix is updated. The next two most similar sectors are then merged and the process continues until a predetermined level of aggregation is reached, or until all sectors are merged into one.

The second algorithm was developed by Ward (1963) and was implemented by Wishart (1969). It employs only squared Euclidean distances as measures of similarity. The objective is to find, at each stage, the two sectors whose merger results in the minimum increase in the total within sector sum of squared deviations of input coefficients from their means. That is, if we define:

\[ a_{ij} \] as the input coefficient from industry i to the \( l^{th} \) of \( n_j \) industries in sector j, and

\[ \bar{a}_{ij} = \frac{\sum_{l=1}^{n_j} a_{ilj}}{n_j} \] as the mean input coefficient from industry i to sector j,

then

\[ E_j = \sum_{i=1}^{n+1} \sum_{l=1}^{n_j} (a_{ilj} - \bar{a}_{ij})^2 \] is the within sector sum of squared deviations of input coefficients to sector j, from their means.

Similarly,

\[ E_k = \sum_{i=1}^{n+1} \sum_{l=1}^{n_k} (a_{ilk} - \bar{a}_{ik})^2 \] is the within sector sum of squared deviations of input coefficients to sector k, from their means.
The total within sector sum of squared deviations of input coefficients from their means is given by:

\[ E = \sum_{j=1}^{m} E_j \]

where \( m \) is the number of sectors.

It can be shown [see Anderberg (1973), p. 143] that the increase in the total within sector sum of squared deviations of input coefficients from their means, resulting from the potential merger of sectors \( j \) and \( k \) is given by:

\[ \Delta E_{jk} = \frac{n_j n_k}{n_j + n_k} \sum_{i=1}^{n+1} (\bar{a}_{ij} - \bar{a}_{ik})^2 \]

and so is proportional to the squared Euclidean distance between the centroids of the two sectors.

Again, at the beginning of the procedure, each sector contains one industry so the initial similarity measures are equal to half the squared Euclidean distance between the input coefficients of pairs of industries.

Following the merger of the two most similar sectors, (say) \( p \) and \( q \), to form a new sector, labelled \( q \), the similarity matrix is updated. Updates may be calculated using equation (80) but a more convenient method is available [see Anderberg (1973), p. 144]:

\[ s_{qr} = \Delta E_{qr} + \frac{1}{n_p + n_q + n_r} \left\{ (n_p + n_r) \Delta E_{pr} + (n_q + n_r) \Delta E_{qr} - n_r \Delta E_{pq} \right\} \quad \ldots (81) \]

\[ s_{pr} < 0 \quad \ldots (82) \]

where \( r \) is any sector other than \( p \) or \( q \).
The two next most similar sectors are merged and the process continues until the desired level of aggregation is reached, or until all sectors have been merged into one.

Blin and Cohen found that the pattern of aggregation produced by the two algorithms conformed with what was expected. They also found that discrimination between sectors was not as sharp when value added was excluded from the measure of technological similarity. However, their use of the correlation coefficient as a measure of similarity is subject to the criticism mentioned earlier, namely, that it will result in the grouping of industries whose input coefficients are proportional.

2.7 Summary

In this chapter the conditions under which aggregation bias, and first-order aggregation bias, vanish, have been discussed. The effect of aggregation on output and income multipliers has also been examined. In addition, existing methodologies for grouping industries into sectors have been reviewed. The methods of Fisher and Neudecker are less than satisfactory in that they incorporate assumptions about future final demands which are unlikely to be satisfied. Neudecker's method is not practicable in that it involves experimentation with various aggregation schemes, from which the best is chosen. Mukherjee, Kossov and Blin and Cohen all use cluster analysis to aggregate industries into sectors, on the basis
of similarity of input structures. In each of these methods, the grouping of industries is determined by the algorithm itself, and no knowledge of, or assumptions about, final demands is required. The use of a distance function, such as the squared Euclidean distance, as a measure of similarity of input structures is more theoretically sound than the use of the correlation coefficient, as the latter may lead to aggregation of industries whose input coefficients are proportional, rather than equal.

In the next chapter, these results will be used to develop an improved method of aggregation.
In this appendix the following are derived:

(a) the matrix $\bar{A}$ which minimises the expected sum of squares of aggregation errors, and

(b) the minimum expected sum of squares of aggregation errors.  

We begin with equation (56) with $\phi$ omitted, that is

$$
\lambda = \text{tr}\{((I-A)^{-1})^T - T(I-A)^{-1}\}F_O\{((I-A)^{-1})^T - T(I-A)^{-1}\}' \quad \ldots (56)
$$

Using $\text{tr}(V+W) = \text{tr}(V) + \text{tr}(W)$

$$
\lambda = \text{tr}\{((I-A)^{-1})^T F_O\{((I-A)^{-1})'T\}' - \text{tr}\{((I-A)^{-1})^T F_O\{((I-A)^{-1})'T\}'
- \text{tr}(T(I-A)^{-1} F_O\{((I-A)^{-1})'T\} + \text{tr}(T(I-A)^{-1} F_O\{((I-A)^{-1})'T\}' \ldots (83)
$$

Since $\text{tr}(VW) = \text{tr}(WV)$ and $F_O$ is symmetric the third term in the above expression can be written as

$$
- \text{tr}\{((I-A)^{-1})^T F_O\{((I-A)^{-1})'T\}' \ldots (84)
$$

Substituting equation (84) into equation (83) gives

$$
\lambda = \text{tr}\{((I-A)^{-1})^T F_O\{((I-A)^{-1})'T\}' - 2\text{tr}\{((I-A)^{-1})^T F_O\{((I-A)^{-1})'T\}'
+ \text{tr}(T(I-A)^{-1} F_O\{((I-A)^{-1})'T\}' \ldots (85)
$$

Since $d(\text{tr}(V)) = \text{tr}(d(V))$ equation (85) becomes

$$
d\lambda = \text{tr}\{d\{((I-A)^{-1})^T F_O\{((I-A)^{-1})'T\}' - 2d\{((I-A)^{-1})^T F_O\{((I-A)^{-1})'T\}'
+ d\{T(I-A)^{-1} F_O\{((I-A)^{-1})'T\}' \ldots (86)
$$

Since $d(VW) = (dV)W + V(dW)$ equation (86) becomes

$$
d\lambda = \text{tr}\{d\{((I-A)^{-1})^T F_O\{((I-A)^{-1})'T\}' + (I-A)^{-1} F_O\{d\{((I-A)^{-1})'T\}'
- 2d\{((I-A)^{-1}) F_O\{((I-A)^{-1})'T\}' \ldots (87)
$$

20. The theorems relating to traces and matrix calculus can be found in Malinvaud [(1970), pp. 196-200].
Since $\text{tr}(VW) = \text{tr}(W'V')$ and $F^D$ and $(TF_D')$ are symmetric, equation (87) becomes

$$d\lambda = \text{tr}\{(I-A)^{-1}(TF_D')\{d(I-A)^{-1}\} + (I-A)^{-1}(TF_D')\{d(I-A)^{-1}\}$$

$$- 2T(I-A)^{-1}F^D'T\{d(I-A)^{-1}\}$$

$$= 2\text{tr}\{(I-A)^{-1}(TF_D') - T(I-A)^{-1}F^D'T\}\{d(I-A)^{-1}\} \quad \ldots (88)$$

From equation (88) we see that $\lambda$ is minimised when

$$(I-A)^{-1}(TF_D') = T(I-A)^{-1}F^D'T.$$ 

Therefore,

$$(I-A)^{-1} = T(I-A)^{-1}F^D'T (TF_D')^{-1} \quad \ldots (89)$$

and

$$(I-A) = (TF_D')\{T(I-A)^{-1}F^D'T\}^{-1}$$

$$= T[I - (I-A)](I-A)^{-1}F^D'T (T(I-A)^{-1}F^D'T)^{-1}$$

$$= TA(I-A)^{-1}F^D'T (T(I-A)^{-1}F^D'T)^{-1} \quad \ldots (90)$$

Equation (90) gives the matrix $\tilde{A}$ which minimises the expected sum of squares of aggregation errors.

Substituting equation (89) into equation (56) gives

$$\lambda = \text{tr}\{(T(I-A)^{-1}F^D'T (TF_D')^{-1}T - T(I-A)^{-1}F^D'T (TF_D')^{-1}T$$

$$- T(I-A)^{-1}F^D'T (TF_D')^{-1}T - T(I-A)^{-1}F^D'T (TF_D')^{-1}T$$

$$= \text{tr}\{(T(I-A)^{-1}F^D'T (TF_D')^{-1}T - I)\{F^D'T (TF_D')^{-1}T - I\}\{T(I-A)^{-1}\}$$

$$= \text{tr}\{(T(I-A)^{-1}F^D'T (TF_D')^{-1}T - T'(TF_D')^{-1}T - T'(TF_D')^{-1}T$$

$$- T'(TF_D')^{-1}T + I\}\{T(I-A)^{-1}\}$$

$$= \text{tr}\{(T(I-A)^{-1}F^D'I - T'(TF_D')^{-1}T)\{T(I-A)^{-1}\} \quad \ldots (91)$$

Equation (91) gives the minimum expected sum of squares of aggregation errors.
CHAPTER 3

A METHODOLOGY FOR AGGREGATION

This chapter begins, in Section 3.1, with a specification of the capabilities, which are required of a methodology for aggregating industries of an input-output model into sectors. A clustering algorithm, which meets these requirements, is developed in Section 3.2. Finally, the implementation of the clustering algorithm, in a form which lends itself to computerization, is presented in Section 3.3.

3.1 The Capabilities Required of a Methodology for Aggregation

3.1.1 The Aggregation Criterion

A methodology is required, which can group industries of an input-output model into sectors, so that aggregation bias in forecasts of gross output, produced by the aggregated model, is as small as possible. In Chapter 2 the theoretical conditions, under which aggregation bias, and first-order aggregation bias, vanish, were reviewed. Broadly speaking, these conditions can be classified into two categories: those which relate to the technical coefficients, and those which relate to final demands. Since, in general, final demands are unknown at the time when the model needs to be condensed, the criterion for aggregation should be based upon theoretical conditions concerning the technical coefficients. On the other
hand, the aggregation procedure should be flexible enough to allow information concerning future final demands to be taken into account, should it be available.

There are two conditions, pertaining to the technical coefficients, under which aggregation bias vanishes:

(a) Industries may be combined to form a single sector if, in order to produce one unit of output, they each require equal, aggregate amounts of input from industries in the same sector. In other words, aggregation bias vanishes if industries, which are grouped into the same sector, have equal, partially aggregated input coefficients.

(b) Industries, with identical input coefficients, may be grouped into the same sector without introducing aggregation bias into forecasts of gross output.

Equality of input coefficients was used as a criterion for aggregation in the procedures reviewed in Section 2.6.3. However, being a sufficient, but not a necessary, condition for consistent aggregation, equality of input coefficients is a more restrictive condition than equality of partially aggregated input coefficients, which is both necessary and sufficient. Therefore, equality of partially aggregated input coefficients is considered to be the better criterion for aggregation.

3.1.2 Aggregation Constraints

The methodology should allow various constraints to be imposed upon the pattern of aggregation, either to take account
of information concerning future final demands, or to allow for particular uses to which the intersectoral model is to be put. More specifically, it should be possible to ensure that:

(a) Certain industries are aggregated into the same sector. For example, the analyst may wish to take advantage of the fact that two industries, one of which consumes the entire output of the other, can be aggregated, without producing biased forecasts of gross outputs.

(b) Certain industries are not aggregated into the same sector, or, equivalently, aggregation takes place only within specific groups of industries. For example, in aggregating industries of an international input-output model, normally one would not wish to aggregate industries of different countries into the same sector. Also, if the objective is to determine the aggregate change in gross output of a set of industries, in response to changes in final demands, then these industries should not be aggregated with others, which are not included in the set.

(c) Certain industries are not aggregated with any others in forming the intersectoral model. For example, if interest is focussed upon the effects of changes in final demands upon the outputs of a few, key industries, then the key industries should remain isolated in the aggregated model. This is, in fact, a special case of requirement (b) above, for each key industry may not be aggregated with any other industry.
3.1.3 Other Requirements

(a) The aggregation of industries into sectors should be performed by the algorithm itself; it should not be necessary for the analyst to experiment with various aggregation schemes, from which the best is chosen.

(b) The methodology should be capable of aggregating large input-output models within reasonable time and cost constraints.

3.2 A Clustering Algorithm for Aggregating Industries of an Input-Output Model

In accordance with the results of theoretical research, industries are to be aggregated into the same sector if they have relatively similar partially aggregated input coefficients. Similarity will be measured by the within sector, sum of squared deviations of all partially aggregated input coefficients from their means.

Let:

\( h \) be the number of sectors into which industries have been aggregated, at a given stage of the clustering procedure.

\( a_{ijkl} \) be the input from the \( i^{th} \) industry of sector \( j \), which is required to produce one unit of output by the \( k^{th} \) industry of sector \( l \).

\( n_j \) be the number of industries in sector \( j \).
\( n_j \) be the number of industries in sector \( j \).

\[ n = \sum_{j=1}^{h} n_j = \sum_{k=1}^{\infty} n_k \] be the total number of industries in the model.

\[ a_{jkl} = \sum_{i=1}^{n_j} a_{ijkl} \] be the total input from all industries of sector \( j \), which is required to produce one unit of output by the \( k \)th industry of sector \( l \). That is, \( a_{jkl} \) are the partially aggregated input coefficients.

\[ \overline{a}_{jkl} = \sum_{k=1}^{n_k} a_{jkl} / n_k \] be the mean of the partially aggregated input coefficients from sector \( j \) to sector \( l \).

Also let:

\[ E_h^l \] be the sum of squared deviations of the partially aggregated input coefficients of sector \( l \) about their means (or the "error sum of squares" for sector \( l \)) in the \( h \) sector model. Therefore,

\[ E_h^l = \sum_{j=1}^{h} \sum_{k=1}^{n_k} (a_{jkl} - \overline{a}_{jkl})^2 \]

\[ = \sum_{j=1}^{h} \sum_{k=1}^{n_k} a_{jkl}^2 - n_k \sum_{j=1}^{h} \overline{a}_{jkl}^2. \]

\( E_{pq}^h = \sum_{l=1}^{h} E_l^h = \sum_{l=1}^{h} \sum_{j=1}^{n_l} \sum_{k=1}^{a_{jkl}} a_{jkl}^2 - \sum_{j=1}^{n_l} \sum_{k=1}^{a_{jkl}} a_{jkl}^2 \]

\[ \sum_{j=1}^{n_l} \sum_{k=1}^{a_{jkl}} a_{jkl}^2 - \sum_{j=1}^{n_l} \sum_{k=1}^{a_{jkl}} a_{jkl}^2 \]

be the total sum of squared deviations of the partially aggregated input coefficients from their means for all \( h \) sectors, which would result from the merger of sectors \( p \) and \( q \), in the \( h+1 \) sector.
model, to form an \( h \) sector model. \( E_{pq}^h \) is referred to as the "total error sum of squares", and the smaller its value, the more similar are the partially aggregated input coefficients within each of the \( h \) sectors.

\[
E_{rs}^h = \min\{E_{pq}^h\} \quad (h=n-1,n-2,\ldots,2) \quad \ldots(93)
\]

be the total error sum of squares, resulting from the merger of sectors \( r \) and \( s \) in the \( h+1 \) sector model, to form the actual \( h \) sector model.

Initially, there are \( n \) sectors, each consisting of one industry, so the input coefficients, \( a_{ijkl} \) \((i=1; j=1,2,\ldots,n; k=1; \ell=1,2,\ldots,n)\), the partially aggregated input coefficients, \( a_{jk\ell} \) \((j=1,2,\ldots,n; k=1; \ell=1,2,\ldots,n)\) and the means of the partially aggregated input coefficients, \( \bar{a}_{j\ell} \) \((j=1,2,\ldots,n; \ell=1,2,\ldots,n)\) are all equal. Consequently, \( E_{pq}^n \) \((\ell=1,2,\ldots,n)\) are all equal to zero and the total error sums of squares, \( E_{pq}^n \), are also equal to zero.

The aggregation of sectors \( p \) and \( q \) \((p>q)\) to form a new sector, labelled \( t(=q) \), would lead to an increased total error sum of squares:

\[
E_{pq}^{n-1} = \frac{1}{(n-1)(n-1)} \sum_{\ell=1}^{n-1} \sum_{j=1}^{n-1} a_{jk\ell}^2 \sum_{k=1}^{n-1} \sum_{j=1}^{n-1} a_{j\ell}^2 - \frac{1}{(n-1)(n-1)} \sum_{\ell=1}^{n-1} \sum_{j=1}^{n-1} a_{j\ell}^2
\]

\[
= \sum_{j=1}^{n-1} \sum_{k=1}^{n-1} a_{jkt}^2 - n_t \sum_{j=1}^{n-1} a_{jt}^2 \quad \ldots(94)
\]

where \( n_t = n_p + n_q \)

Equation (94) is used to construct the initial similarity
matrix. For example, consider the matrix of technical coefficients, given in Table 3.2.1 below.

<table>
<thead>
<tr>
<th>Sector 1 (I1)</th>
<th>Sector 2 (I2)</th>
<th>Sector 3 (I3)</th>
<th>Sector 4 (I4)</th>
<th>Sector 5 (I5)</th>
<th>Sector 6 (I6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>0.05</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.05</td>
</tr>
<tr>
<td>0.40</td>
<td>0.10</td>
<td>0.30</td>
<td>0.20</td>
<td>0.25</td>
<td>0.20</td>
</tr>
<tr>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.00</td>
<td>0.05</td>
<td>0.30</td>
</tr>
<tr>
<td>0.10</td>
<td>0.05</td>
<td>0.05</td>
<td>0.10</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>0.00</td>
<td>0.10</td>
<td>0.05</td>
<td>0.10</td>
<td>0.20</td>
<td>0.25</td>
</tr>
<tr>
<td>0.10</td>
<td>0.20</td>
<td>0.10</td>
<td>0.20</td>
<td>0.20</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Note that in Table 3.2.1, and the tables which follow, the symbol "In" is used to denote the n-th industry in the original model. For example, I3 is the third industry in the original model.

The aggregation of sectors 1 and 2 to form a new sector, labelled 1, would result in the partially aggregated input coefficients, and the means of partially aggregated input coefficients, displayed in brackets in Table 3.2.2 below. Note that in those sectors, where the partially aggregated input coefficients and the means of partially aggregated input coefficients are equal, the latter are not presented.
Table 3.2.2
Partially Aggregated Input Coefficients Resulting from the Potential Merger of Sectors 1 and 2

<table>
<thead>
<tr>
<th>Sector 1 (I1)</th>
<th>Sector 2 (I2)</th>
<th>Sector 3 (I3)</th>
<th>Sector 4 (I4)</th>
<th>Sector 5 (I5)</th>
<th>Sector 6 (I6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Mean)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.20</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.20</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>0.40</td>
<td>0.10</td>
<td>(0.50)</td>
<td>(0.15)</td>
<td>(0.325)</td>
<td></td>
</tr>
<tr>
<td>0.30</td>
<td>0.20</td>
<td>(0.40)</td>
<td>(0.30)</td>
<td>(0.35)</td>
<td>(0.25)</td>
</tr>
<tr>
<td>0.20 (0.20)</td>
<td>(0.20)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.20 (0.05)</td>
<td>(0.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.10 (0.10)</td>
<td>(0.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.20 (0.10)</td>
<td>(0.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.10 (0.20)</td>
<td>(0.10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.20 (0.10)</td>
<td>(0.20)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The total error sum of squares, resulting from the potential merger of sectors 1 and 2, is computed as:

$$E^5_{21} = (0.50^2 + 0.15^2 + 0.20^2 + 0.20^2 + 0.10^2 + 0.05^2 + 0.00^2 + 0.10^2 + 0.10^2 + 0.20^2) - 2(0.325^2 + 0.200^2 + 0.075^2 + 0.050^2 + 0.150^2) = 0.0725$$

The total error sum of squares, resulting from the potential merger of each other pair of sectors, is calculated in the same way and these measures are arranged to form a similarity matrix, as in Table 3.2.3.

Table 3.2.3
Similarity Matrix

<table>
<thead>
<tr>
<th>E^5_{21} = 0.0725</th>
</tr>
</thead>
<tbody>
<tr>
<td>E^5_{31} = 0.0075</td>
</tr>
<tr>
<td>E^5_{41} = 0.0500</td>
</tr>
<tr>
<td>E^5_{51} = 0.0525</td>
</tr>
<tr>
<td>E^5_{61} = 0.0625</td>
</tr>
<tr>
<td>E^5_{32} = 0.0275</td>
</tr>
<tr>
<td>E^5_{42} = 0.0325</td>
</tr>
<tr>
<td>E^5_{52} = 0.0450</td>
</tr>
<tr>
<td>E^5_{62} = 0.0175</td>
</tr>
<tr>
<td>E^5_{43} = 0.0225</td>
</tr>
<tr>
<td>E^5_{53} = 0.0075</td>
</tr>
<tr>
<td>E^5_{63} = 0.0325</td>
</tr>
<tr>
<td>E^5_{44} = 0.0025*</td>
</tr>
<tr>
<td>E^5_{54} = 0.0775</td>
</tr>
<tr>
<td>E^5_{64} = 0.0350</td>
</tr>
</tbody>
</table>
The merger of sectors 4 and 5, in Table 3.2.1, produces the smallest total error sum of squares, so these two sectors are aggregated to form a new sector, labelled 4. The resulting five sector model is depicted in Table 3.2.4 and the total error sum of squares for the five sector model is $E_{54}=0.0025$.

Table 3.2.4
Technical Coefficients of the Five Sector Model

<table>
<thead>
<tr>
<th>Sector</th>
<th>Sector 1</th>
<th>Sector 2</th>
<th>Sector 3</th>
<th>Sector 4</th>
<th>Sector 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(I1)</td>
<td>(I2)</td>
<td>(I3)</td>
<td>(I4)</td>
<td>(I5)</td>
</tr>
<tr>
<td>Sector 1 (I1)</td>
<td>0.10</td>
<td>0.05</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>Sector 2 (I2)</td>
<td>0.40</td>
<td>0.10</td>
<td>0.30</td>
<td>0.20</td>
<td>0.25</td>
</tr>
<tr>
<td>Sector 3 (I3)</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.00</td>
<td>0.05</td>
</tr>
<tr>
<td>Sector 4 (I4)</td>
<td>0.10</td>
<td>0.05</td>
<td>0.05</td>
<td>0.10</td>
<td>0.00</td>
</tr>
<tr>
<td>Sector 5 (I5)</td>
<td>0.00</td>
<td>0.10</td>
<td>0.05</td>
<td>0.10</td>
<td>0.20</td>
</tr>
<tr>
<td>Sector 6 (I6)</td>
<td>0.10</td>
<td>0.20</td>
<td>0.10</td>
<td>0.20</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Having aggregated sectors 4 and 5 to form a new sector 4, the elements of the similarity matrix, given in Table 3.2.3, must be updated. This is performed by considering the potential aggregation of pairs of sectors, which appear in the five sector model, given in Table 3.2.4, and calculating the total error sum of squares for each, using equation (92).
The aggregation of sectors 1 and 2 into a new sector, labelled 1, would result in partially aggregated input coefficients, and means of partially aggregated input coefficients, as displayed in brackets in Table 3.2.5 below.

<table>
<thead>
<tr>
<th>Sector</th>
<th>Sector 1</th>
<th>Sector 3</th>
<th>Sector 4</th>
<th>Sector 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(I1)</td>
<td>(I2)</td>
<td>Mean</td>
<td>(I7)</td>
</tr>
<tr>
<td>Sector 1</td>
<td>0.10</td>
<td>0.05</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>0.40</td>
<td>0.10</td>
<td>0.30</td>
<td>0.20</td>
</tr>
<tr>
<td>Sector 3</td>
<td>0.20</td>
<td>0.20</td>
<td>Mean</td>
<td>(I3)</td>
</tr>
<tr>
<td>Sector 4</td>
<td>0.10</td>
<td>0.05</td>
<td>0.05</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.10</td>
<td>0.05</td>
<td>0.10</td>
</tr>
<tr>
<td>Sector 6</td>
<td>(I6)</td>
<td>(I7)</td>
<td>Mean</td>
<td></td>
</tr>
</tbody>
</table>

The total error sum of squares, resulting from the potential aggregation of sectors 1 and 2, is computed as:

\[
E_{21}^4 = (0.50^2 + 0.15^2 + 0.20^2 + 0.20^2 + 0.10^2 + 0.15^2 + 0.10^2 + 0.20^2) - 2(0.325^2 + 0.200^2 + 0.125^2 + 0.150^2) + (0.30^2 + 0.35^2 + 0.00^2 + 0.05^2 + 0.20^2 + 0.20^2 + 0.20^2) - 2(0.325^2 + 0.025^2 + 0.200^2 + 0.200^2) = 0.0700
\]

\[E_{31}^4, E_{32}^4, E_{61}^4, E_{62}^4 \text{ and } E_{63}^4 \] are calculated in a similar way.
The aggregation of sector 1 with sector 4, to form a new sector, labelled 1, would result in partially aggregated input coefficients, and means of partially aggregated input coefficients, as displayed in brackets in Table 3.2.6 below.

Table 3.2.6
Partially Aggregated Input Coefficients Resulting from the Potential Merger of Sectors 1 and 4

<table>
<thead>
<tr>
<th>Sector</th>
<th>Sector 1</th>
<th>Sector 2</th>
<th>Sector 3</th>
<th>Sector 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(I1)</td>
<td>(I2)</td>
<td>(I3)</td>
<td>(I4)</td>
</tr>
<tr>
<td>Sector 1 (I4)</td>
<td>0.10</td>
<td>0.10</td>
<td>0.00</td>
<td>(II)</td>
</tr>
<tr>
<td>Sector 1 (I5)</td>
<td>0.00</td>
<td>0.10</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.20)</td>
<td>(0.30)</td>
<td>(0.30)</td>
<td>(0.80/3)</td>
</tr>
<tr>
<td>Sector 2 (I2)</td>
<td>(0.40)</td>
<td>(0.20)</td>
<td>(0.25)</td>
<td>(0.85/3)</td>
</tr>
<tr>
<td>Sector 3 (I3)</td>
<td>(0.20)</td>
<td>(0.00)</td>
<td>(0.05)</td>
<td>(0.25/3)</td>
</tr>
<tr>
<td>Sector 6 (I6)</td>
<td>(0.10)</td>
<td>(0.20)</td>
<td>(0.20)</td>
<td>(0.50/3)</td>
</tr>
</tbody>
</table>

The total error sum of squares, resulting from the potential aggregation of sectors 1 and 4, is computed as:

\[
E_{41}^4 = \frac{1}{9}(0.20^2 + 0.30^2 + 0.30^2 + 0.40^2 + 0.20^2 + 0.25^2 + 0.20^2 + 0.00^2 + 0.05^2 + 0.10^2 + 0.20^2 + 0.20^2) - 3(0.80^2 + 0.85^2 + 0.25^2 + 0.50^2)
\]

\[
= 0.0567
\]

\[
E_{42}^4, E_{43}^4 \text{ and } E_{64}^4 \text{ are calculated in a similar manner.}
\]
Finally, all elements involving sector 5 are deleted from the similarity matrix. Its final form is given in Table 3.2.7.

**Table 3.2.7**  
Revised Similarity Matrix

<table>
<thead>
<tr>
<th></th>
<th>$E_{21}^4$ = 0.0700</th>
<th>$E_{31}^4$ = 0.0075</th>
<th>$E_{32}^4$ = 0.0325</th>
<th>$E_{41}^4$ = 0.0567</th>
<th>$E_{42}^4$ = 0.0450</th>
<th>$E_{43}^4$ = 0.0167</th>
<th>$E_{61}^4$ = 0.0400</th>
<th>$E_{62}^4$ = 0.0125</th>
<th>$E_{63}^4$ = 0.0250</th>
<th>$E_{64}^4$ = 0.0567</th>
</tr>
</thead>
</table>

The merger of sectors 1 and 3 produces the smallest total error sum of squares, so these two sectors are aggregated to form a new sector, labelled 1. The resulting four sector model is given in Table 3.2.8 and the total error sum of squares for the four sector model is $E_{31}^4 = 0.0075$.

**Table 3.2.8**  
Technical Coefficients of the Four Sector Model

<table>
<thead>
<tr>
<th></th>
<th>Sector 1</th>
<th>Sector 2</th>
<th>Sector 4</th>
<th>Sector 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(I1)</td>
<td>(I3)</td>
<td>(I2)</td>
<td>(I4)</td>
</tr>
<tr>
<td>-----</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>Sector 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(I1)</td>
<td>0.10</td>
<td>0.10</td>
<td>0.05</td>
<td>0.10</td>
</tr>
<tr>
<td>(I3)</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.00</td>
</tr>
<tr>
<td>Sector 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(I2)</td>
<td>0.40</td>
<td>0.30</td>
<td>0.10</td>
<td>0.20</td>
</tr>
<tr>
<td>(I4)</td>
<td>0.10</td>
<td>0.05</td>
<td>0.05</td>
<td>0.10</td>
</tr>
<tr>
<td>Sector 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(I5)</td>
<td>0.00</td>
<td>0.05</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>Sector 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(I6)</td>
<td>0.10</td>
<td>0.10</td>
<td>0.20</td>
<td>0.20</td>
</tr>
</tbody>
</table>
The similarity matrix is updated again by considering the potential aggregation of pairs of sectors, which appear in the four sector model given in Table 3.2.8, and calculating the total error sum of squares for each, using equation (92). The result is given in Table 3.2.9.

**Table 3.2.9**

Revised Similarity Matrix

<table>
<thead>
<tr>
<th></th>
<th>$E_{21}^3 = 0.0783$</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$E_{31}^3 = 0.0388$</td>
<td>$E_{41}^3 = 0.0383$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$E_{61}^3 = 0.0392$</td>
<td>$E_{62}^3 = 0.0175^2$</td>
<td></td>
<td>$E_{42}^3 = 0.0433$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The merger of sectors 2 and 6 produces the smallest total error sum of squares, so these two sectors are aggregated to form a new sector, labelled 2. The resulting three sector model is given in Table 3.2.10 and the total error sum of squares for the three sector model is $E_{62}^3 = 0.0175$.

**Table 3.2.10**

Technical Coefficients of the Three Sector Model

<table>
<thead>
<tr>
<th></th>
<th>Sector 1</th>
<th>Sector 2</th>
<th>Sector 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(I1)</td>
<td>(I3)</td>
<td>(I2)</td>
</tr>
<tr>
<td>Sector 1</td>
<td>0.10</td>
<td>0.10</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>(I3)</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>0.30</td>
<td>0.00</td>
<td>0.05</td>
</tr>
<tr>
<td>(I2)</td>
<td>0.40</td>
<td>0.30</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>0.20</td>
<td>0.20</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sector 2</td>
<td>0.10</td>
<td>0.10</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>(I6)</td>
<td>0.10</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.10</td>
<td>0.00</td>
</tr>
<tr>
<td>(I4)</td>
<td>0.10</td>
<td>0.05</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.10</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sector 4</td>
<td>0.00</td>
<td>0.05</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.10</td>
<td>0.20</td>
</tr>
</tbody>
</table>
The similarity matrix is revised again by considering the potential merger of all possible pairs of sectors, which appear in the three sector model given in Table 3.2.10. The result is given in Table 3.2.11.

Table 3.2.11
Revised Similarity Matrix

\[
\begin{array}{ccc}
E_{21}^2 = 0.0525 & - & - \\
E_{41}^2 = 0.0338 & E_{42}^2 = 0.0638 & - \\
- & - & - \\
- & - & - \\
- & - & - \\
\end{array}
\]

Sectors 1 and 4 are aggregated and the resulting two sector model, given in Table 3.2.12, has a total error sum of squares, \( E_{41}^2 = 0.0338 \).

Table 3.2.12
Technical Coefficients of the Two Sector Model

<table>
<thead>
<tr>
<th></th>
<th>Sector 1</th>
<th></th>
<th>Sector 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(I1)</td>
<td>(I3)</td>
<td>(I4)</td>
</tr>
<tr>
<td>Sector 1</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>0.20</td>
<td>0.20</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.05</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.05</td>
<td>0.10</td>
</tr>
<tr>
<td>Sector 2</td>
<td>0.40</td>
<td>0.30</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.10</td>
<td>0.20</td>
</tr>
</tbody>
</table>
The final step of the clustering procedure groups the two remaining sectors into a single sector, with a total error sum of squares equal to 0.04.

As with all clustering algorithms, the procedure described above does not guarantee an optimal pattern of aggregation, except at the first stage. However, by significantly reducing the number of aggregation patterns which need to be considered, from the number which would be required with a complete enumeration, it is possible to produce reasonable solutions to large-scale input-output models. For example, in aggregating six industries into three sectors, 31 aggregation patterns need to be considered, compared with 90 which would need to be investigated to obtain the optimal solution using a complete enumeration. Generally, in aggregating \( n \) industries into \( m \) sectors, the number of patterns investigated by the clustering algorithm is

\[
\binom{n}{2} + \binom{n-1}{2} + \cdots + \binom{m+1}{2}
\]

while the number investigated by complete enumeration is

\[
\frac{1}{m!} \sum_{k=0}^{m} (-1)^{m-k} \binom{m}{k} k^n
\]

It is possible to ensure that certain industries are aggregated into the same sector by flagging their similarity measures for immediate aggregation. For example, to ensure that industries 1, 2 and 3 are aggregated into the same sector, the similarity matrix would be set up as shown in
Prior to the commencement of the clustering algorithm itself, industries 1 and 2 are aggregated, the similarity matrix is updated, then sector 1 (containing industries 1 and 2) is aggregated with industry 3. The similarity matrix is updated again and the clustering algorithm takes over.

It is also possible to ensure that certain industries are not aggregated into the same sector, by setting elements of the initial similarity matrix to artificially high values. For example, to ensure that industries 3, 4 and 5 are not aggregated with each other, the initial similarity matrix would be set up as in Table 3.2.14.
The updating process bypasses those elements, which have been set to artificially high levels.

Alternatively, it is possible to ensure that aggregation takes place only within specified groups of industries, by setting the appropriate elements of the initial similarity matrix to artificially large values. For example, aggregation is restricted to within industries 1, 2 and 3 and to within industries 4, 5 and 6, of a six industry table, if the initial similarity matrix is set up as in Table 3.2.15. Elements, which have been set to artificially large values, are bypassed during the updating of the similarity matrix.
Finally, it is possible to ensure that certain key industries are left unaggregated in the intersectoral model, by setting the initial similarity measures between each key industry and all others at artificially large values. Thus, industries 1 and 3 are guaranteed to remain isolated if the initial similarity matrix is constructed as in Table 3.2.16.

Table 3.2.16
Similarity Matrix with Industries 1 and 3 Remaining Isolated

\[
\begin{bmatrix}
E_{11} & E_{12} & \cdots & E_{1n} \\
E_{21} & E_{22} & \cdots & E_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
E_{n1} & E_{n2} & \cdots & E_{nn} \\
\end{bmatrix}
\]
Again the updating process bypasses those elements, which have been set to artificially large values.

3.3 Implementation of the Clustering Algorithm

The successful implementation of the clustering algorithm, described in Section 3.2, requires a more efficient method of setting up the initial similarity matrix, and of updating its elements at each stage of the clustering procedure.

3.3.1 Construction of the Initial Similarity Matrix

Initially, each sector consists of one industry, so the technical coefficients are given by \( a_{jl} \) (\( j=1,2,\ldots,n; \ l=1,2,\ldots,n \)). For simplicity, the first and third subscripts will be dropped temporarily, and the technical coefficients will be written as \( a_{jl} \).

The total error sum of squares, resulting from the potential aggregation of industries \( p \) and \( q \) (\( p>q \)), is given by equation (94), which can be simplified to:

\[
E_{pq}^{n-1} = (a_{pp} + a_{qq})^2 + (a_{pq} + a_{qp})^2 - \frac{2}{4} (a_{pp} + a_{pq} + a_{qp} + a_{qq})^2 + \sum_{j=1}^{n} a_{jp}^2 + \sum_{j=1}^{n} a_{jq}^2 - \frac{2}{4} \sum_{j=1}^{n} (a_{jp} + a_{jq})^2 
\]

\( j \neq p, q \) \( j \neq p, q \) \( j \neq p, q \)
\[
E_{pq}^{n-1} = \sum_{j=1}^{n} a_{jp}^2 + \sum_{j=1}^{n} a_{jq}^2 + 2(a_{pp} a_{qp} + a_{pq} a_{qq}) \\
- \frac{1}{2} \sum_{j=1}^{n} (a_{jp} + a_{jq})^2 - (a_{pp} + a_{pq})(a_{qp} + a_{qq}) \\
= \sum_{j=1}^{n} a_{jp}^2 + \sum_{j=1}^{n} a_{jq}^2 - \sum_{j=1}^{Jp/2} a_{jp}^2 - \sum_{j=1}^{jq/2} a_{jq}^2 \\
- \sum_{j=1}^{n} a_{jp} a_{jq} + (a_{pp} - a_{pq})(a_{qp} - a_{qq}) \\
= \sum_{j=1}^{Jp/2} a_{jp}^2 + \sum_{j=1}^{jq/2} a_{jq}^2 - \sum_{j=1}^{Jp/2} a_{jp} a_{jq} + (a_{pp} - a_{pq})(a_{qp} - a_{qq}) \\
... (95)
\]

Equation (95) does not require that the partially aggregated input coefficients, or the means of the partially aggregated input coefficients, be calculated. Similarity measures, for each pair of industries, can be calculated simply and quickly using equation (95), compared with the general procedure, described in Section 3.2. The general procedure required the construction of $nC_2$ matrices of partially aggregated input coefficients and their means, prior to the application of equation (94).

3.3.2 Updating the Similarity Matrix

Given a similarity matrix showing the total error sums of squares, $E_{pq}^h$, of all potential $h$ sector models ($h=n-1,n-2,...,2$), and given that sectors $r$ and $s$ ($r>s$) are merged to form a new sector, labelled $z(=s)$ in the actual $h$
sector model, the objective is to find the simplest method of updating the elements of the similarity matrix, so that they represent the total error sums of squares, $E_{pq}^{h-1}$, of all potential $h-1$ sector models.

Firstly, an updating equation is developed to find $E_{pq}^{h-1}$ where $p \neq z, r$ and $q \neq z, r$. Ideally, it would be desirable to find a relationship between $E_{pq}^{h-1}$ and $E_{pq}^{h}$, the latter being the element in the current similarity matrix, which needs to be updated.

Suppose the potential aggregation of sectors $p$ and $q$ ($p > q$) results in a new sector, labelled $t (= q)$. Hence, the element, $E_{pq}^{h}$, in the current similarity matrix is given by:

$$E_{pq}^{h} = \sum_{l=1}^{h} \sum_{j=1}^{h} \sum_{k=1}^{n} \sum_{j=1}^{h} \sum_{k=1}^{n} a_{jk}^{l} = \sum_{l=1}^{h} \sum_{j=1}^{h} \sum_{k=1}^{n} a_{jk}^{l}$$

where $l=1,2,\ldots,t,\ldots,r,\ldots,h$ and $j=1,2,\ldots,t,\ldots,r,\ldots,h$.

In equation (96), sectors $r$ and $s$ remain separate. In the updated element, $E_{pq}^{h-1}$, sectors $r$ and $s$ are aggregated to form sector $z$, hence:

$$E_{pq}^{h-1} = \sum_{l=1}^{h-1} \sum_{j=1}^{h-1} \sum_{k=1}^{n} \sum_{j=1}^{h-1} \sum_{k=1}^{n} a_{jk}^{l} = \sum_{l=1}^{h-1} \sum_{j=1}^{h-1} \sum_{k=1}^{n} a_{jk}^{l}$$

where $l=1,2,\ldots,t,\ldots,z,\ldots,h-1$ and $j=1,2,\ldots,t,\ldots,z,\ldots,h-1$. 
In order to simplify equation (97), we will make use of the following relationships:

\[ a_{zk} = a_{rk} + a_{sk} \]  \hspace{1cm} (98)

\[ \Sigma_{k=1}^{n} a_{zk} = \Sigma_{k=1}^{n} a_{rk} + \Sigma_{k=1}^{n} a_{sk} + 2 \Sigma_{k=1}^{n} r_{k} a_{rk} a_{sk} \]  \hspace{1cm} (99)

\[ a_{zt} = a_{rt} + a_{st} \]  \hspace{1cm} (100)

\[ a_{zt} = a_{rt} + a_{st} + 2 a_{rt} a_{st} \]  \hspace{1cm} (101)

Also

\[ n a_{zk} = n a_{rk} + n a_{sk} \]  \hspace{1cm} (102)

\[ n a_{zkz} = n a_{rkr} + n a_{skr} \]  \hspace{1cm} (103)

\[ n a_{z} = n a_{r} + n a_{s} \]  \hspace{1cm} (104)

\[ a_{zt} = a_{rt} + a_{st} + 2 a_{rt} a_{st} \]  \hspace{1cm} (105)

Also

\[ n a_{zkz} = n (a_{rkr} + a_{skr}) + n (a_{rks} + a_{skrs}) \]  \hspace{1cm} (106)

\[ n a_{zkz} = n a_{rkr} + n a_{skr} + 2 n a_{rks} a_{skrs} \]  \hspace{1cm} (107)

\[ n a_{z} = n a_{r} + n a_{s} \]  \hspace{1cm} (108)
\[
\begin{align*}
\frac{a_{zz}^2}{n_z^2} &= n_r^2(a_{rr}^2 + a_{sr}^2 + 2a_{rr}a_{sr}) + n_s^2(a_{rs}^2 + a_{ss}^2 + 2a_{rs}a_{ss}) \\
+ \frac{2n_r n_s (a_{rr} + a_{sr})(a_{rs} + a_{ss})}{n_z^2} \tag{109}
\end{align*}
\]

where \( n_z = n_r + n_s \) \tag{110}

Substituting equations (99), (101), (103), (105), (107), (109) and (110) into equation (97) gives:

\[
E_{pq}^{h-1} = \sum_{\ell=1}^{h} \sum_{j=1}^{h} \sum_{k=1}^{L} a_{jk\ell}^2 - \sum_{\ell=1}^{h} \sum_{j=1}^{h} \sum_{k=1}^{L} \frac{a_{jk\ell}^2}{n_z^2} + 2 \sum_{\ell=1}^{h} \sum_{j=1}^{h} \sum_{k=1}^{L} a_{rk\ell}a_{sk\ell} \times
\frac{h}{n_r + n_s} \tag{111}
\]

\[
+ \frac{n_r n_s}{n_r + n_s} \sum_{j=1}^{h} \frac{a_{js}^2}{n_r + n_s} - 2 \frac{n_r}{n_r + n_s} \frac{a_{rr}a_{sr} + 2n_s a_{rs}a_{ss}}{n_r + n_s} \tag{111}
\]

\[
- 2n_r n_s \sum_{j=1}^{h} \frac{a_{js}^2}{n_r + n_s} + 2n_r n_s \frac{(a_{rr}a_{rs} + a_{sr}a_{ss})}{n_r + n_s} \tag{111}
\]

\[
- 2n_r n_s \frac{(a_{rr} + a_{sr})(a_{rs} + a_{ss})}{n_r + n_s} \tag{111}
\]

where \( \ell=1,2,\ldots,t,\ldots,s,\ldots,r,\ldots,h \) and

\( j=1,2,\ldots,t,\ldots,s,\ldots,r,\ldots,h \).

Comparing equation (111) with equation (96), it is evident that the first two terms of equation (111) equal \( E_{pq}^h \).

The remaining terms of equation (111) can be simplified.
to give:

\[
E^{h-1}_{pq} = E^h_{pq} + 2 \sum_{\ell=1}^{h} \sum_{k=1}^{n_{\ell}} \bar{a}_{\ell k} \bar{a}_{\ell k} - 2 \sum_{\ell=1}^{h} n_{\ell} \bar{a}_{\ell q} \bar{a}_{\ell s} + \frac{n_{r} n_{s}}{n_{r} + n_{s}} \sum_{j=1}^{h} \frac{2 (\bar{a}_{jr} - \bar{a}_{js})^2 + 2n_{r} n_{s} (\bar{a}_{rr} - \bar{a}_{rs})(\bar{a}_{sr} - \bar{a}_{ss})}{n_{r} + n_{s}} \]

...(112)

where \(\ell=1,2,...,t,...,s,...,r,...,h\) and \(j=1,2,...,t,...,s,...,r,...,h\).

Equation (112) expresses the new element, \(E_{pq}^{h-1}\), in the similarity matrix as a function of its previous value, \(E_{pq}^h\). However, equation (112) is not the simplest updating equation. The search for the latter continues by considering the smallest element, \(E_{rs}^h\), in the current similarity matrix.

The element, \(E_{rs}^h\), is the total error sum of squares resulting from the actual aggregation of sectors \(r\) and \(s\), to form a new sector, labelled \(z\), in the actual \(h\) sector model. Its value is given by:

\[
E_{rs}^h = \sum_{\ell=1}^{h} \sum_{j=1}^{h} \sum_{k=1}^{n_{\ell}} a_{j k}^2 - \sum_{\ell=1}^{h} n_{\ell} a_{s}^2 \]

...(113)

where \(\ell=1,2,...,q,...,p,...,z,...,h\) and \(j=1,2,...,q,...,p,...,z,...,h\).

In equation (113) sectors \(p\) and \(q\) remain separate. In the updated element, \(E_{pq}^{h-1}\), sectors \(p\) and \(q\) are aggregated into sector \(t\) (see equation (97)). However, \(E_{pq}^{h-1}\) can be expressed as a function of \(E_{rs}^h\) by substituting the following
relationships into equation (97):

\[ a_{t \ell k} = a_{p k \ell} + a_{q k \ell} \]  \hspace{1cm} (\ell \neq t) \hspace{1cm} \ldots (114)

\[ \sum_{k=1}^{n} a_{t \ell k}^2 = \sum_{k=1}^{n} a_{p k \ell}^2 + \sum_{k=1}^{n} a_{q k \ell}^2 + 2 \sum_{k=1}^{n} a_{p k \ell} a_{q k \ell} \]  \hspace{1cm} (\ell \neq t) \hspace{1cm} \ldots (115)

\[ \ddot{a}_{t \ell} = \ddot{a}_{p \ell} + \ddot{a}_{q \ell} \]  \hspace{1cm} (\ell \neq t) \hspace{1cm} \ldots (116)

\[ \dddot{a}_{t \ell} = \dddot{a}_{p \ell} + \dddot{a}_{q \ell} + 2 \dddot{a}_{p \ell} \dddot{a}_{q \ell} \]  \hspace{1cm} (\ell \neq t) \hspace{1cm} \ldots (117)

Also

\[ \sum_{k=1}^{n} a_{t \ell k} = \sum_{k=1}^{n} a_{p \ell k} + \sum_{k=1}^{n} a_{q \ell k} \]  \hspace{1cm} (\ell \neq t) \hspace{1cm} \ldots (118)

\[ \sum_{k=1}^{n} a_{t \ell k}^2 = \sum_{k=1}^{n} a_{p \ell k}^2 + \sum_{k=1}^{n} a_{q \ell k}^2 \]  \hspace{1cm} (\ell \neq t) \hspace{1cm} \ldots (119)

\[ n_t \dddot{a}_{t \ell} = n_p \dddot{a}_{p \ell} + n_q \dddot{a}_{q \ell} \]  \hspace{1cm} (\ell \neq t) \hspace{1cm} \ldots (120)

\[ \dddot{a}_{t \ell} = \frac{n_t^2}{n_{t}^2} \dddot{a}_{p \ell} + \frac{n_t^2}{n_{t}^2} \dddot{a}_{q \ell} + 2 n_{t}^2 \frac{n_p n_q \dddot{a}_{p \ell} \dddot{a}_{q \ell}}{n_t^2} \]  \hspace{1cm} (\ell \neq t) \hspace{1cm} \ldots (121)

Also

\[ \sum_{k=1}^{n} a_{t \ell k} = \sum_{k=1}^{n} a_{p \ell k} + \sum_{k=1}^{n} a_{q \ell k} \]  \hspace{1cm} (\ell \neq t) \hspace{1cm} \ldots (122)

\[ \sum_{k=1}^{n} a_{t \ell k}^2 = \sum_{k=1}^{n} a_{p \ell k}^2 + \sum_{k=1}^{n} a_{q \ell k}^2 + 2 \sum_{k=1}^{n} a_{p \ell k} a_{q \ell k} \]  \hspace{1cm} (\ell \neq t) \hspace{1cm} \ldots (123)

\[ n_t \dddot{a}_{t \ell} = n_p \dddot{a}_{p \ell} + \dddot{a}_{q \ell} \]  \hspace{1cm} (\ell \neq t) \hspace{1cm} \ldots (124)

\[ \dddot{a}_{t \ell} = \frac{n_t^2}{n_{t}^2} \dddot{a}_{p \ell} + \frac{n_t^2}{n_{t}^2} \dddot{a}_{q \ell} + 2 \dddot{a}_{p \ell} \dddot{a}_{q \ell} \]  \hspace{1cm} (\ell \neq t) \hspace{1cm} \ldots (125)
where \( n_t = n_p + n_q \) \hspace{1cm} \ldots (126)

Substituting equations (115), (117), (119), (121), (123), (125) and (126) into equation (97) gives:

\[
E_{pq}^{h-1} = \sum_{\ell=1}^{h} \sum_{j=1}^{h} n_{\ell} \sum_{k=1}^{n_{\ell}} a_{jk\ell} - \sum_{\ell=1}^{h} n_{\ell} \sum_{j=1}^{n_{\ell}} \sum_{k=1}^{n_{\ell}} a_{j\ell k} + 2 \sum_{\ell=1}^{h} \sum_{k=1}^{n_{\ell}} \sum_{\ell=1}^{h} a_{pk\ell} a_{qk\ell} \\
-2 \sum_{\ell=1}^{h} \sum_{j=1}^{n_{\ell}} \sum_{k=1}^{n_{\ell}} \frac{a_{p\ell q}}{n_{p+q}} + 2n_{p} \frac{a_{p\ell q}}{n_{p+q}} + 2n_{q} \frac{a_{q\ell p}}{n_{p+q}} + \frac{h}{n_{p+q}} \sum_{j=1}^{n_{p+q}} \frac{a_{qj\ell}}{n_{p+q}} \sum_{k=1}^{n_{p+q}} \frac{a_{pkj}}{n_{p+q}} \\
+ \frac{n_{p} n_{q}}{n_{p+q}} \sum_{j=1}^{h} \frac{a_{jq\ell}}{n_{p+q}} - 2n_{p} \frac{a_{p\ell q}}{n_{p+q}} a_{p\ell q} + 2n_{q} \frac{a_{q\ell p}}{n_{p+q}} a_{q\ell p} \\
- 2n_{p} n_{q} \sum_{j=1}^{h} \frac{a_{jpq}}{n_{p+q}} a_{jpq} + \frac{n_{p} n_{q}}{n_{p+q}} \sum_{j=1}^{h} \frac{a_{jpq}}{n_{p+q}} a_{jpq} + \sum_{j=1}^{h} \frac{a_{jpq}}{n_{p+q}} a_{jpq} + \frac{n_{p} n_{q}}{n_{p+q}} \sum_{j=1}^{h} \frac{a_{jpq}}{n_{p+q}} a_{jpq} \\
- 2n_{p} n_{q} (\frac{a_{ppq}}{n_{p+q}} + \frac{a_{qpq}}{n_{p+q}}) (\frac{a_{ppq}}{n_{p+q}} + \frac{a_{qpq}}{n_{p+q}}) \hspace{1cm} \ldots (127)
\]

where \( \ell=1,2,\ldots,q,\ldots,p,\ldots,z,\ldots,h \) and \( j=1,2,\ldots,q,\ldots,p,\ldots,z,\ldots,h \).

Comparing equation (127) with equation (113), it is evident that the first two terms of equation (127) equal \( E_{rs}^{h} \).

The remaining terms can be simplified to give:

\[
E_{pq}^{h-1} = E_{rs}^{h} + 2 \sum_{\ell=1}^{h} \sum_{k=1}^{n_{\ell}} a_{pk\ell} a_{qk\ell} - 2 \sum_{\ell=1}^{h} \sum_{k=1}^{n_{\ell}} a_{pk\ell} a_{qk\ell} + \\
\frac{n_{p} n_{q}}{n_{p+q}} \sum_{j=1}^{h} (\frac{a_{jpq}}{n_{p+q}} - \frac{a_{jqp}}{n_{p+q}})^2 + \frac{n_{p} n_{q}}{n_{p+q}} (\frac{a_{ppq}}{n_{p+q}} - \frac{a_{qpp}}{n_{p+q}}) (\frac{a_{qpp}}{n_{p+q}} - \frac{a_{qpp}}{n_{p+q}}) \hspace{1cm} \ldots (128)
\]

where \( \ell=1,2,\ldots,q,\ldots,p,\ldots,z,\ldots,h \) and \( j=1,2,\ldots,q,\ldots,p,\ldots,z,\ldots,h \).
Equation (128) can be used to express $E_{rs}^h$ as a function of the total error sum of squares in the $h+1$ sector model, $E_{ab}^{h+1}$, "a" and "b" being the two sectors, which are merged prior to the aggregation of sectors $r$ and $s$:

$$E_{rs}^h = E_{ab}^{h+1} + 2 \sum_{k=1}^{h+1} \sum_{l=1}^{n} a_{rl} a_{sk} - 2 \sum_{l=1}^{h+1} n_{rl} \bar{a}_{rl} \bar{a}_{sl}$$

where $l = 1, 2, \ldots, q, \ldots, p, \ldots, s, \ldots, r, \ldots, h$ and

$$j = 1, 2, \ldots, q, \ldots, p, \ldots, s, \ldots, r, \ldots, h.$$

Combining sectors $p$ and $q$ in equation (129) to form sector $t$, gives:

$$E_{rs}^h = E_{ab}^{h+1} + 2 \sum_{k=1}^{h} \sum_{l=1}^{n} a_{rl} a_{sk} - 2 \sum_{l=1}^{h} n_{rl} \bar{a}_{rl} \bar{a}_{sl} +$$

$$\frac{n_{rs}}{n_{r} + n_{s}} \left( \bar{a}_{jr} - \bar{a}_{js} \right)^2 + \frac{2n_{rs}}{n_{r} + n_{s}} (\bar{a}_{rr} - \bar{a}_{rs})(\bar{a}_{sr} - \bar{a}_{ss})$$

$$+ 2(n_{rt} \bar{a}_{st} - n_{rp} \bar{a}_{sp} - n_{rq} \bar{a}_{sq})$$

$$\frac{n_{rs}}{n_{r} + n_{s}} \left( \bar{a}_{pr} - \bar{a}_{ps} \right)^2 + (\bar{a}_{qr} - \bar{a}_{qs})^2 - (\bar{a}_{tr} - \bar{a}_{ts})^2$$

where $l = 1, 2, \ldots, t, \ldots, s, \ldots, r, \ldots, h$ and

$$j = 1, 2, \ldots, t, \ldots, s, \ldots, r, \ldots, h.$$
Comparing equation (130) to equation (112), it is evident that terms two to five in equation (130) are equal to \( E_{pq}^{h-1} - E_{pq}^h \), hence:

\[
E_{rs}^h = E_{ab}^{h+1} + E_{pq}^{h-1} - E_{pq}^h + 2(n_t \tilde{a}_{rt} \tilde{a}_{st} - n_p \tilde{a}_{rp} \tilde{a}_{sp} - n_q \tilde{a}_{rq} \tilde{a}_{sq})
+ \frac{n_r n_s}{n_r + n_s} \left\{ (\tilde{a}_{pr} - \tilde{a}_{ps})^2 + (\tilde{a}_{qr} - \tilde{a}_{qs})^2 - (\tilde{a}_{tr} - \tilde{a}_{ts})^2 \right\}
\]

\[\ldots(131)\]

Now \( n_t \tilde{a}_{rt} = n_p \tilde{a}_{rp} + n_q \tilde{a}_{rq} \) \[\ldots(132)\]
and \( n_t \tilde{a}_{st} = n_p \tilde{a}_{sp} + n_q \tilde{a}_{sq} \) \[\ldots(133)\]
so

\[
\begin{align*}
&n_t \tilde{a}_{rt} \tilde{a}_{st} - n_p \tilde{a}_{rp} \tilde{a}_{sp} - n_q \tilde{a}_{rq} \tilde{a}_{sq} \\
&= \frac{n_p \tilde{a}_{rp} \tilde{a}_{sp}}{n_t} + \frac{n_q \tilde{a}_{rq} \tilde{a}_{sq}}{n_t} + \frac{n_p \tilde{a}_{qr} \tilde{a}_{ts}}{n_t} + \frac{n_q \tilde{a}_{qt} \tilde{a}_{qs}}{n_t}
\end{align*}
\]

\[\ldots(134)\]

Also \( \tilde{a}_{tr} = \tilde{a}_{pr} + \tilde{a}_{qr} \) \[\ldots(135)\]

and \( \tilde{a}_{ts} = \tilde{a}_{ps} + \tilde{a}_{qs} \) \[\ldots(136)\]
so

\[
(\tilde{a}_{pr} - \tilde{a}_{ps})^2 + (\tilde{a}_{qr} - \tilde{a}_{qs})^2 - (\tilde{a}_{tr} - \tilde{a}_{ts})^2
\]

\[= -2(\tilde{a}_{pr} - \tilde{a}_{ps})(\tilde{a}_{qr} - \tilde{a}_{qs})\]

\[\ldots(137)\]

Substituting equations (134) and (137) into equation (131) gives:

\[
E_{rs}^h = E_{ab}^{h+1} + E_{pq}^{h-1} - E_{pq}^h - 2n_t n_s (\tilde{a}_{pr} - \tilde{a}_{ps}) (\tilde{a}_{sp} - \tilde{a}_{sq})
- 2n_r n_s (\tilde{a}_{pr} - \tilde{a}_{ps}) (\tilde{a}_{qr} - \tilde{a}_{qs})
\]

\[\ldots(138)\]
Finally, rearranging equation (138) gives:

\[
E_{pq}^{h-1} = E_{rs}^h + E_{pq}^h - E_{ab}^{h+1} + \frac{2n_r n_q (\bar{a}_{rp} - \bar{a}_{rq}) (\bar{a}_{sp} - \bar{a}_{sq})}{n_p + n_q} + 2n_r n_s (\bar{a}_{pr} - \bar{a}_{ps})(\bar{a}_{qr} - \bar{a}_{qs})
\] ...(139)

Equation (139) expresses the new element, \(E_{pq}^{h-1}\), in the similarity matrix as a function of its previous value, \(E_{pq}^h\), the total error sum of squares for the \(h\) sector model, \(E_{rs}^h\), and the total error sum of squares for the \(h+1\) sector model, \(E_{ab}^{h+1}\). Equation (139) is used to calculate all elements, \(E_{pq}^{h-1}\), in the revised similarity matrix, where \(p \neq z, r\) and \(q \neq z, r\).

An updating equation is also required to find \(E_{zp}^{h-1}\), that is, the total error sum of squares resulting from the potential merger of any sector, labelled \(p\), and the most recently created sector, labelled \(z\). Sector \(z\) is formed as a result of the merger of sectors \(r\) and \(s\) in the \(h+1\) sector model. The potential merger of sectors \(p\) and \(z\) in the \(h\) sector model gives rise to a new sector, labelled \(t(=z)\).

From equation (92) we see that \(E_{zp}^{h-1}\) is of the form:

\[
E_{zp}^{h-1} = \sum_{l=1}^{h-1} \sum_{j=1}^{h-1} \sum_{k=1}^{n_l} \sum_{j=1}^{a_{jl}} a_{jkl}^{2} - \sum_{l=1}^{h-1} \sum_{j=1}^{h-1} a_{jl}^{2} - \sum_{l=1}^{h-1} \sum_{j=1}^{a_{jl}} a_{jkl}^{2} \quad ...(140)
\]

where \(l = 1, 2, \ldots, t, \ldots, h-1\) and \(j = 1, 2, \ldots, t, \ldots, h-1\).
In order to simplify equation (140), we employ the following relationships:

\[ a_{t\ell} = a_{z\ell} + a_{p\ell} \]  \hspace{1cm} (\ell \neq t) \hspace{1cm} \ldots (141)

\[ \sum_{k=1}^{n} a_{t\ell} = \sum_{k=1}^{n} a_{z\ell} + \sum_{k=1}^{n} a_{p\ell} + 2 \sum_{k=1}^{n} a_{z\ell} a_{p\ell} \]  \hspace{1cm} (\ell \neq t) \hspace{1cm} \ldots (142)

\[ \bar{a}_{t\ell} = \bar{a}_{z\ell} + \bar{a}_{p\ell} \]  \hspace{1cm} (\ell \neq t) \hspace{1cm} \ldots (143)

\[ \bar{a}_{t\ell} = \bar{a}_{z\ell}^2 + \bar{a}_{p\ell}^2 + 2\bar{a}_{z\ell} \bar{a}_{p\ell} \]  \hspace{1cm} (\ell \neq t) \hspace{1cm} \ldots (144)

Also

\[ \sum_{k=1}^{n} a_{t\ell} = \sum_{k=1}^{n} a_{z\ell} + \sum_{k=1}^{n} a_{p\ell} \]  \hspace{1cm} (\ell \neq t) \hspace{1cm} \ldots (145)

\[ \sum_{k=1}^{n} a_{t\ell}^2 = \sum_{k=1}^{n} a_{z\ell}^2 + \sum_{k=1}^{n} a_{p\ell}^2 \]  \hspace{1cm} (\ell \neq t) \hspace{1cm} \ldots (146)

\[ n_t \bar{a} = n_z \bar{a} + n_p \bar{a} \]  \hspace{1cm} (\ell \neq t) \hspace{1cm} \ldots (147)

\[ \bar{a}_{t\ell} = \frac{n_z^2 \bar{a}_{z\ell}^2 + n_p^2 \bar{a}_{p\ell}^2 + 2n_z n_p \bar{a}_{z\ell} \bar{a}_{p\ell}}{n_t^2} \]  \hspace{1cm} (\ell \neq t) \hspace{1cm} \ldots (148)

Also

\[ \sum_{k=1}^{n} a_{t\ell} = \sum_{k=1}^{n} (a_{z\ell} + a_{p\ell}) + \sum_{k=1}^{n} (a_{z\ell} + a_{p\ell}) \]  \hspace{1cm} (\ell \neq t) \hspace{1cm} \ldots (149)

\[ \sum_{k=1}^{n} a_{t\ell}^2 = \sum_{k=1}^{n} a_{z\ell}^2 + \sum_{k=1}^{n} a_{p\ell}^2 + 2 \sum_{k=1}^{n} a_{z\ell} a_{p\ell} \]  \hspace{1cm} (\ell \neq t) \hspace{1cm} \ldots (150)

\[ n_t \bar{a} = n_z \bar{a}_{z\ell} + n_p \bar{a}_{p\ell} \]  \hspace{1cm} (\ell \neq t) \hspace{1cm} \ldots (151)

\[ \bar{a}_{t\ell} = \frac{n_z^2 \bar{a}_{z\ell}^2 + n_p^2 \bar{a}_{p\ell}^2 + 2n_z n_p \bar{a}_{z\ell} \bar{a}_{p\ell}}{n_t^2} \]  \hspace{1cm} (\ell \neq t) \hspace{1cm} \ldots (152)
where \( n_{\ell} = n_z + n_p \) \( \ldots \) (153)

Substituting equations (142), (144), (146), (148), (150), (152) and (153) into equation (140) gives:

\[
E_{zp}^{h-1} = \sum_{\ell=1}^{n} \sum_{j=1}^{n} n_{\ell} a_{jk}^{\ell} + \sum_{l=1}^{h} n_{l} \sum_{j=1}^{n} a_{jl}^{n} + 2 \sum_{l=1}^{h} n_{l} a_{zkl} a_{pk\ell}
\]

\[
- 2 \sum_{\ell=1}^{h} n_{\ell} a_{z\ell} a_{p\ell} + 2n_{z} a_{zz} a_{pz} + 2n_{p} a_{zp} a_{pp} + \frac{n_{z} n_{p}}{n_{z} + n_{p}} \sum_{j=1}^{h} \sum_{\ell=1}^{n} a_{\ell j}^{2}
\]

\[
+ \frac{n_{z} n_{p}}{n_{z} + n_{p}} \sum_{j=1}^{h} a_{jp}^{2} - 2n_{z} a_{zz} a_{pz} - 2n_{p} a_{zp} a_{pp}
\]

\[
- 2n_{z} n_{p} \sum_{j=1}^{h} \sum_{\ell=1}^{n} a_{jz} a_{jp} + 2n_{z} n_{p} (a_{zz} a_{zp} + a_{pz} a_{pp}) + \frac{n_{z} n_{p}}{n_{z} + n_{p}} (a_{zz} + a_{pz}) (a_{zp} + a_{pp})
\] \( \ldots \) (154)

where \( \ell=1,2,\ldots,p,\ldots,z,\ldots,h \) and \( j=1,2,\ldots,p,\ldots,z,\ldots,h \).

Comparing equation (154) with equation (113), it is evident that the first two terms of equation (154) equal \( E_{rs}^{h} \).

The remaining terms can be simplified to give:

\[
E_{zp}^{h-1} = E_{rs}^{h} + 2 \sum_{l=1}^{h} n_{l} a_{zkl} a_{pk\ell} - 2 \sum_{l=1}^{h} n_{l} a_{z\ell} a_{p\ell} + \frac{n_{z} n_{p}}{n_{z} + n_{p}} \sum_{j=1}^{h} (a_{jj} - a_{jp})^{2} + 2n_{z} n_{p} (a_{zz} - a_{zp}) (a_{pz} - a_{pp})
\] \( \ldots \) (155)

where \( \ell=1,2,\ldots,p,\ldots,z,\ldots,h \) and \( j=1,2,\ldots,p,\ldots,z,\ldots,h \).
Since equation (155) cannot be simplified any further, it is used to update those elements of the similarity matrix, which involve the most recently created sector, \( z \).

3.3.3 The Computer Routine

The availability of the two updating equations, numbered (139) and (155) above, greatly reduces the amount of computation, which is necessary when the clustering algorithm is applied to a specific input-output model. Note that the general updating procedure, described in Section 3.2, requires the construction of \( hC_2 \) matrices of partially aggregated input coefficients and their means in order to determine the similarity matrix, which is used to reduce the \( h \) sector model to \( h-1 \) sectors \( (h=n-1,n-2,\ldots,2) \). However, using equations (139) and (155) to update the similarity matrix requires the construction of only one matrix of partially aggregated input coefficients and their means at each stage of the clustering procedure.

In fact, the matrices of partially aggregated input coefficients and the means of partially aggregated input coefficients can be updated themselves in a recursive manner at each stage of the clustering procedure. Initially, each of the \( n \) sectors contains one industry, so the partially aggregated input coefficients, the means of the partially aggregated input coefficients and the technical coefficients are all equal.
Therefore, the two matrices are initialised as follows:

\[
a_{ijkl} = a_{ijkl} \quad (i=1, j=1,2,\ldots,n; \quad k=1, \ldots,n)\\
\tilde{a}_{ij} = a_{ijkl} \quad (i=1, j=1,2,\ldots,n; \quad k=1, \ldots,n)
\]

The number of sectors, \( h \), is initially set to \( n \) (the number of industries) and the number of industries in each sector, \( n_{Xl} \), \( (l=1,2,\ldots,h) \) is set to unity.

Following the merger of sectors \( r \) and \( s \) in the \( h \) sector model to form a new sector, labelled \( z(=s) \), in the \( h-1 \) sector model, the partially aggregated input coefficients and their means are updated using the following sequence of equations:

\[
n_z = n_r + n_s \\ a_{zkl} = a_{rkl} + a_{skl} \quad k=1,2,\ldots,n_{z} \quad l=1,2,\ldots,h-1 \\ a_{rkl} = 0 \quad k=1,2,\ldots,n_{z} \quad l=1,2,\ldots,h-1 \\ a_{skl} = a_{rkl} + a_{skl} \quad l=1,2,\ldots,h-1 \\ a_{rkl} = 0 \quad l=1,2,\ldots,h-1 \\ a_{rj} = (n_r \tilde{a}_{rj} + n_s \tilde{a}_{sj})/n_z \quad j=1,2,\ldots,h-1 \\ a_{s} = 0 \quad j=1,2,\ldots,h-1
\]

The entire clustering algorithm is set out, in flowchart form, in Figure 3.3.1. A description and listing of a computer program, incorporating the clustering algorithm, is given in the Appendix at the end of this thesis.
1. Set matrix of partially aggregated input coefficients equal to the initial technical coefficients.

2. Set matrix of means of the partially aggregated input coefficients equal to the initial technical coefficients.

3. Set up the initial similarity matrix, using equation (95).

4. Set elements of the similarity matrix to infinity, as required to prevent aggregation of certain industries. Also flag elements of the similarity matrix, as required to force the aggregation of certain industries.

5. Set the number of sectors in the model, $h$, equal to the number of industries, $n$, in the initial input-output model. That is, $h=n$.

6. Set the number of industries in each sector, $n_s$, equal to unity. That is, $n_s = 1$ for $t=1,2,\ldots,h$.

7. Set the number of sectors required in the final aggregated model to $m$.

8. Set the current total error sum of squares to zero. That is, $E_{0b}^h = 0$.

9. $h = h-1$

10. Search the similarity matrix for the next pair of flagged sectors.

   flag found

   11. Determine the total error sum of squares in the $h$ sector model, $E_{rs}^h$.

   12. Record the aggregation of sectors $r$ and $s$.

   13. Remove flag.

   14. Search the similarity matrix for the smallest element and note the corresponding sectors $r$ and $s$.

   15. Determine the total error sum of squares in the $h$ sector model, $E_{rs}^h$.

flag not found

   16. Record aggregation of sectors $r$ and $s$.

17. $h=m$?

   yes --- STOP

   no

18. Update those elements of the similarity matrix which do not involve sectors $r$ and $s$, using equation (135).

19. Update the number of industries in the newly formed sector and in sector $r$. That is, $n_r + n_s$ and $n_r > 0$.

20. Update the matrix of partially aggregated input coefficients, using equations (156) to (158) and update the matrix of means of the partially aggregated input coefficients, using equations (159) to (162).

21. Update elements of the similarity matrix, involving sector $s$, using equation (155) and delete elements involving sector $r$ from the similarity matrix.

22. Update the current total error sum of squares. That is, $E_{ab}^h + E_{rs}^h$. 

Figure 3.3.1
Flowchart of the Clustering Algorithm
A methodology has been developed to aggregate industries of an input-output model into sectors, using similarity of partially aggregated input coefficients as the criterion for aggregation. The methodology takes the form of a clustering algorithm. Beginning with an initial input-output model, consisting of n sectors of one industry each, a measure of similarity of partially aggregated input coefficients is calculated for each possible pair of sectors. The two sectors with the most similar partially aggregated input coefficients are merged to form a new sector, reducing the number of sectors in the model to n-1. Similarity measures are recalculated for each possible pair of sectors in the n-1 sector model, the most similar pair of sectors are merged, and the process continues until the model is aggregated into the desired number of sectors, or until all the original sectors have been merged into one.

The implementation of the clustering algorithm is aided by the development of three equations, labelled (95), (139) and (155) in Section 3.3 above. The first of these equations provides a convenient way of calculating the initial similarity measures between all possible pairs of industries in the original input-output model. The latter two equations simplify the process of revising the similarity matrix at each stage of the clustering procedure.
Finally, the aggregation procedure allows the analyst to place constraints upon the intersectoral model. It is possible to force the aggregation of certain industries into the same sector, to ensure that specific industries are not aggregated into the same sector and to ensure that certain key industries remain isolated in the intersectoral model.
CHAPTER 4
APPLICATION OF THE METHODOLOGY FOR AGGREGATION

4.1 The Results of Aggregation Based on Similarity of Partially Aggregated Input Coefficients

The 1965-66 input-output model of the New Zealand economy, constructed by the New Zealand Department of Statistics, contains 109 industries. For easy reference, these industries are listed in the fold-out Appendix A at the end of this chapter. The clustering algorithm, developed in Chapter 3, was used to aggregate these industries, in 65 successive stages, until they were merged into 44 sectors. In this section, the sectors so produced are compared to those used by the New Zealand Department of Statistics [(1974), Parts 1 and 2] in their 44 sector model of the 1965-66 New Zealand economy.

The sectors, which were formed in the process of aggregating the 109 industries into 44 sectors, are listed in Table 4.1.1. This table also gives the value of the total, within sector, sum of squared deviations of partially aggregated input coefficients from their means (that is, the total error sum of squares) for each of the 108 to 44 sector models, formed during the aggregation procedure. An examination of Table 4.1.1 reveals the degree of discrimination between sectors. The larger the percentage of sectors, which are formed at high levels of similarity, out of the total of 65 mergers
Based on Similarity of Partially Aggregated Input Coefficients

<table>
<thead>
<tr>
<th>Stage</th>
<th>Number of Sectors</th>
<th>Aggregation of Industries</th>
<th>Total Error Sum of Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>108</td>
<td>108 109</td>
<td>0.00000</td>
</tr>
<tr>
<td>2</td>
<td>107</td>
<td>78 81</td>
<td>0.00032</td>
</tr>
<tr>
<td>3</td>
<td>106</td>
<td>62 86</td>
<td>0.00078</td>
</tr>
<tr>
<td>4</td>
<td>105</td>
<td>71 72</td>
<td>0.00127</td>
</tr>
<tr>
<td>5</td>
<td>104</td>
<td>71 72 78 81</td>
<td>0.00185</td>
</tr>
<tr>
<td>6</td>
<td>103</td>
<td>53 56</td>
<td>0.00242</td>
</tr>
<tr>
<td>7</td>
<td>102</td>
<td>71 72 78 81 76</td>
<td>0.00306</td>
</tr>
<tr>
<td>8</td>
<td>101</td>
<td>62 86 83</td>
<td>0.00375</td>
</tr>
<tr>
<td>9</td>
<td>100</td>
<td>25 26</td>
<td>0.00446</td>
</tr>
<tr>
<td>10</td>
<td>99</td>
<td>54 67</td>
<td>0.00539</td>
</tr>
<tr>
<td>11</td>
<td>98</td>
<td>85 106</td>
<td>0.00635</td>
</tr>
<tr>
<td>12</td>
<td>97</td>
<td>49 75</td>
<td>0.00735</td>
</tr>
<tr>
<td>13</td>
<td>96</td>
<td>27 34</td>
<td>0.00838</td>
</tr>
<tr>
<td>14</td>
<td>95</td>
<td>22 24</td>
<td>0.00954</td>
</tr>
<tr>
<td>15</td>
<td>94</td>
<td>1 21</td>
<td>0.01078</td>
</tr>
<tr>
<td>16</td>
<td>93</td>
<td>61 92</td>
<td>0.01203</td>
</tr>
<tr>
<td>17</td>
<td>92</td>
<td>105 108 109</td>
<td>0.01336</td>
</tr>
<tr>
<td>18</td>
<td>91</td>
<td>28 33</td>
<td>0.01470</td>
</tr>
<tr>
<td>19</td>
<td>90</td>
<td>82 84</td>
<td>0.01614</td>
</tr>
<tr>
<td>20</td>
<td>89</td>
<td>49 75 99</td>
<td>0.01757</td>
</tr>
<tr>
<td>21</td>
<td>88</td>
<td>79 87</td>
<td>0.01907</td>
</tr>
<tr>
<td>22</td>
<td>87</td>
<td>19 31</td>
<td>0.02068</td>
</tr>
<tr>
<td>23</td>
<td>86</td>
<td>53 56 54 67</td>
<td>0.02232</td>
</tr>
<tr>
<td>24</td>
<td>85</td>
<td>49 75 99 85 106</td>
<td>0.02390</td>
</tr>
<tr>
<td>25</td>
<td>84</td>
<td>90 91</td>
<td>0.02509</td>
</tr>
<tr>
<td>26</td>
<td>83</td>
<td>20 65</td>
<td>0.02663</td>
</tr>
<tr>
<td>27</td>
<td>82</td>
<td>27 34 30</td>
<td>0.02825</td>
</tr>
<tr>
<td>28</td>
<td>81</td>
<td>88 95</td>
<td>0.03003</td>
</tr>
<tr>
<td>29</td>
<td>80</td>
<td>62 86 83 105 108 109</td>
<td>0.03188</td>
</tr>
<tr>
<td>30</td>
<td>79</td>
<td>61 92 90 91</td>
<td>0.03388</td>
</tr>
<tr>
<td>31</td>
<td>78</td>
<td>71 72 78 81 76 100</td>
<td>0.03588</td>
</tr>
<tr>
<td>32</td>
<td>77</td>
<td>13 18</td>
<td>0.03792</td>
</tr>
<tr>
<td>33</td>
<td>76</td>
<td>46 47</td>
<td>0.03996</td>
</tr>
<tr>
<td>34</td>
<td>75</td>
<td>73 77</td>
<td>0.04204</td>
</tr>
<tr>
<td>35</td>
<td>74</td>
<td>53 56 54 67 104</td>
<td>0.04433</td>
</tr>
<tr>
<td>36</td>
<td>73</td>
<td>17 89</td>
<td>0.04614</td>
</tr>
<tr>
<td>37</td>
<td>72</td>
<td>27 34 30 42</td>
<td>0.04856</td>
</tr>
<tr>
<td>38</td>
<td>71</td>
<td>59 61 92 90 91</td>
<td>0.05104</td>
</tr>
<tr>
<td>39</td>
<td>70</td>
<td>64 66</td>
<td>0.05354</td>
</tr>
<tr>
<td>40</td>
<td>69</td>
<td>19 31 25 26</td>
<td>0.05606</td>
</tr>
<tr>
<td>41</td>
<td>68</td>
<td>71 72 78 81 76 100 79 87</td>
<td>0.05861</td>
</tr>
<tr>
<td>42</td>
<td>67</td>
<td>73 77 101</td>
<td>0.06118</td>
</tr>
<tr>
<td>43</td>
<td>66</td>
<td>71 72 78 81 76 100 79 87 98</td>
<td>0.06407</td>
</tr>
<tr>
<td>44</td>
<td>65</td>
<td>28 33 49 75 99 85 106</td>
<td>0.06713</td>
</tr>
<tr>
<td>45</td>
<td>64</td>
<td>45 46 47</td>
<td>0.07091</td>
</tr>
<tr>
<td>46</td>
<td>63</td>
<td>15 59 61 92 90 91</td>
<td>0.07477</td>
</tr>
<tr>
<td>47</td>
<td>62</td>
<td>1 21 22 24</td>
<td>0.07853</td>
</tr>
<tr>
<td>48</td>
<td>61</td>
<td>19 31 25 26 29</td>
<td>0.08016</td>
</tr>
<tr>
<td>49</td>
<td>60</td>
<td>20 65 62 86 83 105 108 109</td>
<td>0.08405</td>
</tr>
<tr>
<td>50</td>
<td>59</td>
<td>64 66</td>
<td>0.08805</td>
</tr>
<tr>
<td>51</td>
<td>58</td>
<td>3 44 48</td>
<td>0.09050</td>
</tr>
<tr>
<td>52</td>
<td>57</td>
<td>3 44 48 71 72 78 81 76 100 79 87 98</td>
<td>0.09240</td>
</tr>
<tr>
<td>53</td>
<td>56</td>
<td>1 21 22 24 23</td>
<td>0.09642</td>
</tr>
<tr>
<td>54</td>
<td>55</td>
<td>3 44 48 71 72 78 81 76 100 79 87 98 102</td>
<td>0.10042</td>
</tr>
<tr>
<td>55</td>
<td>54</td>
<td>17 89 28 33 49 75 99 85 106</td>
<td>0.10444</td>
</tr>
<tr>
<td>56</td>
<td>53</td>
<td>4 88 95</td>
<td>0.10860</td>
</tr>
<tr>
<td>57</td>
<td>52</td>
<td>3 44 48 71 72 78 81 76 100 79 87 98 102 4 88 95</td>
<td>0.10930</td>
</tr>
<tr>
<td>58</td>
<td>51</td>
<td>64 66 70</td>
<td>0.11280</td>
</tr>
<tr>
<td>59</td>
<td>50</td>
<td>13 18 15 59 61 92 90 91</td>
<td>0.11735</td>
</tr>
<tr>
<td>60</td>
<td>49</td>
<td>64 66 70 82 84</td>
<td>0.12213</td>
</tr>
<tr>
<td>61</td>
<td>48</td>
<td>2 27 34 30 42</td>
<td>0.12730</td>
</tr>
<tr>
<td>62</td>
<td>47</td>
<td>2 27 34 30 42 53 56 54 67 104</td>
<td>0.13274</td>
</tr>
<tr>
<td>63</td>
<td>46</td>
<td>33 18 15 59 61 92 90 91 52</td>
<td>0.13846</td>
</tr>
<tr>
<td>64</td>
<td>45</td>
<td>39 73 77 101</td>
<td>0.14434</td>
</tr>
<tr>
<td>65</td>
<td>44</td>
<td>20 65 62 86 83 105 108 109 107</td>
<td>0.15058</td>
</tr>
</tbody>
</table>
required to produce the 44 sector model, the greater is the degree of discrimination between sectors. On the other hand, if approximately the same percentage of sectors are formed over the entire range of the similarity scale, then there is little discrimination between sectors. The number and percentage of sectors formed within various ranges of the similarity scale are presented in Table 4.1.2. A high degree of discrimination exists between sectors, as the 109 industries are aggregated into 68 sectors, but the degree of discrimination between sectors formed later in the aggregation procedure, is not very great.

Table 4.1.2
Number and Percentage of Sectors Formed Within Various Ranges of the Similarity Scale

<table>
<thead>
<tr>
<th>Similarity Measure</th>
<th>Number of Sectors Formed</th>
<th>Percentage of Sectors Formed</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00-&lt;0.02</td>
<td>21</td>
<td>32.3</td>
</tr>
<tr>
<td>0.02-&lt;0.04</td>
<td>12</td>
<td>18.5</td>
</tr>
<tr>
<td>0.04-&lt;0.06</td>
<td>8</td>
<td>12.3</td>
</tr>
<tr>
<td>0.06-&lt;0.08</td>
<td>6</td>
<td>9.2</td>
</tr>
<tr>
<td>0.08-&lt;0.10</td>
<td>6</td>
<td>9.2</td>
</tr>
<tr>
<td>0.10-&lt;0.12</td>
<td>6</td>
<td>9.2</td>
</tr>
<tr>
<td>0.12-&lt;0.14</td>
<td>4</td>
<td>6.2</td>
</tr>
<tr>
<td>0.14-&lt;0.16</td>
<td>2</td>
<td>3.1</td>
</tr>
</tbody>
</table>
The 44 sector classification of industries, given by the New Zealand Department of Statistics and hereafter referred to as Model 1, is given in Table 4.1.3. The 44 sector model, produced by the clustering algorithm and hereafter referred to as Model 2, is given in Table 4.1.4. The total, within sector, sum of squared deviations of the partially aggregated input coefficients from their means for Model 2 is equal to 0.15058. The same measure was calculated for Model 1, by forcing the algorithm to produce the Department of Statistics' classification. Its value is 1.46481. Hence, it is concluded that the grouping of industries, produced by the aggregation methodology, results in partially aggregated input coefficients within sectors, which are more similar than those associated with the Department of Statistics' grouping. Consequently, Model 2 should result in less aggregation bias in forecasts of gross output than Model 1.

There are, however, a number of similarities and differences between the two models, which require some comment.

Firstly, focussing upon the similarities, it is observed that the following industries remain isolated in both models:

10. Fruit and Vegetable Preserving
32. Footwear - not Rubber
93. Residential Building
94. Commercial Building
97. Electricity and Gas
103. Air Transport
<table>
<thead>
<tr>
<th>Sector</th>
<th>Industries Comprising the Sector</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>5  6</td>
</tr>
<tr>
<td>6</td>
<td>7  11  12  13  14  15</td>
</tr>
<tr>
<td>7</td>
<td>8  9</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>16  17  18</td>
</tr>
<tr>
<td>10</td>
<td>19</td>
</tr>
<tr>
<td>11</td>
<td>20  21</td>
</tr>
<tr>
<td>12</td>
<td>22  23  24</td>
</tr>
<tr>
<td>13</td>
<td>25  26  27  28  29  30  31  33  34</td>
</tr>
<tr>
<td>14</td>
<td>32</td>
</tr>
<tr>
<td>15</td>
<td>35  36  37  38  39  40</td>
</tr>
<tr>
<td>16</td>
<td>41  42  43</td>
</tr>
<tr>
<td>17</td>
<td>44  45  46</td>
</tr>
<tr>
<td>18</td>
<td>47  48  49</td>
</tr>
<tr>
<td>19</td>
<td>50  51  52</td>
</tr>
<tr>
<td>20</td>
<td>53  54  55</td>
</tr>
<tr>
<td>21</td>
<td>56</td>
</tr>
<tr>
<td>22</td>
<td>57  58  59  60  61  62  63</td>
</tr>
<tr>
<td>23</td>
<td>64  65  66  67  68  69  70</td>
</tr>
<tr>
<td>24</td>
<td>71  72  73  74  75  76</td>
</tr>
<tr>
<td>25</td>
<td>77  78</td>
</tr>
<tr>
<td>26</td>
<td>79  80  81</td>
</tr>
<tr>
<td>27</td>
<td>82  84  85  86  87</td>
</tr>
<tr>
<td>28</td>
<td>83</td>
</tr>
<tr>
<td>29</td>
<td>88  89  90  91  92</td>
</tr>
<tr>
<td>30</td>
<td>93</td>
</tr>
<tr>
<td>31</td>
<td>94</td>
</tr>
<tr>
<td>32</td>
<td>95  96</td>
</tr>
<tr>
<td>33</td>
<td>97</td>
</tr>
<tr>
<td>34</td>
<td>98</td>
</tr>
<tr>
<td>35</td>
<td>99</td>
</tr>
<tr>
<td>36</td>
<td>100</td>
</tr>
<tr>
<td>37</td>
<td>101</td>
</tr>
<tr>
<td>38</td>
<td>102</td>
</tr>
<tr>
<td>39</td>
<td>103</td>
</tr>
<tr>
<td>40</td>
<td>104</td>
</tr>
<tr>
<td>41</td>
<td>105</td>
</tr>
<tr>
<td>42</td>
<td>106</td>
</tr>
<tr>
<td>43</td>
<td>107</td>
</tr>
<tr>
<td>44</td>
<td>108  109</td>
</tr>
</tbody>
</table>
Table 4.1.4

Model 2: Industry Classification Produced by Aggregating According to Similarity of Partially Aggregated Input Coefficients

<table>
<thead>
<tr>
<th>Sector</th>
<th>Industries Comprising the Sector</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1 21 22 23 24</td>
</tr>
<tr>
<td>2</td>
<td>2 27 30 34 42 53 54 56 67 104</td>
</tr>
<tr>
<td>3</td>
<td>3 4 44 48 71 72 76 78 79 81</td>
</tr>
<tr>
<td></td>
<td>87 88 95 98 100 102</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>12</td>
<td>13 15 18 52 59 61 90 91 92</td>
</tr>
<tr>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>14</td>
<td>16</td>
</tr>
<tr>
<td>15</td>
<td>17 28 33 49 75 85 89 99 106</td>
</tr>
<tr>
<td>16</td>
<td>19 25 26 29 31</td>
</tr>
<tr>
<td>17</td>
<td>20 62 65 83 86 105 107 108 109</td>
</tr>
<tr>
<td>18</td>
<td>32</td>
</tr>
<tr>
<td>19</td>
<td>35</td>
</tr>
<tr>
<td>20</td>
<td>36</td>
</tr>
<tr>
<td>21</td>
<td>37</td>
</tr>
<tr>
<td>22</td>
<td>38</td>
</tr>
<tr>
<td>23</td>
<td>39 73 77 101</td>
</tr>
<tr>
<td>24</td>
<td>40</td>
</tr>
<tr>
<td>25</td>
<td>41</td>
</tr>
<tr>
<td>26</td>
<td>43</td>
</tr>
<tr>
<td>27</td>
<td>45 46 47</td>
</tr>
<tr>
<td>28</td>
<td>50</td>
</tr>
<tr>
<td>29</td>
<td>51</td>
</tr>
<tr>
<td>30</td>
<td>55</td>
</tr>
<tr>
<td>31</td>
<td>57</td>
</tr>
<tr>
<td>32</td>
<td>58</td>
</tr>
<tr>
<td>33</td>
<td>60</td>
</tr>
<tr>
<td>34</td>
<td>63</td>
</tr>
<tr>
<td>35</td>
<td>64 66 70 82 84</td>
</tr>
<tr>
<td>36</td>
<td>68</td>
</tr>
<tr>
<td>37</td>
<td>69</td>
</tr>
<tr>
<td>38</td>
<td>74</td>
</tr>
<tr>
<td>39</td>
<td>80</td>
</tr>
<tr>
<td>40</td>
<td>93</td>
</tr>
<tr>
<td>41</td>
<td>94</td>
</tr>
<tr>
<td>42</td>
<td>96</td>
</tr>
<tr>
<td>43</td>
<td>97</td>
</tr>
<tr>
<td>44</td>
<td>103</td>
</tr>
</tbody>
</table>
In fact, these are the only sectors, for which there is a complete one-to-one correspondence between the two models. Nevertheless, other groupings of industries are common to sectors of both models, namely:

13. Chocolate and Confectionery,
and 15. Food Preparations N.E.I.

22. Other Spinning and Weaving,
23. Hosiery and Other Knitting,
and 24. Textiles N.E.I.

25. Men's Outerwear,
26. Women's Outerwear,
29. Corsetry,
and 31. Apparel N.E.I.

27. Underclothing,
30. Shirts and Pyjamas,
and 34. Made-up Textiles.

28. Millinery and Hats,
and 33. Canvas Goods.

45. Cartons and Paper Bags,
and 46. Paper Products N.E.I.

53. Tyres and Tubes,
and 54. Other Rubber Goods.

59. Paint and Varnish,
and 61. Chemical Products N.E.I.
64. Structural Clay Products,
66. Cement,
and 70. Mineral Products N.E.I.

71. Basic Metal Industries,
72. Sheetmetal Working,
and 76. Metal Products N.E.I.

79. Range Making,
and 81. Electrical Goods N.E.I.

82. Boat Building and Repairs,
and 84. Body Building.

90. Toys and Sports Goods,
91. Manufacturing N.E.I.,
and 92. Plastics Manufacturing.

108. Services to Households etc.,
and 109. Services to Government.

An examination of the differences between the two models reveals some unexpected industry aggregations in the model generated by the clustering algorithm. However, it is difficult to interpret why certain combinations of industries have been chosen to form sectors, because the merger of two existing sectors affects, not only the error sum of squares for the resulting sector, as opposed to those of the original two sectors, but also the error sums of squares of all other existing sectors in the model. It would be misleading,
therefore, to attempt to reconcile the sectors produced by the clustering methodology on the basis of expected similarity of the input structures of their industries. Consequently, the following discussion will be limited to pointing out those aggregations, which might be unacceptable to the input-output analyst, in that they contain industries, which have no readily identifiable relationship to one another.

(a) The aggregation of industry 1 (Farming) with the following industries, to form sector 1, is displayed in Figure 4.1.1.

21. Wool Milling
22. Other Spinning and Weaving
23. Hosiery and Other Knitting
24. Textiles N.E.I.

Figure 4.1.1
The Formation of Sector 1 in Model 2

<table>
<thead>
<tr>
<th>Industry</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>21</td>
</tr>
<tr>
<td>22</td>
</tr>
<tr>
<td>24</td>
</tr>
<tr>
<td>23</td>
</tr>
</tbody>
</table>

No. of Sectors in Model when Industries are Merged

|         | 95 | 94 | 62 | 56 |

Note that Farming is aggregated with Wool Milling at an early stage of the aggregation procedure, when there is a high degree of discrimination between sectors.
(b) The aggregation of the following industries, to form sector 2, is displayed in Figure 4.1.2.

2. Hunting and Fishing
27. Underclothing
30. Shirts and Pyjamas
34. Made-up Textiles
42. Mattresses
53. Tyres and Tubes
54. Other Rubber Goods
56. Chemical Fertilizers
67. Glass Products
104. Road Transport

Figure 4.1.2
The Formation of Sector 2 in Model 2

<table>
<thead>
<tr>
<th>Industry</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
</tr>
<tr>
<td>27</td>
</tr>
<tr>
<td>34</td>
</tr>
<tr>
<td>30</td>
</tr>
<tr>
<td>42</td>
</tr>
<tr>
<td>53</td>
</tr>
<tr>
<td>56</td>
</tr>
<tr>
<td>54</td>
</tr>
<tr>
<td>67</td>
</tr>
<tr>
<td>104</td>
</tr>
</tbody>
</table>

No. of Sectors in Model when Industries are Merged
103  99  96  86  82  74  72  48  47
The most unexpected mergers occur in the formation of the 48 sector model, when Hunting and Fishing is aggregated with Underclothing, Shirts and Pyjamas, Made-up Textiles and Mattresses, and in the formation of the 47 sector model, when this same group of industries is merged with Tyres and Tubes, Other Rubber Goods, Chemical Fertilizers, Glass Products and Road Transport. Note, however, that both of these mergers occur late in the aggregation procedure, when the degree of discrimination between sectors is very low.

(c) The aggregation of the following industries, to form sector 3, is displayed in Figure 4.1.3.

3. Forestry
4. Mining and Quarrying
44. Pulp and Paper
48. Job and General Printing
71. Basic Metal Industries
72. Sheetmetal Working
76. Metal Products N.E.I.
78. Machinery N.E.I.
79. Range Making
81. Electrical Goods N.E.I.
87. Transport Equipment N.E.I.
88. Jewellery
95. Civil Engineering
98. Water and Sanitation
100. Banking and Insurance
102. Shipping Transport
Again, the most unexpected aggregations take place late in the clustering procedure, when the degree of discrimination between sectors is low, namely, when models of 59 sectors or less are formed.
(d) The aggregation of the following industries, to form sector 12, is displayed in Figure 4.1.4.

13. Chocolate and Confectionery
15. Food Preparations N.E.I.
18. Aerated Waters and Cordials
52. Leather Goods
59. Paint and Varnish
61. Chemical Products N.E.I.
90. Toys and Sports Goods
91. Manufacturing N.E.I.
92. Plastics Manufacturing

Figure 4.1.4
The Formation of Sector 12 in Model 2

Industry

13
18
15
59
61
92
90
91
52

No. of Sectors in Model when Industries are Merged

93 84 79 77 71 63 50 46

Most surprising, perhaps, is the merger of Chemical Products N.E.I. and Plastics Manufacturing at such an early stage in the aggregation procedure. Other unusual groupings take place when the degree of discrimination between sectors is small, namely, in the 63, 50 and 46 sector models.
(e) The aggregation of the following industries, to form sector 15, is displayed in Figure 4.1.5.

17. Malting and Brewing
28. Millinery and Hats
33. Canvas Goods
49. Printing and Trade Services
75. Electro-Plating
85. Vehicle Repair
89. Brushes and Brooms
99. Trade
106. Services

Figure 4.1.5

The Formation of Sector 15 in Model 2

A number of peculiar groupings of industries are found to occur relatively early in the aggregation process. In fact, only the merger of Millinery and Hats and Canvas Goods could have been foreseen.
The aggregation of industry 19 (Tobacco and Cigarettes) with the following industries, to form sector 16, is displayed in Table 4.1.6.

25. Men's Outerwear
26. Women's Outerwear
29. Corsetry
31. Apparel N.E.I.

It is surprising to find Tobacco and Cigarettes aggregated with Apparel N.E.I. at such an early stage of the clustering process.
(g) The aggregation of the following industries, to form sector 17, is displayed in Figure 4.1.7.

- 20. Wool Scouring
- 62. Petroleum and Coal Products
- 65. Pottery Clay Products
- 83. Vehicle Assembly
- 86. Aircraft Repair
- 105. Communications
- 107. Ownership of Property
- 108. Services to Households etc.
- 109. Services to Government

**Figure 4.1.7**

The Formation of Sector 17 in Model 2

The merger of Petroleum and Coal Products with Aircraft Repair occurs very early in the aggregation process. The merger of Wool Scouring with Pottery Clay Products also occurs surprisingly early, as does the grouping of industries 62, 86, 83, 108, 109 and 105.
The aggregation of the following industries, to form sector 23, is displayed in Figure 4.1.8.

39. Plywood and Veneer
73. Wire Working
77. Farm Machinery
101. Rail Transport

![Diagram showing the formation of Sector 23 in Model 2]

Note that the most unexpected merger, namely, that of Plywood and Veneer with the other three industries, occurs very late in the aggregation process, when the degree of discrimination between the sectors is low.
(i) The aggregation of the following industries, to form sector 35, is displayed in Figure 4.1.9.

64. Structural Clay Products
66. Cement
70. Mineral Products N.E.I.
82. Boat Building and Repairs
84. Body Building

Figure 4.1.9
The Formation of Sector 35 in Model 2

<table>
<thead>
<tr>
<th>Industry</th>
<th>No. of Sectors in Model when Industries are Merged</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>90</td>
</tr>
<tr>
<td>66</td>
<td>70</td>
</tr>
<tr>
<td>70</td>
<td>51</td>
</tr>
<tr>
<td>82</td>
<td>49</td>
</tr>
<tr>
<td>84</td>
<td></td>
</tr>
</tbody>
</table>

The only totally unexpected merger takes place in the formation of the 49 sector model, when industries 64, 66 and 70 are aggregated with industries 82 and 84. At this stage of the clustering procedure, the degree of discrimination between sectors is low.
4.2 The Results of Constrained Aggregation, Based on Similarity of Partially Aggregated Input Coefficients

Since some of the sectors, produced by the clustering algorithm, might be regarded as unsatisfactory according to a priori notions, it was decided to impose the following constraints upon the intersectoral model.

(a) Industries 1, 2, 3 and 4 must remain isolated.

(b) Aggregation may take place within the following groups of industries, but not between groups:

- industries 5 to 19 inclusive,
- industries 20 to 24 inclusive,
- industries 25 to 34 inclusive,
- industries 35 to 43 inclusive,
- industries 44 to 49 inclusive,
- industries 50 to 52 inclusive,
- industries 53 to 55 inclusive,
- industries 56 to 63 inclusive,
- industries 64 to 70 inclusive,
- industries 71 to 76 inclusive,
- industries 77 to 87 inclusive,
- industries 88 to 92 inclusive,
- industries 93 to 96 inclusive,
- industries 97 to 100 inclusive,
- industries 101 to 109 inclusive.

Note that these constraints do not prohibit the formation of sectors, consistent with the Department of Statistics'
classification. In fact, they are likely to lead to an intersectoral model, which resembles the Department's model more closely than that produced when no constraints were imposed upon the clustering process.

The final grouping of the 109 industries into 44 sectors, produced by the clustering algorithm under the constraints listed above, is referred to hereafter as Model 3 and is given in Table 4.2.1. The industry groupings, being consistent with the constraints imposed upon the model, are satisfactory on a priori grounds. The sum of squared deviations of partially aggregated input coefficients from their means is 0.40350, compared with 0.15058 for the model produced by the algorithm when no constraints were imposed, and 1.46484 for the Department of Statistics' model. Hence, even when stringent restrictions were placed upon the grouping of industries, the clustering algorithm was able to produce a 44 sector model, for which the similarity between partially aggregated input coefficients, within sectors, is considerably greater than in the model given by the Department of Statistics.

The composition of the sectors, appearing in the three models, will now be compared. The reader is reminded that the Department of Statistics' model is Model 1, the model produced by the unconstrained clustering algorithm is Model 2, and the model produced by the clustering algorithm, with constraints imposed, is Model 3.
Table 4.2.1
Model 3: The Constrained 44 Sector Model Produced by Clustering According to Similarity of Partially Aggregated Input Coefficients

<table>
<thead>
<tr>
<th>Sector</th>
<th>Industries Comprising the Sector</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>5 11</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>10 14 16</td>
</tr>
<tr>
<td>11</td>
<td>12 13 15 17 18 19</td>
</tr>
<tr>
<td>12</td>
<td>20</td>
</tr>
<tr>
<td>13</td>
<td>21 22 24</td>
</tr>
<tr>
<td>14</td>
<td>23</td>
</tr>
<tr>
<td>15</td>
<td>25 26 27 28 29 30 31 33 34</td>
</tr>
<tr>
<td>16</td>
<td>32</td>
</tr>
<tr>
<td>17</td>
<td>35</td>
</tr>
<tr>
<td>18</td>
<td>36 38</td>
</tr>
<tr>
<td>19</td>
<td>37</td>
</tr>
<tr>
<td>20</td>
<td>39 41 42</td>
</tr>
<tr>
<td>21</td>
<td>40</td>
</tr>
<tr>
<td>22</td>
<td>43</td>
</tr>
<tr>
<td>23</td>
<td>44 49</td>
</tr>
<tr>
<td>24</td>
<td>45 46 47 48</td>
</tr>
<tr>
<td>25</td>
<td>50 51</td>
</tr>
<tr>
<td>26</td>
<td>52</td>
</tr>
<tr>
<td>27</td>
<td>53 54</td>
</tr>
<tr>
<td>28</td>
<td>55</td>
</tr>
<tr>
<td>29</td>
<td>56 59 60 61 62</td>
</tr>
<tr>
<td>30</td>
<td>57 58</td>
</tr>
<tr>
<td>31</td>
<td>63</td>
</tr>
<tr>
<td>32</td>
<td>64 65 66 67 69 70</td>
</tr>
<tr>
<td>33</td>
<td>68</td>
</tr>
<tr>
<td>34</td>
<td>71 72 73 75 76</td>
</tr>
<tr>
<td>35</td>
<td>74</td>
</tr>
<tr>
<td>36</td>
<td>77 79 80 87</td>
</tr>
<tr>
<td>37</td>
<td>78 81 82 83 84 85 86</td>
</tr>
<tr>
<td>38</td>
<td>88 89 90 91 92</td>
</tr>
<tr>
<td>39</td>
<td>93 94</td>
</tr>
<tr>
<td>40</td>
<td>95 96</td>
</tr>
<tr>
<td>41</td>
<td>97</td>
</tr>
<tr>
<td>42</td>
<td>98 99 100</td>
</tr>
<tr>
<td>43</td>
<td>101 105 107 108 109</td>
</tr>
<tr>
<td>44</td>
<td>102 103 104 106</td>
</tr>
</tbody>
</table>
Table 4.2.2 lists those sectors, for which there is a
one-to-one correspondence in the three possible pairs of models.

Table 4.2.2

Sectors with a One-To-One Correspondence

<table>
<thead>
<tr>
<th>Model 1 &amp; Model 3</th>
<th>Model 1 &amp; Model 2</th>
<th>Model 2 &amp; Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>32</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>93</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>94</td>
<td>9</td>
</tr>
<tr>
<td>25, 26, 27, 28, 29</td>
<td>97</td>
<td>32</td>
</tr>
<tr>
<td>30, 31, 33, 34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>103</td>
<td>35</td>
</tr>
<tr>
<td>88, 89, 90, 91, 92</td>
<td></td>
<td>37</td>
</tr>
<tr>
<td>95, 96</td>
<td></td>
<td>40</td>
</tr>
<tr>
<td>97</td>
<td></td>
<td>43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>55</td>
</tr>
<tr>
<td></td>
<td></td>
<td>63</td>
</tr>
<tr>
<td></td>
<td></td>
<td>68</td>
</tr>
<tr>
<td></td>
<td></td>
<td>74</td>
</tr>
<tr>
<td></td>
<td></td>
<td>97</td>
</tr>
</tbody>
</table>

The most striking feature of Table 4.2.2 is the number of single
industry sectors produced by the clustering algorithm, regardless
of whether or not constraints are imposed. However, the only
sectors, which appear in all three models, are the two single
industry sectors, Footwear - not Rubber (Industry 32) and
Electricity and Gas (Industry 97). The constrained clustering
algorithm produced nine sectors with a one-to-one correspondence
in the Department's model, but since industries 1, 2, 3 and 4
were prohibited from being aggregated with other industries,
only five of the nine were formed by the algorithm. This is a
surprisingly small number, when one considers the nature of the
constraints imposed upon Model 3.
However, a large number of common industry groupings appear in the three models. These are listed in Table 4.2.3.

Table 4.2.3
Common Industry Groupings

<table>
<thead>
<tr>
<th>Model 1 &amp; Model 3</th>
<th>Model 1 &amp; Model 2</th>
<th>Model 2 &amp; Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13, 15</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>22, 23, 24</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>25, 26, 29, 31</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>27, 30, 34</td>
<td>9</td>
</tr>
<tr>
<td>12, 13, 15</td>
<td>28, 33</td>
<td>13, 15, 18</td>
</tr>
<tr>
<td>17, 18</td>
<td>45, 46</td>
<td>21, 22, 24</td>
</tr>
<tr>
<td>22, 24</td>
<td>53, 54</td>
<td>25, 26, 29, 31</td>
</tr>
<tr>
<td>25, 26, 27, 28, 29</td>
<td>59, 61</td>
<td>27, 30, 34</td>
</tr>
<tr>
<td>30, 31, 33, 34</td>
<td>64, 66, 70</td>
<td>28, 33</td>
</tr>
<tr>
<td>36, 38</td>
<td>71, 72, 76</td>
<td>32</td>
</tr>
<tr>
<td>41, 42</td>
<td>79, 81</td>
<td>35</td>
</tr>
<tr>
<td>45, 46</td>
<td>82, 84</td>
<td>37</td>
</tr>
<tr>
<td>47, 48</td>
<td>90, 91, 92</td>
<td>40</td>
</tr>
<tr>
<td>50, 51</td>
<td>108, 109</td>
<td>43</td>
</tr>
<tr>
<td>53, 54</td>
<td>45, 46, 47</td>
<td></td>
</tr>
<tr>
<td>57, 58</td>
<td>53, 54</td>
<td>55</td>
</tr>
<tr>
<td>59, 60, 61, 62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>64, 65, 66, 67</td>
<td>59, 61</td>
<td></td>
</tr>
<tr>
<td>69, 70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>71, 72, 73, 75, 76</td>
<td></td>
<td>63</td>
</tr>
<tr>
<td>79, 80</td>
<td>64, 66, 70</td>
<td></td>
</tr>
<tr>
<td>82, 84, 85, 86</td>
<td></td>
<td>68</td>
</tr>
<tr>
<td>88, 89, 90, 91, 92</td>
<td>71, 72, 76</td>
<td></td>
</tr>
<tr>
<td>95, 96</td>
<td>74</td>
<td></td>
</tr>
<tr>
<td>108, 109</td>
<td>78, 81</td>
<td></td>
</tr>
<tr>
<td></td>
<td>79, 87</td>
<td></td>
</tr>
<tr>
<td></td>
<td>82, 84</td>
<td></td>
</tr>
<tr>
<td></td>
<td>83, 86</td>
<td></td>
</tr>
<tr>
<td></td>
<td>90, 91, 92</td>
<td></td>
</tr>
<tr>
<td></td>
<td>97</td>
<td></td>
</tr>
<tr>
<td></td>
<td>98, 100</td>
<td></td>
</tr>
<tr>
<td></td>
<td>105, 107, 108, 109</td>
<td></td>
</tr>
</tbody>
</table>
As expected, there is a considerable overlap between Model 3 and Model 1, and between Model 3 and Model 2. Industry groupings, which are common to all three models, are:

13. Chocolate and Confectionery,
and 15. Food Preparations N.E.I.

22. Other Spinning and Weaving,
and 24. Textiles N.E.I.

25. Men's Outerwear,
26. Women's Outerwear,
29. Corsetry,
and 31. Apparel N.E.I.

27. Underclothing,
30. Shirts and Pyjamas,
and 34. Made-up Textiles.

28. Millinery and Hats,
and 33. Canvas Goods.

45. Cartons and Paper Bags,
and 46. Paper Products N.E.I.

53. Tyres and Tubes,
and 54. Other Rubber Goods.

59. Paint and Varnish,
and 61. Chemical Products N.E.I.
64. Structural Clay Products, 
66. Cement, 
and 70. Mineral Products N.E.I. 

71. Basic Metal Industries, 
72. Sheetmetal Working, 
and 76. Metal Products N.E.I. 

82. Boat Building and Repairs, 
and 84. Body Building. 

90. Toys and Sports Goods, 
91. Manufacturing N.E.I., 
and 92. Plastics Manufacturing. 

108. Services to Households etc., 
and 109. Services to Government. 

The main differences between the model produced under constrained aggregation (Model 3) and the Department of Statistics' model (Model 1) are as follows: 

(a) In Model 3, industry 5 (Meat Freezing and Preserving) is aggregated with industry 11 (Grain Milling), rather than with industry 6 (Ham, Bacon and Smallgoods), as occurs in Model 1. This probably occurs because both Meat Freezing and Preserving and Grain Milling receive a high proportion of their inputs from the single-industry sector, Farming, of both Models 1 and 3.
(b) The following industries are aggregated to form a sector in Model 3:

10. Fruit and Vegetable Preserving,
14. Animal Feed,

In Model 1, industry 10 remains isolated, while industries 14 and 16 are placed in different sectors.

(c) The aggregation, in Model 3, of industries:

12. Biscuits and Bread Baking,
13. Chocolate and Confectionery,
and 15. Food Preparations N.E.I.,

with industries:

17. Malting and Brewing,
18. Aerated Waters and Cordials,
and 19. Tobacco and Cigarettes,

to form a sector, differs from the groupings found in Model 1. In the latter, industry 19 remains isolated, industries 12, 13 and 15 are aggregated with industries:

7. Ice Cream,
11. Grain Milling,
and 14. Animal Feed,

and industries 17 and 18 are grouped with industry 16 (Wine Making) to form a sector.

(d) Industry 8 (Butter, Cheese and Milk-Powder etc.) and industry 9 (Milk Treatment) remain isolated in Model 3, but are combined to form a sector in Model 1.
(e) There is a difference in the grouping of the following industries in Model 1 and Model 3:

20. Wool Scouring,
21. Wool Milling,
22. Other Spinning and Weaving,
23. Hosiery and Other Knitting,
and 24. Textiles N.E.I.

In Model 1, industries 20 and 21 form a sector, and industries 22, 23 and 24 form another sector. In Model 3, industries 20 and 23 remain isolated, while industries 21, 22 and 24 are aggregated to form a sector.

(f) The following industries are isolated in Model 3:

35. Sawmills,
37. Joinery,
40. Wood Products N.E.I.,
and 43. Venetian Blinds.

Also in Model 3, industries:

36. Planing Mills,
and 38. Wooden Containers

combine to form a sector, and industries:

39. Plywood and Veneer,
41. Furniture,
and 42. Mattresses

comprise another sector. Model 1 contains a sector consisting of industries 35 to 40 inclusive, and another sector containing industries 41, 42 and 43.
(g) Industries 44 to 49 inclusive are grouped differently in the two models. Model 3 contains a sector, comprised of industries:

   44. Pulp and Paper,
   and 49. Printing and Trade Services,

and another sector, containing industries:

   45. Cartons and Paper Bags,
   46. Paper Products N.E.I.,
   47. Printing and Publishing,
   and 48. Job and General Printing.

By comparison, Model 1 groups industries 44, 45 and 46 into one sector, and industries 47, 48 and 49 into another sector.

(h) Model 3 groups industries 50 to 55 inclusive into four sectors. Two of these are single-industry sectors, namely, industry 52 (Leather Goods) and industry 55 (Tyre Retreading). The third sector is made up of industries:

   50. Tanning,
   and 51. Fellmongery,

and the fourth sector contains industries:

   53. Tyres and Tubes,
   and 54. Other Rubber Goods.

Model 1 contains just two groupings, the first of which consists of industries 50, 51 and 52, and the second is comprised of industries 53, 54 and 55.
(i) In Model 1, industry 56 (Chemical Fertilizers) remains isolated, while the following industries are aggregated to form a sector:

57. Vegetable and Animal Oils,
58. Soap and Candle,
59. Paint and Varnish,
60. Medical and Toilet Goods,
61. Chemical Products N.E.I.,
62. Petroleum and Coal Products,
and 63. Bituminous Materials.

Industry 63 remains isolated in Model 3, industries 56, 59, 60, 61 and 62 form a sector, and industries 57 and 58 form another sector.

(j) Both models group the following industries into the same sector:

64. Structural Clay Products,
65. Pottery Clay Products,
66. Cement,
67. Glass Products,
69. Lime,
and 70. Mineral Products N.E.I.

However, industry 68 (Concrete Products) is included in the same sector in Model 1, but remains isolated in Model 3.
(k) Similarly, both models contain a sector consisting of industries:

71. Basic Metal Industries,
72. Sheetmetal Working,
73. Wire Working,
75. Electro-Plating,
and 76. Metal Products N.E.I.,

but in Model 1, industry 74 (Nail Making) is part of the same sector, whereas it forms a single-industry sector in Model 3.

(l) A substantial regrouping of industries 77 to 87 inclusive takes place in Model 3, compared with Model 1. The former contains two sectors, the first consisting of industries:

77. Farm Machinery,
79. Range Making,
80. Radio and TV Assembly,
and 87. Transport Equipment N.E.I.

The second sector contains industries:

78. Machinery N.E.I.,
81. Electrical Goods N.E.I.,
82. Boat Building and Repairs,
83. Vehicle Assembly,
84. Body Building,
85. Vehicle Repair,
and 86. Aircraft Repair.

Model 1 arranges these industries into four sectors. Industry 83 forms a single-industry sector, industries 77 and 78 are grouped to form a second sector, industries 79, 80 and 81 form a third sector, and industries 82, 84, 85, 86 and 87 make up the fourth sector.
(m) Industry 93 (Residential Building) and Industry 94 (Commercial Building) are aggregated, in Model 3, to form a sector, whereas they remain isolated in Model 1.

(n) Industries 98 to 109 inclusive are grouped into the following sectors in Model 3. The first sector consists of industries:

98. Water and Sanitation,
99. Trade,
and 100. Banking and Insurance.

The second sector consists of industries:

101. Rail Transport,
105. Communications,
107. Ownership of Property,
108. Services to Households etc.,
and 109. Services to Government.

The third sector is comprised of industries:

102. Shipping Transport,
103. Air Transport,
104. Road Transport,
and 106. Services.

In Model 1, however, industries 108 and 109 are aggregated to form a sector, but industries 98 to 107 inclusive each remain isolated.
4.3 Results of Aggregation Based on Similarity of Input Coefficients

It is interesting to compare both the Department of Statistics' model and the model produced by clustering according to similarity of partially aggregated input coefficients with the 44 sector model, produced using Ward's clustering algorithm. The latter is one of the methods employed by Blin and Cohen and it has been discussed already in Section 2.6.3. The criterion used to aggregate industries into sectors is similarity of input coefficients within sectors, as measured by the total, within sector, sum of squared deviations of input coefficients from their means. In particular, it is interesting to observe whether or not the sectors produced by Ward's algorithm are more intuitively appealing than those produced using the algorithm developed in Chapter 3, with no constraints imposed.

Table 4.3.1 shows the successive mergers of sectors, according to similarity of input coefficients, down to the formation of the 44 sector model. The total, within sector, sum of squared deviations of input coefficients from their means is called the "total error sum of squares" in Table 4.3.1. The information, extracted from Table 4.3.1 to form Table 4.3.2, reveals a high degree of discrimination between sectors in that the proportion of sectors formed within various ranges of similarity, of the total of 65 mergers, decreases as the
<table>
<thead>
<tr>
<th>Stage of Sectors</th>
<th>Aggregation of Industries</th>
<th>Total Error Sum of Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>108 109</td>
<td>0.00000</td>
</tr>
<tr>
<td>2</td>
<td>107 72 76</td>
<td>0.00045</td>
</tr>
<tr>
<td>3</td>
<td>106 62 86</td>
<td>0.00092</td>
</tr>
<tr>
<td>4</td>
<td>105 53 56</td>
<td>0.00155</td>
</tr>
<tr>
<td>5</td>
<td>104 25 26</td>
<td>0.00226</td>
</tr>
<tr>
<td>6</td>
<td>103 62 86 83</td>
<td>0.00297</td>
</tr>
<tr>
<td>7</td>
<td>102 49 75</td>
<td>0.00397</td>
</tr>
<tr>
<td>8</td>
<td>101 27 34</td>
<td>0.00500</td>
</tr>
<tr>
<td>9</td>
<td>100 72 76 78</td>
<td>0.00621</td>
</tr>
<tr>
<td>10</td>
<td>99 62 86 83 105</td>
<td>0.00747</td>
</tr>
<tr>
<td>11</td>
<td>98 53 56 67</td>
<td>0.00874</td>
</tr>
<tr>
<td>12</td>
<td>97 99 106</td>
<td>0.01005</td>
</tr>
<tr>
<td>13</td>
<td>96 45 46</td>
<td>0.01139</td>
</tr>
<tr>
<td>14</td>
<td>95 28 33</td>
<td>0.01273</td>
</tr>
<tr>
<td>15</td>
<td>94 53 56 67 54</td>
<td>0.01408</td>
</tr>
<tr>
<td>16</td>
<td>93 30 91</td>
<td>0.01546</td>
</tr>
<tr>
<td>17</td>
<td>92 85 104</td>
<td>0.01683</td>
</tr>
<tr>
<td>18</td>
<td>91 62 86 83 105 108 109</td>
<td>0.01822</td>
</tr>
<tr>
<td>19</td>
<td>90 22 24</td>
<td>0.01961</td>
</tr>
<tr>
<td>20</td>
<td>89 25 26 31</td>
<td>0.02128</td>
</tr>
<tr>
<td>21</td>
<td>88 59 61</td>
<td>0.02297</td>
</tr>
<tr>
<td>22</td>
<td>87 20 65</td>
<td>0.02468</td>
</tr>
<tr>
<td>23</td>
<td>86 71 72 76 78</td>
<td>0.02640</td>
</tr>
<tr>
<td>24</td>
<td>85 90 92</td>
<td>0.02839</td>
</tr>
<tr>
<td>25</td>
<td>84 71 72 76 78 81</td>
<td>0.03042</td>
</tr>
<tr>
<td>26</td>
<td>83 30 91 42</td>
<td>0.03253</td>
</tr>
<tr>
<td>27</td>
<td>82 64 95</td>
<td>0.03466</td>
</tr>
<tr>
<td>28</td>
<td>81 20 65 49 75</td>
<td>0.03721</td>
</tr>
<tr>
<td>29</td>
<td>80 47 48</td>
<td>0.03983</td>
</tr>
<tr>
<td>30</td>
<td>79 27 24 53 56 67 54</td>
<td>0.04249</td>
</tr>
<tr>
<td>31</td>
<td>78 30 91 42 89</td>
<td>0.04527</td>
</tr>
<tr>
<td>32</td>
<td>77 28 33 90 92</td>
<td>0.04814</td>
</tr>
<tr>
<td>33</td>
<td>76 13 18</td>
<td>0.05118</td>
</tr>
<tr>
<td>34</td>
<td>75 73 88</td>
<td>0.05433</td>
</tr>
<tr>
<td>35</td>
<td>74 17 19</td>
<td>0.05766</td>
</tr>
<tr>
<td>36</td>
<td>73 82 84</td>
<td>0.06112</td>
</tr>
<tr>
<td>37</td>
<td>72 59 61 60</td>
<td>0.06461</td>
</tr>
<tr>
<td>38</td>
<td>71 28 33 90 92 30 91 42 89</td>
<td>0.06825</td>
</tr>
<tr>
<td>39</td>
<td>70 62 86 83 105 108 109 107</td>
<td>0.07205</td>
</tr>
<tr>
<td>40</td>
<td>69 1 21</td>
<td>0.07627</td>
</tr>
<tr>
<td>41</td>
<td>68 20 65 49 75 99 106</td>
<td>0.08059</td>
</tr>
<tr>
<td>42</td>
<td>67 64 95 66</td>
<td>0.08500</td>
</tr>
<tr>
<td>43</td>
<td>66 3 44</td>
<td>0.08948</td>
</tr>
<tr>
<td>44</td>
<td>65 79 87</td>
<td>0.09398</td>
</tr>
<tr>
<td>45</td>
<td>64 13 18 15</td>
<td>0.09870</td>
</tr>
<tr>
<td>46</td>
<td>63 79 87 82 84</td>
<td>0.10360</td>
</tr>
<tr>
<td>47</td>
<td>62 27 34 53 56 67 54 28 33 90 92 30</td>
<td>0.10882</td>
</tr>
<tr>
<td>48</td>
<td>61 91 42 89</td>
<td>0.11456</td>
</tr>
<tr>
<td>49</td>
<td>60 20 65 49 75 99 106 62 86 83 105 108 109 107</td>
<td>0.12069</td>
</tr>
<tr>
<td>50</td>
<td>59 64 95 66 70</td>
<td>0.12697</td>
</tr>
<tr>
<td>51</td>
<td>58 27 34 53 56 67 54 28 33 90 92 30 91 42 89 52</td>
<td>0.13328</td>
</tr>
<tr>
<td>52</td>
<td>57 71 72 76 78 81 79 87 82 84</td>
<td>0.13984</td>
</tr>
<tr>
<td>53</td>
<td>56 93 94</td>
<td>0.14670</td>
</tr>
<tr>
<td>54</td>
<td>55 20 65 49 75 99 106 62 86 83 105 108 109 107 85 104</td>
<td>0.15384</td>
</tr>
<tr>
<td>55</td>
<td>54 57 58</td>
<td>0.16116</td>
</tr>
<tr>
<td>56</td>
<td>53 3 44 39</td>
<td>0.16648</td>
</tr>
<tr>
<td>57</td>
<td>52 73 88 74</td>
<td>0.17599</td>
</tr>
<tr>
<td>58</td>
<td>51 4 69</td>
<td>0.18353</td>
</tr>
<tr>
<td>59</td>
<td>50 1 21 10</td>
<td>0.19123</td>
</tr>
<tr>
<td>60</td>
<td>49 36 38</td>
<td>0.19903</td>
</tr>
<tr>
<td>61</td>
<td>48 2 17 19 20 65 49 75 59 106 62 86 83 105 108 109 107 85 104</td>
<td>0.20715</td>
</tr>
<tr>
<td>62</td>
<td>47 23 25 26 31</td>
<td>0.21538</td>
</tr>
<tr>
<td>63</td>
<td>46 27 34 53 56 67 54 28 33 90 92 30 91 42 89 52 29</td>
<td>0.22108</td>
</tr>
<tr>
<td>64</td>
<td>45 1 21 10 14</td>
<td>0.23462</td>
</tr>
<tr>
<td>65</td>
<td>44 27 34 53 56 67 54 28 33 90 92 30 91 42 89 52 29 41</td>
<td>0.24578</td>
</tr>
</tbody>
</table>
similarity measure increases, particularly as the 109 industries are merged into 74 sectors.

### Table 4.3.2

**Number and Percentage of Sectors Formed Within Various Ranges of the Similarity Scale**

<table>
<thead>
<tr>
<th>Similarity Measure</th>
<th>Number of Sectors Formed</th>
<th>Percentage of Sectors Formed</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00-&lt;0.02</td>
<td>19</td>
<td>29.2</td>
</tr>
<tr>
<td>0.02-&lt;0.04</td>
<td>10</td>
<td>15.4</td>
</tr>
<tr>
<td>0.04-&lt;0.06</td>
<td>6</td>
<td>9.2</td>
</tr>
<tr>
<td>0.06-&lt;0.08</td>
<td>5</td>
<td>7.7</td>
</tr>
<tr>
<td>0.08-&lt;0.10</td>
<td>5</td>
<td>7.7</td>
</tr>
<tr>
<td>0.10-&lt;0.12</td>
<td>3</td>
<td>4.6</td>
</tr>
<tr>
<td>0.12-&lt;0.14</td>
<td>4</td>
<td>6.2</td>
</tr>
<tr>
<td>0.14-&lt;0.16</td>
<td>2</td>
<td>3.1</td>
</tr>
<tr>
<td>0.16-&lt;0.18</td>
<td>3</td>
<td>4.6</td>
</tr>
<tr>
<td>0.18-&lt;0.20</td>
<td>3</td>
<td>4.6</td>
</tr>
<tr>
<td>0.20-&lt;0.22</td>
<td>2</td>
<td>3.1</td>
</tr>
<tr>
<td>0.22-&lt;0.24</td>
<td>2</td>
<td>3.1</td>
</tr>
<tr>
<td>0.24-&lt;0.26</td>
<td>1</td>
<td>1.5</td>
</tr>
</tbody>
</table>

The final 44 sector model produced by Ward's algorithm is given in Table 4.3.3 and has a total, within sector, sum of squared deviations of input coefficients from their means equal to 0.24578. For brevity, this model is called Model 4 in the discussion which follows.
Table 4.3.3
Model 4: 44 Sector Model Produced by Ward's Algorithm

<table>
<thead>
<tr>
<th>Sector</th>
<th>Industries Comprising the Sector</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1 10 14 21</td>
</tr>
<tr>
<td>2</td>
<td>2 17 19 20 49 62 65 75 83 85 86</td>
</tr>
<tr>
<td></td>
<td>99 104 108 106 107 108 109</td>
</tr>
<tr>
<td>3</td>
<td>3 39 44</td>
</tr>
<tr>
<td>4</td>
<td>4 69</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>12</td>
<td>13 15 18</td>
</tr>
<tr>
<td>13</td>
<td>16</td>
</tr>
<tr>
<td>14</td>
<td>22 24</td>
</tr>
<tr>
<td>15</td>
<td>23 25 26 31</td>
</tr>
<tr>
<td>16</td>
<td>27 28 29 30 33 34 41 42 52 53 54</td>
</tr>
<tr>
<td></td>
<td>56 67 89 90 91 92</td>
</tr>
<tr>
<td>17</td>
<td>32</td>
</tr>
<tr>
<td>18</td>
<td>35</td>
</tr>
<tr>
<td>19</td>
<td>36 38</td>
</tr>
<tr>
<td>20</td>
<td>37</td>
</tr>
<tr>
<td>21</td>
<td>40</td>
</tr>
<tr>
<td>22</td>
<td>43</td>
</tr>
<tr>
<td>23</td>
<td>45 46</td>
</tr>
<tr>
<td>24</td>
<td>47 48</td>
</tr>
<tr>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td>26</td>
<td>51</td>
</tr>
<tr>
<td>27</td>
<td>55</td>
</tr>
<tr>
<td>28</td>
<td>57 58</td>
</tr>
<tr>
<td>29</td>
<td>59 60 61</td>
</tr>
<tr>
<td>30</td>
<td>63</td>
</tr>
<tr>
<td>31</td>
<td>64 66 70 95</td>
</tr>
<tr>
<td>32</td>
<td>68</td>
</tr>
<tr>
<td>33</td>
<td>71 72 76 78 79 81 82 84 87</td>
</tr>
<tr>
<td>34</td>
<td>73 74 88</td>
</tr>
<tr>
<td>35</td>
<td>77</td>
</tr>
<tr>
<td>36</td>
<td>80</td>
</tr>
<tr>
<td>37</td>
<td>93 94</td>
</tr>
<tr>
<td>38</td>
<td>96</td>
</tr>
<tr>
<td>39</td>
<td>97</td>
</tr>
<tr>
<td>40</td>
<td>98</td>
</tr>
<tr>
<td>41</td>
<td>100</td>
</tr>
<tr>
<td>42</td>
<td>101</td>
</tr>
<tr>
<td>43</td>
<td>102</td>
</tr>
<tr>
<td>44</td>
<td>103</td>
</tr>
</tbody>
</table>
Model 4, like Model 2, contains a number of sectors, which appear heterogeneous on a priori grounds; in particular, sectors 1, 2, 3, 4, 15, 16, 31, 33 and 34. The differences and similarities between the intersectoral models 1, 2 and 4 will now be examined. Table 4.3.4 shows the industry composition of sectors for which there is a one-to-one correspondence in the three possible pairs of models.

Table 4.3.4

<table>
<thead>
<tr>
<th>Sectors with a One-To-One Correspondence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1 &amp; Model 2</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>32</td>
</tr>
<tr>
<td>93</td>
</tr>
<tr>
<td>94</td>
</tr>
<tr>
<td>97</td>
</tr>
<tr>
<td>103</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
All of these sectors contain just one industry. Note that Models 2 and 4 have the greatest number of sectors in common. Only industries 32 (Footwear - not Rubber), 97 (Electricity and Gas) and 103 (Air Transport) are left unaggregated in all three models. However, certain groupings of industries are common to two or more of the models. These are given in Table 4.3.5.

### Table 4.3.5

**Common Industry Groupings**

<table>
<thead>
<tr>
<th>Model 1 &amp; Model 2</th>
<th>Model 1 &amp; Model 4</th>
<th>Model 2 &amp; Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>13 15</td>
<td>13 15</td>
<td>1 21</td>
</tr>
<tr>
<td>22 23 24</td>
<td>22 24</td>
<td>2 104</td>
</tr>
<tr>
<td>25 26 29 31</td>
<td>25 26 31</td>
<td>3 44</td>
</tr>
<tr>
<td>27 30 34</td>
<td>27 28 29 30</td>
<td>13 15 18</td>
</tr>
<tr>
<td>28 33</td>
<td>36 38</td>
<td>17 49 75 85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>99 106</td>
</tr>
<tr>
<td>45 46</td>
<td>41 42</td>
<td>20 62 65 83 86</td>
</tr>
<tr>
<td></td>
<td></td>
<td>105 107 108 109</td>
</tr>
<tr>
<td>53 54</td>
<td>45 46</td>
<td>22 24</td>
</tr>
<tr>
<td>59 61</td>
<td>47 48</td>
<td>25 26 31</td>
</tr>
<tr>
<td>64 66 70</td>
<td>53 54</td>
<td>27 30 34 42 53</td>
</tr>
<tr>
<td></td>
<td></td>
<td>54 56 67</td>
</tr>
<tr>
<td>71 72 76</td>
<td>57 58</td>
<td>28 33 89</td>
</tr>
<tr>
<td>79 81</td>
<td>59 60 61</td>
<td>45 46</td>
</tr>
<tr>
<td>82 84</td>
<td>64 66 70</td>
<td>59 61</td>
</tr>
<tr>
<td>90 91 92</td>
<td>71 72 76</td>
<td>64 66 70</td>
</tr>
<tr>
<td>108 109</td>
<td>73 74</td>
<td>71 72 76 78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>79 81</td>
</tr>
<tr>
<td></td>
<td>79 81</td>
<td>82 84</td>
</tr>
<tr>
<td></td>
<td>82 84 87</td>
<td>90 91 92</td>
</tr>
<tr>
<td></td>
<td>85 86</td>
<td></td>
</tr>
<tr>
<td></td>
<td>89 90 91 92</td>
<td></td>
</tr>
<tr>
<td></td>
<td>108 109</td>
<td></td>
</tr>
</tbody>
</table>
The following industry groupings are found in all three models:

13. Chocolate and Confectionery,
and 15. Food Preparations N.E.I.

22. Other Spinning and Weaving,
and 24. Textiles N.E.I.

25. Men's Outerwear,
26. Women's Outerwear,
and 31. Apparel N.E.I.

27. Underclothing,
30. Shirts and Pyjamas,
and 34. Made-up Textiles.

28. Millinery and Hats,
and 33. Canvas Goods.

45. Cartons and Paper Bags,
and 46. Paper Products N.E.I.

53. Tyres and Tubes,
and 54. Other Rubber Goods.

59. Paint and Varnish,
and 61. Chemical Products N.E.I.

64. Structural Clay Products,
66. Cement,
and 70. Mineral Products N.E.I.
71. Basic Metal Industries,
72. Sheetmetal Working,
and 76. Metal Products N.E.I.

79. Range Making,
and 81. Electrical Goods N.E.I.

82. Boat Building and Repairs,
and 84. Body Building.

90. Toys and Sports Goods,
91. Manufacturing N.E.I.,
and 92. Plastics Manufacturing.

108. Services to Households etc.,
and 109. Services to Government.

More interesting, perhaps, are those industry groupings, which are produced by both clustering algorithms and yet do not appear in the Department of Statistics' model. Since Ward's algorithm aggregates on the basis of similarity of input coefficients, these industry groupings can now be considered in relation to their input structures.
(a) The aggregation of the two industries:

1. Farming
21. Wool Milling

takes place much later when clustering is performed on the basis of similarity of input coefficients (see Figure 4.3.1) than when clustering is performed on the basis of similarity of partially aggregated input coefficients (see Figure 4.1.1). Nevertheless, the degree of similarity between the input structures of these two industries may appear surprisingly high. This is probably due to the level of aggregation of the Farming industry in the 109 industry, interindustry table. Farming buys a high proportion of inputs from itself and Wool Milling buys a large proportion of inputs from the Farming industry.

**Figure 4.3.1**

The Formation of Sector 1 in Model 4

<table>
<thead>
<tr>
<th>Industry</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>21</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>14</td>
</tr>
</tbody>
</table>

No. of Sectors in Model when Industries are Merged

69  50  45
(b) The following groups of industries appear in Model 2, are grouped into a single sector in Model 4, but do not appear in Model 1.

1. Hunting and Fishing,
2. Road Transport.

19. Tobacco and Cigarettes.
17. Malting and Brewing,
49. Printing and Trade Services,
75. Electro-Plating,
85. Vehicle Repair,
99. Trade,
106. Services.

20. Wool Scouring,
62. Petroleum and Coal Products,
65. Pottery Clay Products,
83. Vehicle Assembly,
86. Aircraft Repair,
105. Communications,
107. Ownership of Property,
108. Services to Households etc.,
109. Services to Government.

The order in which these industries are aggregated, to form sector 2 in Model 4, is displayed in Figure 4.3.2.
Figure 4.3.2 reveals that the following groups of industries have surprisingly similar input structures:

   62. Petroleum and Coal Products,
   86. Aircraft Repair,
   83. Vehicle Assembly,
   105. Communications,
   108. Services to Household etc.,
   and 109. Services to Government.
   49. Printing and Trade Services,
   and 75. Electro-Plating.
   20. Wool Scouring,
   and 65. Pottery Clay Products.
   99. Trade,
   and 106. Services.
   85. Vehicle Repair,
   and 104. Road Transport.

All but the last of these groups of industries are also formed early in the derivation of Model 2. Further aggregations take place at a stage when the discrimination between industries is not very high. In particular, the aggregation of industries 2 and 104 and of industries 17, 49, 75, 85, 99 and 106 do not occur until the 48 sector model is formed. Industries 20, 62, 65, 83, 86, 105, 107, 108 and 109 are aggregated to form the 60 sector model.

(c) Industries 13 (Chocolate and Confectionery) and 18 (Aerated Waters and Cordials) are aggregated to form the 76 sector model, and are joined with industry 15 (Food Preparations N.E.I.) relatively late in the clustering process, when the 64 sector model is formed. This grouping also appears in Model 2.
(d) It is surprising to find industry 3 (Forestry) aggregated with industry 44 (Pulp and Paper) in the 66 sector version of Model 4 and in the 59 sector version of Model 2.

(e) The following groups of industries appear in Model 2, they form part of a single sector in Model 4, but they do not appear in Model 1.

27. Underclothing,
30. Shirts and Pyjamas,
34. Made-up Textiles,
42. Mattresses,
53. Tyres and Tubes,
54. Other Rubber Goods,
56. Chemical Fertilizers,
67. Glass Products.

28. Millinery and Hats,
33. Canvas Goods,
89. Brushes and Brooms.

The order in which these industries, along with industries:

29. Corsetry
41. Furniture
52. Leather Goods
90. Toys and Sports Goods
91. Manufacturing N.E.I.
92. Plastics Manufacturing

are aggregated, to form sector 16 in Model 4, is displayed in Figure 4.3.3.
The amalgamation of industries, prior to the formation of the 79 sector model, meets with expectations, with the exception of the merger of industries 30 (Shirts and Pyjamas), 91 (Manufacturing N.E.I.) and 42 (Mattresses). The groupings which emerge at higher levels of aggregation were all unexpected.

(f) The following industries are grouped into the same sector in Model 2 and in Model 4.

71. Basic Metal Industries 78. Machinery N.E.I.
72. Sheetmetal Working 79. Range Making

This grouping is considered to be acceptable on intuitive grounds.
The sectors produced by aggregating industries according to similarity of input coefficients are not significantly more appealing than those produced by grouping industries according to similarity of partially aggregated input coefficients. Model 2 and Model 4 contain a considerable number of overlapping sectors, and both contain sectors which would be unacceptable for most analyses.

A fifth model, Model 5, was produced by aggregating industries according to similarity of input coefficients, but with the same constraints imposed upon it as those imposed upon Model 3 (see Table 4.3.6). All but nine of the sectors in Model 5 also appear in Model 3.

In order to compare the similarity of input structures of the sectors in Models 1, 2, 3, 4 and 5, the within sector, sums of squared deviations of input coefficients from their means have been calculated also for the former three models. The similarity measures for Models 2, 3, 4 and 5 are 0.42537, 0.43270, 0.24578 and 0.39525, respectively, compared with Model 1's value of 1.45733. Hence, the official 44 sector classification of the 109 industries is less homogeneous, in terms of input structures, than the models produced by clustering either on the basis of similarity of input coefficients, or on the basis of similarity of partially aggregated input coefficients, with or without constraints being imposed.
Table 4.3.6
Model 5: Constrained 44 Sector Model Produced by Ward's Algorithm

<table>
<thead>
<tr>
<th>Sector</th>
<th>Industries Comprising the Sector</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>5 11</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>10 14</td>
</tr>
<tr>
<td>11</td>
<td>12 13 15 17 18 19</td>
</tr>
<tr>
<td>12</td>
<td>16</td>
</tr>
<tr>
<td>13</td>
<td>20 23</td>
</tr>
<tr>
<td>14</td>
<td>21 22 24</td>
</tr>
<tr>
<td>15</td>
<td>25 26 27 28 29 30 31 33 34</td>
</tr>
<tr>
<td>16</td>
<td>32</td>
</tr>
<tr>
<td>17</td>
<td>35</td>
</tr>
<tr>
<td>18</td>
<td>36 38</td>
</tr>
<tr>
<td>19</td>
<td>37</td>
</tr>
<tr>
<td>20</td>
<td>39 41 42</td>
</tr>
<tr>
<td>21</td>
<td>40</td>
</tr>
<tr>
<td>22</td>
<td>43</td>
</tr>
<tr>
<td>23</td>
<td>44 49</td>
</tr>
<tr>
<td>24</td>
<td>45 46 47 48</td>
</tr>
<tr>
<td>25</td>
<td>50 51</td>
</tr>
<tr>
<td>26</td>
<td>52</td>
</tr>
<tr>
<td>27</td>
<td>53 54</td>
</tr>
<tr>
<td>28</td>
<td>55</td>
</tr>
<tr>
<td>29</td>
<td>56 59 60 61 62</td>
</tr>
<tr>
<td>30</td>
<td>57 58</td>
</tr>
<tr>
<td>31</td>
<td>63</td>
</tr>
<tr>
<td>32</td>
<td>64 65 66 67 69 70</td>
</tr>
<tr>
<td>33</td>
<td>68</td>
</tr>
<tr>
<td>34</td>
<td>71 72 75 76</td>
</tr>
<tr>
<td>35</td>
<td>73 74</td>
</tr>
<tr>
<td>36</td>
<td>77 78 79 80 81 82 83 84 85 86 87</td>
</tr>
<tr>
<td>37</td>
<td>88 89 90 91 92</td>
</tr>
<tr>
<td>38</td>
<td>93 94</td>
</tr>
<tr>
<td>39</td>
<td>95 96</td>
</tr>
<tr>
<td>40</td>
<td>97</td>
</tr>
<tr>
<td>41</td>
<td>98 99 100</td>
</tr>
<tr>
<td>42</td>
<td>101 104 105 106 107 108 109</td>
</tr>
<tr>
<td>43</td>
<td>102</td>
</tr>
<tr>
<td>44</td>
<td>103</td>
</tr>
</tbody>
</table>
4.4 An Assessment of Aggregation Bias in Forecasts of Gross Output

The main advantage of using a formalised procedure to determine the intersectoral model is that sectors are formed on the basis of a criterion, such as similarity of partially aggregated input coefficients or similarity of input coefficients, which is known to be theoretically sound, rather than on the basis of subjective judgement. Whether the criterion is similarity of partially aggregated input coefficients or similarity of input coefficients, the aim is the same; to group industries into sectors so that the elements of the matrix \( (\bar{A}T - TA) \) are as close to zero as possible. As a result, the elements of the matrix

\[
(I - \bar{A})^{-1}T - (I - A)^{-1}(\bar{A}T - TA)(I - A)^{-1}
\]

will be close to zero, and aggregation bias, itself,

\[
\bar{X} - TX = \{(I - \bar{A})^{-1}T - (I - A)^{-1}\}F
\]

will be as close to zero as possible.

The question, which remains to be answered, is whether or not grouping industries according to similarity of partially aggregated input coefficients results in significantly less aggregation bias than grouping industries according to similarity of input coefficients. The objective of this section is to compare the five models, presented in previous sections of this chapter, in terms of the amount of aggregation bias, which is likely to be present in each of their sectors.
The task is complicated by the fact that aggregation bias in each sector gross output is a function of the specific final demand vector being employed. However, aggregation bias vanishes when final demands are proportional to those of the base period, so each row of the matrix

\[(I-\bar{A})^{-1}T - T(I-A)^{-1}\]

must contain both positive and negative elements (unless all of its elements are zero). Consequently, even if the elements of \((I-\bar{A})^{-1}T - T(I-A)^{-1}\) are not close to zero, there will exist many final demand vectors, for which one sector has zero aggregation bias, although aggregation bias is unlikely to vanish in the remaining sectors. However, if final demands are regarded as random variables, small amounts of aggregation bias are likely to be present in the majority of sectors if each row of the matrix \((I-\bar{A})^{-1}T - T(I-A)^{-1}\) contains elements, which have a mean of zero and a small variance.

Hence, the measure, which has been chosen to compare the sectors of Models 1, 2, 3, 4 and 5 in terms of their probable degrees of aggregation bias, is the sum of squares of elements in each row of the matrix \((I-\bar{A})^{-1}T - T(I-A)^{-1}\). This measure is also the squared Euclidean distance, from the origin, of a point in 109 dimensional space. The larger its value, the larger is the amount of probable aggregation bias for the sector. The total sums of squares for all 44 sectors is a measure of the degree of probable aggregation bias for all 44 sectors of a given model.
The index of aggregation bias, referred to above, was calculated for each of the 44 sectors in the five models presented earlier in this chapter, namely, the Department of Statistics' model (Model 1), the model produced by aggregating according to similarity of partially aggregated input coefficients (Model 2), the model produced by aggregating according to partially aggregated input coefficients, but with constraints imposed (Model 3), the model produced by aggregating according to similarity of input coefficients (Model 4) and the model produced by aggregating according to similarity of input coefficients but with constraints imposed (Model 5). The 44 sectors of each model were ranked in descending order, according to the index of aggregation bias, and the results are presented in Table 4.4.1.

A comparison of the indices of aggregation bias between sectors of the five different models (as ordered in Table 4.4.1) reveals that:

(a) Model 1 has a larger index of aggregation bias than Model 2 in 38 of the 44 sectors, while the remaining six sectors have the same indices in both models. Model 1 has a larger index of aggregation bias than Model 3 in 33 of the 44 sectors, Model 3 has the larger index in six sectors and the remaining five sectors have equal values for the index in both models. Model 1 has higher values for the index of aggregation bias than Model 4 in 38 of the 44 sectors, Model 4 has the larger value in one sector, and the index is the same
### Table 4.4.1

Indices of Aggregation Bias in Forecasts of Gross Output

<table>
<thead>
<tr>
<th>Sector</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Index</td>
<td>Sector 1 Index</td>
<td>Sector 2 Index</td>
<td>Sector 3 Index</td>
<td>Sector 4 Index</td>
</tr>
<tr>
<td>1</td>
<td>1.29435</td>
<td>0.05205</td>
<td>0.11742</td>
<td>0.17343</td>
<td>0.12359</td>
</tr>
<tr>
<td>15</td>
<td>0.25037</td>
<td>0.05008</td>
<td>0.08439</td>
<td>0.07536</td>
<td>0.10975</td>
</tr>
<tr>
<td>17</td>
<td>0.17062</td>
<td>0.03689</td>
<td>0.07987</td>
<td>0.06557</td>
<td>0.09005</td>
</tr>
<tr>
<td>5</td>
<td>0.15075</td>
<td>0.02916</td>
<td>0.05948</td>
<td>0.03209</td>
<td>0.08051</td>
</tr>
<tr>
<td>24</td>
<td>0.10830</td>
<td>0.02891</td>
<td>0.04549</td>
<td>0.02623</td>
<td>0.07255</td>
</tr>
<tr>
<td>35</td>
<td>0.08532</td>
<td>0.02166</td>
<td>0.04088</td>
<td>0.02680</td>
<td>0.05458</td>
</tr>
<tr>
<td>3</td>
<td>0.08127</td>
<td>0.01704</td>
<td>0.03084</td>
<td>0.02036</td>
<td>0.04645</td>
</tr>
<tr>
<td>7</td>
<td>0.06321</td>
<td>0.01300</td>
<td>0.03512</td>
<td>0.01790</td>
<td>0.04050</td>
</tr>
<tr>
<td>23</td>
<td>0.05664</td>
<td>0.01172</td>
<td>0.02508</td>
<td>0.01498</td>
<td>0.03514</td>
</tr>
<tr>
<td>6</td>
<td>0.04944</td>
<td>0.00853</td>
<td>0.02278</td>
<td>0.01270</td>
<td>0.02289</td>
</tr>
<tr>
<td>22</td>
<td>0.03944</td>
<td>0.00814</td>
<td>0.02219</td>
<td>0.01039</td>
<td>0.01938</td>
</tr>
<tr>
<td>20</td>
<td>0.03822</td>
<td>0.00654</td>
<td>0.02162</td>
<td>0.00993</td>
<td>0.01791</td>
</tr>
<tr>
<td>11</td>
<td>0.03424</td>
<td>0.00422</td>
<td>0.02007</td>
<td>0.00780</td>
<td>0.01610</td>
</tr>
<tr>
<td>40</td>
<td>0.02896</td>
<td>0.00351</td>
<td>0.01870</td>
<td>0.00609</td>
<td>0.01483</td>
</tr>
<tr>
<td>25</td>
<td>0.02628</td>
<td>0.00256</td>
<td>0.01786</td>
<td>0.00606</td>
<td>0.01418</td>
</tr>
<tr>
<td>12</td>
<td>0.01305</td>
<td>0.00173</td>
<td>0.01520</td>
<td>0.00501</td>
<td>0.00964</td>
</tr>
<tr>
<td>4</td>
<td>0.01274</td>
<td>0.00125</td>
<td>0.01455</td>
<td>0.00441</td>
<td>0.00820</td>
</tr>
<tr>
<td>33</td>
<td>0.01264</td>
<td>0.00121</td>
<td>0.01358</td>
<td>0.00430</td>
<td>0.00733</td>
</tr>
<tr>
<td>18</td>
<td>0.00951</td>
<td>0.00094</td>
<td>0.00962</td>
<td>0.00409</td>
<td>0.00494</td>
</tr>
<tr>
<td>42</td>
<td>0.00946</td>
<td>0.00076</td>
<td>0.00642</td>
<td>0.00343</td>
<td>0.00487</td>
</tr>
<tr>
<td>27</td>
<td>0.00799</td>
<td>0.00046</td>
<td>0.00496</td>
<td>0.00292</td>
<td>0.00396</td>
</tr>
<tr>
<td>29</td>
<td>0.00708</td>
<td>0.00045</td>
<td>0.00398</td>
<td>0.00277</td>
<td>0.00281</td>
</tr>
<tr>
<td>37</td>
<td>0.00662</td>
<td>0.00044</td>
<td>0.00295</td>
<td>0.00248</td>
<td>0.00321</td>
</tr>
<tr>
<td>43</td>
<td>0.00531</td>
<td>0.00041</td>
<td>0.00231</td>
<td>0.00089</td>
<td>0.00202</td>
</tr>
<tr>
<td>16</td>
<td>0.00486</td>
<td>0.00041</td>
<td>0.00180</td>
<td>0.00078</td>
<td>0.00155</td>
</tr>
<tr>
<td>26</td>
<td>0.00428</td>
<td>0.00040</td>
<td>0.00156</td>
<td>0.00065</td>
<td>0.00104</td>
</tr>
<tr>
<td>32</td>
<td>0.00417</td>
<td>0.00037</td>
<td>0.00103</td>
<td>0.00062</td>
<td>0.00091</td>
</tr>
<tr>
<td>38</td>
<td>0.00366</td>
<td>0.00033</td>
<td>0.00094</td>
<td>0.00059</td>
<td>0.00074</td>
</tr>
<tr>
<td>36</td>
<td>0.00300</td>
<td>0.00029</td>
<td>0.00077</td>
<td>0.00050</td>
<td>0.00055</td>
</tr>
<tr>
<td>13</td>
<td>0.00279</td>
<td>0.00024</td>
<td>0.00039</td>
<td>0.00044</td>
<td>0.00053</td>
</tr>
<tr>
<td>19</td>
<td>0.00275</td>
<td>0.00022</td>
<td>0.00031</td>
<td>0.00024</td>
<td>0.00038</td>
</tr>
<tr>
<td>9</td>
<td>0.00148</td>
<td>0.00012</td>
<td>0.00029</td>
<td>0.00024</td>
<td>0.00031</td>
</tr>
<tr>
<td>21</td>
<td>0.00134</td>
<td>0.00007</td>
<td>0.00020</td>
<td>0.00012</td>
<td>0.00028</td>
</tr>
<tr>
<td>41</td>
<td>0.00098</td>
<td>0.00002</td>
<td>0.00012</td>
<td>0.00012</td>
<td>0.00020</td>
</tr>
<tr>
<td>2</td>
<td>0.00034</td>
<td>0.00001</td>
<td>0.00011</td>
<td>0.00007</td>
<td>0.00012</td>
</tr>
<tr>
<td>39</td>
<td>0.00006</td>
<td>0.00000</td>
<td>0.00002</td>
<td>0.00001</td>
<td>0.00001</td>
</tr>
<tr>
<td>34</td>
<td>0.00003</td>
<td>0.00000</td>
<td>0.00001</td>
<td>0.00001</td>
<td>0.00001</td>
</tr>
<tr>
<td>8</td>
<td>0.00003</td>
<td>0.00000</td>
<td>0.00001</td>
<td>0.00001</td>
<td>0.00001</td>
</tr>
<tr>
<td>28</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00001</td>
<td>0.00001</td>
<td>0.00001</td>
</tr>
<tr>
<td>14</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00001</td>
<td>0.00001</td>
<td>0.00001</td>
</tr>
<tr>
<td>10</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00001</td>
<td>0.00001</td>
<td>0.00001</td>
</tr>
<tr>
<td>30</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00001</td>
<td>0.00001</td>
<td>0.00001</td>
</tr>
<tr>
<td>31</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00001</td>
<td>0.00001</td>
<td>0.00001</td>
</tr>
<tr>
<td>44</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00001</td>
<td>0.00001</td>
<td>0.00001</td>
</tr>
</tbody>
</table>

Total 2.59366  Total 0.30495  Total 0.73066  Total 0.53631  Total 0.81135
in the remaining five sectors of both models. Model 1 has a larger aggregation bias index than Model 5 in 36 of the 44 sectors, Model 5 has the larger index in three sectors, and the remaining five sectors have the same index in both models. Hence, it is concluded that the Department of Statistics' model is likely to result in larger aggregation errors in forecasts of gross output than both models produced by aggregating according to similarity of partially aggregated input coefficients (unconstrained or constrained) and both models produced by aggregating according to similarity of input coefficients (unconstrained or constrained). The superiority of Model 3 and Model 5 over Model 1 is particularly noteworthy, since Models 3 and 5 consist of sectors which are acceptable on \textit{a priori} grounds.

(b) Model 2 has a larger index of aggregation bias than Model 4 in only one sector, Model 4's index is larger than Model 2's index in 38 sectors and the index is the same in the remaining five sectors. Consequently, it is concluded that the model produced by aggregating according to similarity of partially aggregated input coefficients is likely to result in less aggregation bias, in forecasts of gross output, than the model produced by aggregating according to similarity of input coefficients.

(c) Model 3 has larger values for the index of aggregation bias than Model 5 in 19 sectors, Model 5's index is larger
than Model 3's in 18 sectors, and the index is the same in the remaining seven sectors. Hence, the two constrained intersectoral models are likely to give rise to approximately the same amounts of aggregation bias in forecasts of gross outputs. This result was to be expected since Model 3 and Model 5 are very similar in terms of the composition of their sectors.

(d) The totals at the bottom of Table 4.4.1 indicate that, when all sectors are taken into account, Model 2 is likely to result in the smallest amount of aggregation bias, followed by Model 4, Model 3, Model 5 and then Model 1.
4.5 Summary

This chapter has been devoted to a comparison of the following five models of the 1965-66 New Zealand economy:

Model 1: The 44 sector classification of industries used by the New Zealand Department of Statistics.

Model 2: The 44 sector model produced by aggregating into the same sector industries which have similar partially aggregated input coefficients.

Model 3: The 44 sector model produced by aggregating industries with similar partially aggregated input coefficients, but with constraints imposed to ensure that aggregation took place only within certain groups of industries.

Model 4: The 44 sector model produced by aggregating into the same sector industries which have similar input coefficients.

Model 5: The 44 sector model produced by aggregating industries with similar input coefficients, but with the same constraints imposed upon it as those imposed upon Model 3.

Appendices B and C, at the end of this chapter, provide a pictorial comparison of the composition of sectors in Models 1, 2 and 4 and Models 1, 3 and 5, respectively.

For each model the total, within sector, sum of squared deviations of partially aggregated input coefficients from their means, and the total, within sector, sum of
squared deviations of input coefficients from their means, were calculated and are summarised in Table 4.5.1 below.

Table 4.5.1

<table>
<thead>
<tr>
<th></th>
<th>Total Sum of Squared Deviations of Partially Aggregated Input Coefficients from their Means</th>
<th>Total Sum of Squared Deviations of Input Coefficients from their Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>1.46484</td>
<td>1.45733</td>
</tr>
<tr>
<td>Model 2</td>
<td>0.15058</td>
<td>0.42537</td>
</tr>
<tr>
<td>Model 3</td>
<td>0.40350</td>
<td>0.43270</td>
</tr>
<tr>
<td>Model 4</td>
<td>0.28237</td>
<td>0.24578</td>
</tr>
<tr>
<td>Model 5</td>
<td>0.43620</td>
<td>0.39525</td>
</tr>
</tbody>
</table>

Since aggregation bias in forecasts of gross sector outputs vanishes when the partially aggregated input coefficients, of industries of the same sector, are equal, it may be concluded that Model 2, which has the most similar partially aggregated input coefficients, is likely to result in the smallest amount of aggregation bias, followed by Model 4, Model 3, Model 5 and then Model 1. However, equality of input coefficients is a sufficient, although not a necessary, condition for aggregation bias to vanish, so it is interesting to rank the five models according to this criterion. Table 4.5.1 reveals that Model 4 has the most similar input coefficients of industries within sectors, followed by Model 5, Model 2, Model 3 and then Model 1. Consequently, if Model 1 has been formed by aggregating, into the same
sector, industries, which are expected intuitively to have similar input coefficients, it may be argued that subjective judgement does not always correspond to reality.

The justification for using either similarity of partially aggregated input coefficients, or similarity of input coefficients, as the criterion for aggregation is that, when it is completely satisfied, aggregation bias vanishes. Aggregation bias is given by the elements of the vector

$$\bar{x} - TX = \{(I-\bar{A})^{-1}T - T(I-A)^{-1}\}F$$

and so is a function of the levels of final demand. Although it can be measured only for a specific set of final demands, the closer are the elements of the matrix

$$(I-\bar{A})T - T(I-A)^{-1}$$

to zero, the smaller is the amount of aggregation bias, which is likely to be present for most final demand vectors. According to this criterion, Model 2 is likely to result in less aggregation bias than Model 4, followed by Models 3 and 5, and then by Model 1. Hence it is concluded that aggregation on the basis of partially aggregated input coefficients is more effective in reducing aggregation bias than aggregation on the basis of input coefficients. Although Model 2 and Model 4 both contain sectors, the composition of which may not be acceptable for many analyses, Model 3 and Model 5 do not contain such sectors, and yet they are superior to Model 1, according to all three criteria discussed above.
Industry Classification in the 1965-66 Input-Output Model of the New Zealand Economy

1. Farming
2. Hunting and Fishing
3. Forestry
4. Mining and Quarrying
5. Meat Freezing and Preserving
6. Ham, Bacon and Smallgoods
7. Ice Cream
8. Butter, Cheese, Milk-Powder etc.
9. Milk Treatment
10. Fruit and Vegetable Preserving
11. Grain Milling
12. Biscuits and Bread Baking
13. Chocolate and Confectionery
14. Animal Feed
15. Food Preparations N.E.I.
16. Wine Making
17. Malting and Brewing
18. Aerated Waters and Cordials
19. Tobacco and Cigarettes
20. Woolscouring
21. Wool Milling
22. Other Spinning and Weaving
23. Hosiery and Other Knitting
24. Textiles N.E.I.
25. Men's Outerwear
26. Women's Outerwear
27. Underclothing
28. Millinery and Hats
29. Corsetry
30. Shirts and Pyjamas
31. Apparel N.E.I.
32. Footwear - not Rubber
33. Canvas Goods
34. Made-up Textiles
35. Sawmills
36. Planing Mills
37. Joinery
38. Wooden Containers
39. Plywood and Veneer
40. Wood Products N.E.I.
41. Furniture
42. Mattresses
43. Venetian Blinds
44. Pulp and Paper
45. Cartons and Paper Bags
46. Paper Products N.E.I.
47. Printing and Publishing
48. Job and General Printing
49. Printing and Trade Services
50. Tanning
51. Fellmongery
52. Leather Goods
53. Tyres and Tubes
54. Other Rubber Goods
55. Tyre Retreading
56. Chemical Fertilizers
57. Vegetable and Animal Oils
58. Soap and Candle
59. Paint and Varnish
60. Medical and Toilet Goods
61. Chemical Products N.E.I.
62. Petroleum and Coal Products
63. Bituminous Materials
64. Structural Clay Products
65. Pottery Clay Products
66. Cement
67. Glass Products
68. Concrete Products
69. Lime
70. Mineral Products N.E.I.
71. Basic Metal Industries
72. Sheetmetal Working
73. Wire Working
74. Nail Making
75. Electro-Plating
76. Metal Products N.E.I.
77. Farm Machinery
78. Machinery N.E.I.
79. Range Making
80. Radio and TV Assembly
81. Electrical Goods N.E.I.
82. Boat Building and Repairs
83. Vehicle Assembly
84. Body Building
85. Vehicle Repair
86. Aircraft Repair
87. Transport Equipment N.E.I.
88. Jewellery
89. Brushes and Brooms
90. Toys and Sports Goods
91. Manufacturing N.E.I.
92. Plastics Manufacturing
93. Residential Building
94. Commercial Building
95. Civil Engineering
96. Other Building Activities
97. Electricity and Gas
98. Water and Sanitation
99. Trade
100. Banking and Insurance
101. Rail Transport
102. Shipping Transport
103. Air Transport
104. Road Transport
105. Communications
106. Services
107. Ownership of Property
108. Services to Households etc.
109. Services to Government
### Classification of Industries in Model 1, Model 2 and Model 4

<table>
<thead>
<tr>
<th>Industry Type</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>10. Fruit and Vegetable Preserving</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Farming</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20. Wool scouring</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21. Wool Milling</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22. Other Spinning and Weaving</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24. Textiles N.E.I.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23. Hosiery and other Knitting</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. Tobacco and Cigarettes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25. Men's OUTERWEAR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26. Women's OUTERWEAR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27. Umbrella Clothing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28. Millinery and Hats</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29. Corsetry</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30. Shirts and Pyjamas</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31. Apparel N.E.I.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33. Canvas Goods</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34. Made-up Textiles</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>83. Vehicle Assembly</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>107. Ownership of Property</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>109. Services to Households etc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>109. Services to Government</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>106. Services</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>104. Road Transport</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Hunting and Fishing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>99. Trade</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>53. Tyres and Tubes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54. Other Rubber Goods</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55. Tyre Retreading</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>57. Vegetable and Animal Oils</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>58. Soap and Candle</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>59. Paint and Varnish</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60. Medical and toilet Goods</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>61. Chemical Products N.E.I.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>62. Petroleum and Coal Products</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>63. Bituminous Materials</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>56. Chemical Fertilizers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41. Furniture</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42. Mattressess</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43. Venetian Blinds</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>88. Jewellery</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>89. Brushes and Brooms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90. Toys and Sports Goods</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>91. Manufacturing N.E.I.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>92. Pastics Manufacturing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Wire Making</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Melting and brewing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. Aerated Waters and Cordials</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. Food Preparations N.E.I.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Animal Feed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Chocolate and Confectionery</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Biscuits and Bread Baking</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Grain Milling</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Ice Cream</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32. Footwear - not Rubber</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Classification of Industries in Model 1, Model 3 and Model 5

<table>
<thead>
<tr>
<th>Model 1</th>
<th>Model 3</th>
<th>Model 5</th>
<th>Model 1</th>
<th>Model 3</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Farming</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Hunting and Fishing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Forestry</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Mining and Quarrying</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Meat Freezing and Preserving</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Ham, Bacon and Smallgoods</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Ice Cream</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Grain Milling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Biscuits and Bread Baking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Chocolate and Confectionary</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Animal Feed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Food Preparations N.E.I.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Butter, Cheese, Milk-Powder etc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Milk Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Fruit and Vegetable Preserving</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Wine Making</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Melting and Browing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. Aerated Waters and Cordials</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. Tobacco and Cigarettes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20. Woollen yarns</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21. Wool Milling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22. Other Spinning and Weaving</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23. Hosey and Other Knitting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24. Textiles N.E.I.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25. Men's Outerwear</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26. Women's Outerwear</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27. Underclothing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28. Millinery and Hats</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29. Corsetry</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30. Shirts and Pyjamas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31. Apparel N.E.I.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33. Canvas Goods</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34. Made-up Textiles</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32. Footwear - not Rubber</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35. Sawmills</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36. Planing Mills</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37. Joinery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38. Wooden Containers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39. Plywood and Veneer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40. Wood Products N.E.I.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41. Furniture</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42. Mattresses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43. Venetian Blinds</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44. Pulp and Paper</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45. Cartons and Paper Bags</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46. Paper Products N.E.I.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47. Printing and Publishing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48. Job and General Printing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>49. Printing and Trade Services</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50. Tanning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>51. Polynonery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52. Leather Goods</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>53. Tyres and Tubes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54. Other Rubber Goods</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55. Tyre Retreading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>56. Chemical Fertilizers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>57. Vegetable and Animal Oils</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>58. Soap and Candle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>59. Paint and Varnish</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60. Medical and Toilet Goods</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>61. Chemical Products N.E.I.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>62. Petroleum and Coal Products</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>63. Immiscible Materials</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64. Structural Clay Products</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65. Pottery Clay Products</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>66. Cement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>67. Glass Products</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>68. Concrete Products</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>69. Lime</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70. Mineral Products N.E.I.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>71. Basic Metal Industries</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>72. Sheetmetal Working</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>73. Wire Working</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>74. Nail Making</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75. Electro-Plating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>76. Metal Products N.E.I.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>77. Farm Machinery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>78. Machinery N.E.I.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>79. Range Making</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80. Radio and TV Assembly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>81. Electrical Goods N.E.I.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>82. Boat Building and Repairs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>83. Vehicle Assembly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>84. Body Building</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>85. Vehicle Repair</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>86. Aircraft Repair</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>87. Transport Equipment N.E.I.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>88. Jewellery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>89. Brushes and Brooms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90. Toys and Sports Goods</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>91. Manufacturing N.E.I.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>92. Plastics Manufacturing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>93. Residential Building</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>94. Commercial Building</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95. Civil Engineering</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>96. Other Building Activities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>97. Electricity and Gas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>98. Water and Sanitation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>99. Trade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100. Banking and Insurance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>101. Rail Transport</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>102. Shipping Transport</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>103. Air Transport</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>104. Road Transport</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>105. Communications</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>106. Services</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>107. Ownership of Property</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>108. Services to Households etc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>109. Services to Government</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Aggregation, as it applies to input-output analysis, is the process of combining industries of an original input-output model into sectors, in order to form a hybrid model, the dimensions of which are smaller than those of the original model. The problem, which arises as a result of the aggregation process, is that, in general, forecasts of sector gross outputs, obtained from the intersectoral model, are biased in that they differ from forecasts of sector gross outputs, obtained by aggregating forecasts of industry gross outputs, produced by the original interindustry model.

There are, however, a number of conditions under which aggregation bias vanishes in static, open, input-output models, namely:

(a) When industries, which are grouped into the same sector, have equal partially aggregated input coefficients.
(b) When industries, which are grouped into the same sector, have equal input coefficients.
(c) When the outputs of industries, which are grouped into the same sector, are used in fixed proportions by other industries and by final consumers.
(d) When \( k \) industries, which are grouped into the same sector, are such that the output of industry \( i \) is absorbed entirely by industry \( i+1 \) (\( i=1,2,...,k-1 \)).
(e) When final demands are equal or proportional to those of the base period.
In addition, aggregation bias is small under the following conditions in that first-order aggregation bias vanishes:

(f) When final demands are proportional to base period gross outputs.

(g) When the structure of final demands, within each sector, corresponds to that of base period gross outputs.

(h) When the only final demands, which differ from their base period values, are those of industries which are not aggregated with other industries in the intersectoral model.

Most of the existing methodologies, for aggregating industries of an input-output model, employ a cluster analysis approach, where the criterion for aggregation is similarity of input coefficients of industries within sectors. In Chapter 3 of this thesis, an alternative clustering algorithm was developed, which aggregates into the same sector industries which have similar partially aggregated input coefficients. Similarity of partially aggregated input coefficients is considered to be a better criterion for aggregation than similarity of input coefficients, since equality of partially aggregated input coefficients is a necessary and sufficient condition for aggregation bias to vanish. Equality of input coefficients is a sufficient, but not a necessary, condition for aggregation bias to vanish.

The algorithm allows the input-output analyst to impose constraints upon the intersectoral model to ensure that
specific industries remain isolated, to force certain industries to be aggregated into the same sector, or to ensure that certain industries are not aggregated into the same sector. Consequently, the analyst can ensure that the intersectoral model is operationally meaningful and is produced using a formalised procedure, which eliminates the need for subjective judgement.

In Chapter 4, the algorithm was used to aggregate the 109 industries of the 1965-66 input-output model of the New Zealand economy into 44 sectors. The resulting intersectoral model was compared to the official 44 sector classification of industries, published by the New Zealand Department of Statistics. It was found that the composition of sectors differed considerably in the two models, so constraints were imposed in an effort to produce an intersectoral model, which is more intuitively appealing. The sectors appearing in both the models, produced by the algorithm, were found to be more homogeneous than the Department of Statistics' model, in terms of partially aggregated input coefficients and in terms of input coefficients.

A third model was produced by aggregating on the basis of similarity of input coefficients, a method used by previous authors. This model also contained sectors, which differed markedly from the official classification, and its sectors had more similar input coefficients and partially aggregated input coefficients than the Department of Statistics' model.
A fourth model was produced by aggregating on the basis of similarity of input coefficients, but with the same set of constraints imposed upon it as those specified for the model produced by constrained aggregation according to similarity of partially aggregated input coefficients. The two constrained models contained 35 identical sectors. As expected, constrained aggregation according to similarity of partially aggregated input coefficients produced sectors with more similar partially aggregated input coefficients and constrained aggregation according to similarity of input coefficients resulted in sectors with more similar input coefficients.

Finally, the five models were compared using an index of probable aggregation bias. The model produced by aggregating on the basis of similarity of partially aggregated input coefficients was found to have the smallest amount of probable aggregation bias, followed by the model produced by aggregating using similarity of input coefficients as the criterion. These were followed by the model produced by constrained aggregation on the basis of partially aggregated input coefficients and the model produced by constrained aggregation using similarity of input coefficients, and finally the official 44 sector model.

The methodology, which has been developed and applied in this thesis, is suitable for aggregating industries of static, open, input-output models; it cannot be applied to
dynamic input-output models. The development of an aggregation procedure for use with dynamic input-output models remains a topic for further research.
REFERENCES


APPENDIX

PROGRAM Input-Output Analysis:
Aggregation of Industries into Sectors.

PROGRAMMER J.R. Rodgers,
Lincoln College.


LANGUAGE FORTRAN IV.

INSTALLATION Burroughs' 6700,
University of Canterbury,
Christchurch,
New Zealand.

DEVICES REQUIRED Card Reader,
Line Printer,
Magnetic Tape.
1. FUNCTION OF THE PROGRAM

The function of the program is to aggregate industries of an input-output model into sectors, so that aggregation bias in forecasts of sector gross outputs is minimized.

2. GENERAL DESCRIPTION

2.1 The Interindustry Model

A matrix of interindustry transaction flows and a set of industry gross outputs must be input to the program, using one of the following options:

(a) The data may be read from cards and written to magnetic tape (from which it can be read during later runs).
(b) The data may be read from cards, but not written to magnetic tape.
(c) The data may be read from a magnetic tape, on which it has been stored during a previous run.

The desired option is specified by setting a flag, which is read from cards, to one of three values. In general, the user will find it most convenient to enter the data from cards and store it on magnetic tape during the first run, and read it from magnetic tape during subsequent runs.

Interindustry transactions and industry gross outputs are used to construct a matrix of input-output coefficients for use by the aggregation procedure. The interindustry
transactions, input-output coefficients and industry gross outputs are written to the line printer, unless the user specifies that their printing should be suppressed.

2.2 Imposition of Constraints upon the Intersectoral Model

Various constraints can be placed upon the intersectoral model to ensure that certain industries are, or are not, aggregated into the same sector. The user may constrain the intersectoral model in one or more of the following ways:

(a) One or more sets of industries, where aggregation between members of the same set is prohibited, may be read from cards.

(b) A list of industries, each of which is to remain isolated in the intersectoral model, may be read from cards.

(c) One or more sets of industries, where aggregation may take place only between industries of the same set, may be read from cards.

(d) One or more sets of industries, where all industries of the same set must be aggregated into a single sector, may be read from cards.

Note that more than one of these options may achieve the same result. For example, given a six industry model, the user can ensure that aggregation takes place only between industries one to three and between industries four to six by specifying, using option (a) above, that aggregation is prohibited between
the following sets of industries:
\{1,4\} \{1,5\} \{1,6\} \{2,4\} \{2,5\} \{2,6\} \{3,4\} \{3,5\} \{3,6\}

The simpler method, however, is to specify, using option (c) above, that aggregation is restricted to sets of industries:
\{1,2,3\} and \{4,5,6\}

2.3 Construction of the Intersectoral Model

Industries are aggregated into sectors using one of two clustering algorithms:

(a) Aggregation may be performed on the basis of similarity of input coefficients, as measured by the total, within sector, sum of squared deviations of input coefficients from their means, (that is, using Ward's method).

(b) Aggregation may be performed on the basis of similarity of partially aggregated input coefficients, as measured by the total, within sector, sum of squared deviations of partially aggregated input coefficients from their means, (that is, using the algorithm developed in this thesis).

The desired option is specified by setting a flag, which is read from cards, to one of two values.

The user specifies, via card input, the number of sectors into which industries are to be grouped. If however, constraints imposed upon the intersectoral model are such that this level of aggregation cannot be reached, each algorithm will terminate when all possible mergers have been completed.
2.4 The Intersectoral Model

The composition of each sector is printed as it is formed and, at the conclusion of the clustering algorithm, the final composition of each sector is printed. A listing of intersector transactions, aggregated input-output coefficients and sector gross outputs may be written to the line printer, provided interindustry transaction flows have been written to magnetic tape during the current, or during a previous, execution of the program.

3. DATA INPUT FROM CARDS

Card 1

Cols 1-80: title

where:

title is an alphanumeric description of the job.

Card 2

Cols 1-5: N (right justified)
Cols 6-10: M (right justified)
Cols 11-15: N1 (right justified)
Cols 16-20: N2 (right justified)
Cols 21-25: N3 (right justified)
Cols 26-30: N4 (right justified)
Cols 31-35: N5 (right justified)
Cols 36-40: N6 (right justified)
Cols 41-80: blank
where:

N  is an integer, equal to the number of industries in the interindustry model, ($N \leq 110$).

M  is an integer, equal to the number of sectors in the intersectoral model, ($1 \leq M \leq N$).

N1 is an integer, equal to 1 if industries are to be aggregated according to similarity of input coefficients within sectors, and equal to 2 if industries are to be aggregated according to similarity of partially aggregated input coefficients within sectors.

N2 is an integer, equal to -1 if interindustry transactions are to be read from magnetic tape, equal to 1 if interindustry transactions are to be read from cards and written to magnetic tape, and equal to 0 if interindustry transactions are to be read from cards but not written to magnetic tape.

N3 is an integer which, if set to nonzero, suppresses the printing of interindustry transactions and industry gross outputs.

N4 is an integer which, if set to nonzero, suppresses the printing of input-output coefficients of the interindustry model.

N5 is an integer which, if set to nonzero, indicates that constraints are to be imposed upon the intersectoral model.

N6 is an integer which, if set to nonzero, allows the printing of intersector transactions, aggregated
input-output coefficients and sector gross outputs.

Note, however, that these listings will not be produced if N2 is equal to zero, even if N6 is nonzero.

Cards 3, 4 and 5

These cards are used to read interindustry transactions and industry gross outputs, if N2 is equal to 0 or 1. If these data are to be read from magnetic tape (and N2=-1), these cards should not appear in the input deck. Transactions are read, column by column, ignoring zero elements. Each column of data is preceded by its column number and a 20 character description. Elements within each column are preceded by their row numbers. The end of a column of data is indicated by an asterisk, as is the end of all columns of data.

Card 3

<table>
<thead>
<tr>
<th>Col.</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:</td>
<td>C or *</td>
</tr>
<tr>
<td>Cols 2-5:</td>
<td>Q (right justified)</td>
</tr>
<tr>
<td>Cols 6-10:</td>
<td>blank</td>
</tr>
<tr>
<td>Cols 11-30:</td>
<td>Name</td>
</tr>
<tr>
<td>Cols 31-80:</td>
<td>blank</td>
</tr>
</tbody>
</table>

where:

The character "C" indicates that a new column of transactions is to be read. The character "*" indicates that all columns of transactions have been read.

Q is an integer, equal to the column number of the set of transactions punched on the following set of cards of type 4.
Name is an alphanumeric description of the industry corresponding to column Q.

<table>
<thead>
<tr>
<th>Card 4</th>
<th>Col.</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1: R</td>
<td>or *</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2-5:</td>
<td>P</td>
<td>(right justified)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6:</td>
<td>blank</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7-13:</td>
<td>X(P,Q)</td>
<td>(right justified)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14-15:</td>
<td>blank</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16:</td>
<td>R or *</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>17-20:</td>
<td>P</td>
<td>(right justified)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>21:</td>
<td>blank</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>22-28:</td>
<td>X(P,Q)</td>
<td>(right justified)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>29-30:</td>
<td>blank</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>31:</td>
<td>R or *</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>32-35:</td>
<td>P</td>
<td>(right justified)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>36:</td>
<td>blank</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>37-43:</td>
<td>X(P,Q)</td>
<td>(right justified)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>44-45:</td>
<td>blank</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>46:</td>
<td>R or *</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>47-50:</td>
<td>P</td>
<td>(right justified)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>51:</td>
<td>blank</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>52-58:</td>
<td>X(P,Q)</td>
<td>(right justified)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>59-60:</td>
<td>blank</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>61:</td>
<td>R or *</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>62-65:</td>
<td>P</td>
<td>(right justified)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>66:</td>
<td>blank</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>67-73:</td>
<td>X(P,Q)</td>
<td>(right justified)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>74-80:</td>
<td>blank</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
where:

The character "R" indicates that a row element within column Q (as specified on the most recent card of type 3) is to be read. The character "*" indicates that all elements within column Q have been read.

P is an integer, equal to the row number of the next transaction which is to be read.

X(P,Q) is a real number, equal to the transaction in row P and column Q.

Card 4 is repeated as many times as are necessary to accommodate all nonzero transactions in a given column. The last item on the last of these cards must be an asterisk. Card 3 and the accompanying sets of cards of type 4 are repeated as many times as are required to accommodate all columns of transactions. The last card of type 3 must have an asterisk punched in column 1 and has no accompanying type 4 cards.

Card 5

Cols 1-15: X_i (right justified)
Cols 16-30: X_i (right justified)
Cols 31-45: X_i (right justified)
Cols 46-60: X_i (right justified)
Cols 61-75: X_i (right justified)
Cols 76-80: blank

where:

X_i is the gross output of industry i, (i=1,2,...,N).
Card 5 is repeated as many times as are required to accommodate all industry gross outputs.

Cards 6, 7, 8 and 9

These cards are used to place constraints upon the intersectoral model, if N5 is nonzero. If no constraints are to be imposed (and N5=0) these cards should not appear in the input deck. Each card of type 6 lists a set of industries, none of which may be aggregated with the other. Cards of type 7 specify industries which are to remain isolated in the intersectoral model. Each card of type 8 lists a set of industries, between which aggregation may take place. However, any industry listed on a card of type 8 may not be aggregated with any other industry which does not appear on the same card. Each card of type 9 specifies a set of industries, all of which must be aggregated into the same sector.

Card 6  
Cols 1-5: \( I_i \) or 999 (right justified)  
Cols 6-10: \( I_i \) (right justified)  
Cols 11-15: \( I_i \) (right justified)  

etc.

Cols 76-80: \( I_i \) (right justified)

where:

\( I_i \) is an integer, equal to the column number of any industry, which may not be aggregated with other industries on the same card.
Card 6 is repeated as many times as are required to specify all groups of industries, within which aggregation may not take place. Note that a maximum of sixteen industries may be listed on any card of type 6. However, a group, which is larger than sixteen, can be specified using a number of cards of type 6. For example, a group consisting of industries 1 to 17 would be specified by punching the following sets of industries on three cards of type 6:

\{17, 16, \ldots, 3, 2\} \\
\{17, 16, \ldots, 3, 1\} \\
\{2, 1\}

The last of these cards, of type 6, should contain the number 999 punched in columns 3 to 5 inclusive. If the user does not wish to specify any groups of industries, within which aggregation may not take place, then a single card of type 6, with 999 punched in columns 3 to 5, should be included in the input deck.

Card 7

Cols 1-5: $I_i$ or 999 (right justified) \\
Cols 6-10: $I_i$ or 999 (right justified) \\
Cols 11-15: $I_i$ or 999 (right justified) \\

\[ \vdots \]

Cols 76-80: $I_i$ or 999 (right justified)

where:

$I_i$ is an integer, equal to the column number of any industry, which is to remain isolated in the
intersectoral model.

Card 7 is repeated as many times as are required to accommodate all industries which are to remain isolated. The last item on the last card of type 7 must be the integer 999. If the user does not wish to specify any such industries, then a single card of type 7, with 999 punched in columns 3 to 5, should be included in the input deck.

Card 8

| Cols 1-5: | I_1 or 999 (right justified) |
| Cols 6-10: | I_1 (right justified) |
| Cols 11-15: | I_1 (right justified) |

etc.

| Cols 76-80: | I_1 (right justified) |

where:

I_1 is an integer, equal to the column number of any industry which may be aggregated with other industries appearing on the same card.

Card 8 is repeated as many times as are required to accommodate all groups of industries, to which aggregation is limited. Note that the maximum number of industries, which may be listed on any card, is sixteen. However, a group, which is larger than sixteen, can be specified using a number of cards of type 8. For example, a group consisting of industries 1 to 18 would be specified by punching the following sets of industries on
four cards of type 8:

\[\{18, 17, \ldots, 4, 3\}\]
\[\{18, 17, \ldots, 4, 2\}\]
\[\{18, 17, \ldots, 4, 1\}\]
\[\{3, 2, 1\}\]

The last of these cards, of type 8, must contain the integer 999 punched in columns 3 to 5. If the user does not wish to constrain the intersectoral model in this way, then a single card of type 8, with 999 punched in columns 3 to 5, should be included in the input deck.

Card 9

<table>
<thead>
<tr>
<th>Cols 1-5:</th>
<th>(I_i) or 999</th>
<th>(right justified)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cols 6-10:</td>
<td>(I_i)</td>
<td>(right justified)</td>
</tr>
<tr>
<td>Cols 11-15:</td>
<td>(I_i)</td>
<td>(right justified)</td>
</tr>
</tbody>
</table>

... etc...

| Cols 76-80: | \(I_i\) | (right justified) |

where:

\(I_i\) is an integer, equal to the column number of any industry, which must be aggregated with other industries appearing on the same card.

Card 9 is repeated as many times as are required to specify all sets of industries, which must be aggregated into the same sector. Note that the maximum number of industries per card is sixteen. However, a group of more than sixteen industries may be aggregated into the same sector, using a number of cards of
type 9. For example, industries 1 to 19 may be aggregated into a single sector, by punching the following sets of industries onto five cards of type 9:

\[
\begin{align*}
&\{19, 18, \ldots, 5, 4\} \\
&\{19, 18, \ldots, 5, 3\} \\
&\{19, 18, \ldots, 5, 2\} \\
&\{19, 18, \ldots, 5, 1\} \\
&\{4, 3, 2, 1\}
\end{align*}
\]

The last of these cards, of type 9, must contain the integer 999 punched in columns 3 to 5. If the user does not wish to constrain the intersectoral model in this way, a single card of type 9, with 999 punched in columns 3 to 5, should be included in the input deck.

Users should note that execution time will be minimized if cards 6, 7, 8 and 9 list industry numbers in descending order, as has been done in the examples given above.

4. OUTPUT TO THE LINE PRINTER

(a) The number of industries, \(N\), the number of sectors, \(M\), and the values of the flags, \(N_1, N_2, N_3, N_4, N_5\) and \(N_6\), as read from card 2.

(b) Interindustry transactions and industry gross outputs, if \(N_3 = 0\). Note that if \(N_3 \neq 0\) this output is suppressed.

(c) Input-output coefficients of the interindustry model, if \(N_4 = 0\). Note that if \(N_4 \neq 0\) this output is suppressed.
(d) The constraints imposed upon the intersectoral model, as read from cards 6, 7, 8 and 9, if N5≠0.

(e) If aggregation is performed on the basis of similarity of input coefficients, the following are printed following each merger:

(i) The stage number.
(ii) The number of sectors in the model.
(iii) The change in the total, within sector, sum of squared deviations of input coefficients from their means, that is, the "change in error s.s."
(iv) The total, within sector, sum of squared deviations of input coefficients from their means, that is, the "total error s.s."
(v) The column numbers of industries in the most recently formed sector.

When all mergers are completed, the composition of each sector is printed.

(f) If aggregation is performed on the basis of similarity of partially aggregated input coefficients, the following are printed following each merger:

(i) The stage number.
(ii) The number of sectors in the model.
(iii) The total, within sector, sum of squared deviations of partially aggregated input coefficients from their means, that is, the "error s.s."
(iv) The "error s.s." per sector.
(v) The column numbers of industries in the most recently formed sector.

When all mergers are completed, the composition of each sector is printed.

(g) Intersector transactions, aggregated input-output coefficients and sector gross outputs are printed, if $N_6 \neq 0$. Note that if $N_6 = 0$, this output is suppressed.
FILE 5=INPUT,UNIT=READER
FILE 6=OUTPUT,UNIT=PRINTER
FILE 7=CLUSTER/SAVEDATA,UNIT=DISK

SUBROUTINE TO WRITE TRANSACTIONS & INPUT OUTPUT TABLES

KEY=1 (TRANSACTION FLOWS)
KEY=2 (INPUT-OUTPUT COEFFICIENTS)
KEY=3 (AGREGATED TRANSACTION FLOWS)
KEY=4 (AGREGATED INPUT-OUTPUT COEFFICIENTS)

SUBROUTINE PRINT
INTEGER CK,LP,NN,MS,KEY,SDM(5900),GROUP(109),L(109)
COMMON CK,LP,NN,MS,KEY,SDM,GROUP,L

1

WRITECLP,1001)(TITLF(I),I=1,16)
GO TO (15,25,16,26),KEY

WRITE(LP,1002)
GO TO 17
WRITE(LP,1004)(I,J=1,NN)
DO 30 1-1,M
WRITE(LP,1005)(NAME(I,J),J=1,4),(DATA(I,J),J=JJ,NN)

CONIINUE
GO TO 35

WRITE(LP,1003)
GO TO 27
WRITE(LP,1009)
DO 26 1-1,NN
WRITE(LP,1006)(NAME(I,J),J=1,4),(DATA(I,J),J=JJ,NN)

CONTINUE

WRITE(LP,1007)(DATA(I,NN),I=JJ,NN)
RETURN

END
SUBROUTINE TO SORT INDUSTRIES INTO DESCENDING ORDER

SUBROUTINE SORT(O,NO)
INTEGER T(16),P,N0
DO 20 J=1,15
P=O(J)
IF(P.EQ.0)GO TO 25
DO 10 I=J+1,16 IF(O(I).LE.P)GO TO 10
P=O(I)
K=J
CONTINUE
O(K)=O(J)
CONTINUE
NO=16 IF(O(16).LE.0)NO=15
RETURN
20 NO=J-1
RETURN
END

SUBROUTINE TO IMPOSE CONSTRAINTS ON AGGREGATION BY SETTING ELEMENTS OF SIMILARITY MATRIX TO INFINITY OR ZERO

SUBROUTINE CONANG
INTEGER CN,LP,N,M,KP,NS,KEY,SIM(5900),GROUP(109),L(109)
REAL TITLE(C124,4),DATA(110,125),SIM(5900)
COMMON CN,LP,N,M,NS,KEY,SIM,COMMON
WRITE(10,1000)
READ(CF,1001)(O(I),I=1,16)
IF(O(1).EQ.999)GO TO 15
CALL SORT2000
DO 10 J=1,NO-1
IF(O(J).EQ.999)GO TO 120
JJ=(J+1)*(J+2)/2
DO 10 I=J+1,NO
JJ=(I-1)*I/2+P
SIM(JJ)=999999
CONTINUE
GO TO 5
10
READ INDUSTRIES WHICH ARE TO REMAIN ISOLATED IN INTERSECTORAL MODEL
READ(CF,1001)(O(I),I=1,16)
WRITE(10,2003)(O(I),I=1,16)
DO 35 J=1,16
P=O(J)
IF(P.EQ.999)GO TO 40
IF(P.EQ.1)IF(J,J,I)GO TO 120
II=(J+1)*(J+1)/2
DO 30 I=J+1,P
JJ=II+I
SIM(JJ)=999999
CONTINUE
IF(P.EQ.999)GO TO 35
DO 30 I=J+1,P
JJ=II+I
SIM(JJ)=999999
CONTINUE
GO TO 15
READ GROUPS OF INDUSTRIES WITHIN WHICH AGGREGATION MAY TAKE PLACE BUT
NOT AGGREGATION MAY NOT TAKE PLACE

READ (CR, 1001) (O(I), I=1, 16)
WRITE (LP, 1004) (O(I), I=1, 16)
IF (O(I).EQ.0.999) GO TO 50
CALL READ (Q, N)
DO 45 J=1, N-1
IF (O(J).GT.0) GO TO 120
I=O(J)-1*(O(J)-2)/2
DO 45 J=1, N-1
45 CONTINUE
J=1
GO TO 40
50 IF (J.K.EQ.0) GO TO 70
J=O(N-1)*(N-2)/2*(N-1)
DO 55 J=1, N-1
IF (N(J).EQ.0) GO TO 120
55 CONTINUE
GO TO 70

READ GROUPS OF INDUSTRIES WHICH MUST BE AGGREGATED

WRITE (LP, 1005) (O(I), I=1, 16)
IF (Q(I).EQ.0.999) RETURN
KEY=1
DO 75 J=1, N-1
IF (O(J).GT.0) GO TO 120
I=O(J)-1*(O(J)-2)/2
DO 75 J=1, N-1
75 CONTINUE
GO TO 70

WRITE (LP, 1006)
RETURN

1000 FORMAT (1H1, 10X, 'CONSTRAINTS ON THE INTERSECTORAL MODEL',/)
1001 FORMAT (1H1, 10X, 'THE FOLLOWING INDUSTRIES MAY NOT BE AGGREGATED',2X,
1002 FORMAT (/2X, 'THE FOLLOWING INDUSTRIES MUST REMAIN ISOLATED',3X,
1003 FORMAT (/2X, 'AGGREGATION RESTRICTED TO INDUSTRIES',12X,16I5)
1004 FORMAT (/2X, 'THE FOLLOWING INDUSTRIES MUST BE AGGREGATED',5X,16I5)
1005 FORMAT (/2X, 'ERROR - ILLEGAL INDUSTRY NUMBER ON LAST CARD READ')
END
SUBROUTINE TO AGGREGATE INDUSTRIES BASED ON SIMILARITY OF INPUT COEFFICIENTS

INTEGER CR,LP,N,N5,KEY,SDM(5900),GROUP(109),L(109)

READ TITLE(16),NAME(124),DATA(116,125),SDM(5900)

COMMON CP,LP,N,M,N5,KEY,SDM,GROUP,L

SUBROUTINE CLUST1

SET UP SIMILARITY MATRIX USING SQUARED EUCLIDEAN DISTANCES

IMPOSE CONSTRAINTS ON AGGREGATION

INITIALIZE FUSION RECORDING MATRIX "GROUP".

FIND MOST SIMILAR INDUSTRIES IN EACH ROW OF SIMILARITY MATRIX

FIND MINIMUM OF ROW MINIMA "EPQ", CHECKING FOR PREVIOUS Merges.

******** MAIN LOOP

10 DO 10 I=2,N

30 IF(KF.Y.EQ.9)RETURN

10 WRITE(LP,1001)(TITLE(I),I=1,16)

20 DO 20 I=2,N

40 IF(LCI1.FO.01GO TO 40

50 CONTINUE

********** MAIN LOOP

40 DO 150 KE=2,N-N-1

50 CONTINUE

********** MAIN LOOP

150 CONTINUE

1001 FORMAT(26A10)

1000 FORMAT(26A10)

END
C UPDATE ERROR SUM OF SQUARES "ERSSQ"
ERSSQ=ERSSQ+EPQ/2.0

C UPDATE SIMILARITY MATRIX & ROW MINIMA, CHECKING FOR PREVIOUS MERGES
TOTAL(L(O)+L(P))
1 IF(L(K).EQ.1)GO TO 65
2 SMALL=999999
3 JJ=(P-1)*(P-2)/2
4 DO 60 K=1,P-1
5 IF(L(K).EQ.0)GO TO 60
6 JJ=JJ+K
7 SIM(I)=SIM(I)+(L(K)*L(P))*(TOTAL+L(K))
8 SIM(I)=SIM(I)+L(P)*EPQ/(TOTAL+L(K))
9 IF(SIM(I)-SMALL)55,55,60
10 SMALL=SIM(I)
11 CONTINUE
12 DO 65 JJ=(P-1)*(P-2)/2
13 IF(L(K).EQ.0)GO TO 65
14 JJ=JJ+K
15 IF(JJ(P)-110,110,70
16 J=(K-1)*(K-2)/2+P
17 DO 80 J=JJ+K
18 IF(SIM(T)-SMALL)95,95,100
19 MIN(K)=K
20 DO 95 J=JJ+K
21 IF(SIM(T)-SMALL)105,105,110
22 V=HMIN(J)
23 IF(V.HMIN(K).EQ.O)GO TO 95
24 IF(SIM(I)-V)GO TO 95
25 SIM(I)=SIM(I)+L(P)*EPQ/(TOTAL+L(K))
26 IF(SIM(I)-SMALL)10,10,115
27 JJ=(K-1)*(K-2)/2
28 JJ=JJ+K
29 DO 100 I=JJ+K
30 IF(SIM(T)-SMALL)115,115,120
31 MIN(K)=K
32 CONTINUE
33 DO 105 I=JJ+K
34 IF(SIM(T)-SMALL)120,120,125
35 GROUP(K)=GROUP(J)
36 K=JJ+I
37 IF(K.LE.I)GO TO 125
38 J=JJ+I
39 DO 130 I=JJ+I,J+L(P)
40 TEMP(I)-GROUP(I)
41 CONTINUE
42 WRITE(LP,1103)(GROUP(I),I=J,K)
43 J=JJ
44 GROUP(K)=GROUP(J)
45 K=JJ+I
46 IF(K.LE.I)GO TO 135
47 J=JJ+I
48 DO 140 I=JJ+I,J+L(P)
49 GROUP(I)=TEMP(I+I)
50 CONTINUE
51 WRITE(LP,1002)(K,(N=KK+11),EPQ,ERSSQ)
52 30 IF(SIM(I)-V)GO TO 95
53 SIM(I)=SIM(I)+L(P)*EPQ/(TOTAL+L(K))
54 IF(SIM(I)-SMALL)10,10,115
55 JJ=(P-1)*(P-2)/2
56 DO 95 J=JJ+K
57 IF(SIM(T)-SMALL)10,10,115
58 MIN(K)=K
59 DO 105 I=JJ+K
60 IF(SIM(T)-SMALL)115,115,120
61 MIN(K)=K
62 CONTINUE
63 DO 105 I=JJ+K
UPDATE NUMBER OF INDUSTRIES IN EACH SECTOR  
L(Q)=TOTAL  
L(P)=P  
CONTINUE  
WRITE(LP,1005)  
II=0  
JJ=0  
DO 160 J=1,N  
IF(L(J).EQ.0)GO TO 160  
II=II+L(I)  
M=M+1  
WRITE(LP,1004)M  
WRITE(LP,1003)(GROUP(J),J=JJ,II)  
JJ=II+1  
CONTINUE  
RETURN  
FORMAT(1H1,20X,'TITLE(1),I=1,16)  
FORMAT(1H1,20X,'TOTAL ERROR S.S.=',100,5)  
FORMAT(1H,20X,'SECTOR',I4)  
FORMAT(1H1,20X,'FINAL COMPOSITION OF SECTORS')  
END  
SUBROUTINE TO AGGREGATE INDUSTRIES BASED ON SIMILARITY OF PARTIALLY AGGREGATED INPUT COEFFICIENTS  
SUBROUTINE CLUST2  
INTEGER CP,L,LP,N,N5,KEY,SDM(5900),GROUP(109),L(109)  
REAL TITLE(16),NAME(1?4),DATA(110,125),SIM(5900)  
COMMON RP,L5,LP5,NAME,DATA,SDM,GROUP,L  
SET UP SIMILARITY MATRIX OF ERROR SUMS OF SQUARES  
DO 10 I=2,N  
II=(I-I)*(I-2)/2  
DO 10 J=1,II-1  
A=0.0  
B=0.0  
JJ=I+J  
DO 5 K=1,N  
A=A+DATA(K,I)*DATA(K,J)  
B=B+DATA(K,I)*DATA(K,J)  
5 CONTINUE  
SIM(JJ)=A/2.0-B*(DATA(J,J)-DATA(J,I))*(DATA(I,J)-DATA(I,I))  
CONTINUE  
IF(N5)15,20,15  
CALL CUNAG  
IF(KEY,EQ.9)RETURN  
SET UP MATRIX OF MEANS  
DO 25 I=1,N  
MEAN(I,J)=DATA(I,J)  
25 CONTINUE  
CALL CUNAG  
CONTINUE  
CALL CUNAG  
WRITE(LP,1001)(TITLE(1),I=1,16)
**MAIN LOOP**

```fortran
DO 150 KK=2,N+1

**FIND THE MOST SIMILAR PAIR OF SECTORS, "R" AND "S", RECORD ERROR SUM OF SQUARES "ERS", CHECK FOR PREVIOUS MERGES**

```if(key,eq,0)go to 34
``` 43  
```do 31 i=2,N
```
```i=(i-1)*(i-2)/2
```
```do 31 j=i,i-1
```
```jd=i+j
```
```if(gdm(jd),ne,0)go to 32
```
```31 continue
```
```key=0
```
```go to 34
```
```32 p=1
```
```s=1
```
```ers=sim(jd)
```
```do 33 j=1,i-1
```
```sdm(jd)=0
```
```continue
```
```go to 41
```
```35 ers=sim(jd)
```
```r=i
```
```s=j
```
```go to 40
```
```40 continue
```
```if(ers,eq,999999)go to 155
```
```do 45 i=1,N
```
```if(l(i),eq,0)go to 40
```
```45 continue
```
```do 40 j=1,i-1
```
```jd=i+j
```
```do 45 1=j+1,r=1
```
```temp(i-j)=group(i)
```
```75 continue
```
```go to 65
```
```80 group(k)=group(j)
```
```if(j,le,11)go to 65
```
```90 continue
```
```70 continue
```
```75 write(ip,1002)kk,(n-kk+1),ers,erssq
```
```10 j=11-11+1
```
```11 k=11-11+(1)
```
```12 write(ip,1003)(group(t),i=j,k)
```
```22 update elements of similarity matrix not involving rows and columns "R" AND "S", checking for previous merges
```
```total=lp+ll(r)
```
```26 xL=2.0*L(R)*L(S)/TOTAL
```
DO 105 I=2,N
IP(I,EQ.0)GO TO 105
II=(I-1)*(I-2)/2
IF(I.EQ.0)GO TO 95
DO 90 =II,1,-1
IP(I,J),EQ.0,0)GO TO 90
IF(J.EQ.S)GO TO 105 II=(I-1)*(J-2)/2+1
GO TO 90
95 DO 100 J=1,1-1
SIM(JJ)=9999999
90 CONTINUE
GO TO 105
105 CONTINUE

C UPDATE ELEMENTS OF SIMILARITY MATRIX INVOLVING ROWS AND COLUMNS "N"
C AND "S" CHECKING FOR PREVIOUS MERGES
C UPDATE PARTIALLY AGGREGATED INPUT COEFFICIENTS AND THEIR MEANS
DO 110 I=1,N
DATA(S,1)*DATA(R,1)
MEAN(S,1)=MEAN(S,1)+MEAN(R,1)
MEAN(R,1)=0.0
110 CONTINUE
DO 145 J=1,N
IP(I,J),EQ.0,0)GO TO 145
IF(J.EQ.0)GO TO 25,145,125
120 JJ=II+J
GO TO 130
125 JJ=(J-1)*(J-2)/2+S
130 IF(S(MJ),EQ.999999)GO TO 145
XL=TOTAL*L(J)/(TOTAL+L(J))
A=0.0
B=0.0
YJ=0.0
1 DO 135 I=1,N
YJ=YL+DATA(S,1)*DATA(R,1)
A=A+YJ*MEAN(S,1)*MEAN(R,1)
135 CONTINUE
SIM(JJ)=ERS+2*YL+2*A+XL*B+2*X*(MEAN(S,1)=MEAN(S,1))
145 CONTINUE
EAB=ERS
150 CONTINUE
155 WRITE(LP,1005)
160 CONTINUE
160 RETURN
SUBROUTINE TO CALCULATE MATRIX OF AGGREGATED TRANSACTIONS FLOWS AND MATRIX OF AGGREGATED TECHNICAL COEFFICIENTS

INTEGER CR, LP, N, N5, N6, KEY, SDM(5900), GROUP(109), L(109)
REAL TITLE(16), NAME(124, 4), DATA(110, 125), SIM(5900)
COMMON CR, LP, N, N5, N6, KEY, SDM, GROUP, L

READ ORIGINAL MATRIX OF INTERINDUSTRY TRANSACTION FLOWS FROM TAPE
DO 10 J=1,125
READ(7)(DATA(I, J), I=1,110)
CONTINUE

M=0
II=0
JJ=1
DO 25 I=1, N
IF(I(II).EQ.0) GO TO 25
II=II+L(I)
M=M+1
25 JJ=II+1
CONTINUE

II=0
JJ=1
DO 40 I=1, N
IF(I(II).EQ.0) GO TO 40
II=II+L(I)
JJ=II+1
40 CONTINUE

LJ=0
II=0
JJ=1
DO 50 J=1, M
IF(J(JJ).EQ.0) GO TO 50
II=II+L(J)
JJ=JJ+1
50 CONTINUE

LI=0
II=0
JJ=1
DO 60 I=1, M
IF(I(I).EQ.0) GO TO 60
II=II+I
LI=LI+1
60 CONTINUE

LJ=0
II=0
JJ=1
DO 70 K=1, N
K=GROUP(K)
LJ=LJ+1
70 CONTINUE

DATA(J, K)=TEMP
CONTINUE
JJ=JJ+1
DO 300 J=1, N
DO 30 K=1, M
DATA(J, K)=TEMP
K=GROUP(K)
30 CONTINUE
300 CONTINUE

PRINT MATRIX OF AGGREGATED TRANSACTION FLOWS
DO 55 J=1, M
DATA(J, I)=DATA(J, HH)
NAMP(I, 1)="SECTOR"
NAME(I, 2)="R"
NAME(I, 3)=" "
55 CONTINUE
N=M
KEY=3
CALL PRINT
CALCULATE MATRIX OF AGGREGATED TECHNICAL COEFFICIENTS AND PRINT

DO 60 J=1,N
   DUMMY=DATA(J,NN)
   DO 60 I=1,N
      DATA(I,J)=DATA(I,J)/DUMMY
      KEY=4
      CALL PRINT
   CONTINUE
60   RETURN

PROGRAM TO AGGREGATE INDUSTRIES OF AN INTERINDUSTRY TABLE

PROGRAMMER: JOAN RODGERS
- DEPT AGRICULTURAL ECONOMICS & MARKETING,
- LINCOLN COLLEGE.

INSTALLATION: BURNOUGHS 6700, UNIVERSITY OF CANTERBURY

INTEGER CR,LP,N,M,N5,KEY,SON(5900),GROUP(109),I(109)
REAL TITLE(16),NAME(125),DATA(110,125),SIM(5900)
COMMON CR,IP,LP,N,M,KEY,SON,GROUP,L
    TITLE,NAME,DATA,SIM

CR=5
LP=6

READ PROBLEM TITLE CARD
READ(CR,1001)(TITLE(I),I=1,16)

READ PROBLEM DESCRIPTOR CARD
N = NO OF INDUSTRIES IN INTERINDUSTRY MODEL
M = NO OF SECTORS IN INTERSECTOR MODEL
N1 = CLUSTERING ALGORITHM (1=INPUT COEFFICIENTS,
   2=PARTIALLY AGGREGATED INPUT COEFFICIENTS)
N2 = READ DATA FROM TAPE, 0 (READ DATA FROM CARDS & WRITE
   TO TAPE), 0 (READ DATA FROM CARDS BUT DO NOT WRITE TO TAPE)
N3 = NON ZERO TO SUPPORT PRINTING OF INTERINDUSTRY TABLE
N4 = NON ZERO TO SUPPORT PRINTING OF INPUT-OUTPUT TABLE
N5 = NON ZERO IF CONSTRAINED AGGREGATION IS REQUIRED
N6 = NON ZERO TO CALCULATE AND PRINT MATRICES OF AGGREGATED
   TRANSACTION FLOWS AND AGGREGATED TECHNICAL COEFFICIENTS
READ(CR,1002)(N,N1,N2,N3,N4,N5,N6)

WRITE(LP,1003)N,N1,N2,N3,N4,N5,N6
NN=N+1

IF(N)400,5,5

READ INTERINDUSTRY TABLE "DATA" FROM CARDS COLUMN BY COLUMN &
IGNORING ZERO ELEMENTS,
EACH COLUMN IS PRECEDED BY A COLUMN NO "O" & DESCRIPTOR "NAME",
ROW ELEMENTS WITHIN EACH COLUMN ARE PRECEDED BY A ROW NO "P",
THE END OF A COLUMN IS INDICATED BY AN "**",
THE END OF ALL COLUMNS IS INDICATED BY AN **
READ(CR,1004)(DUMMY,O,CARD(0,1),I=1,4)
   IF(DUMMY.EQ.'**') GO TO 20
   IF(CARD.GT.N.OR.O.LT.1) GO TO 95
   READ(CR,1005)(ROW(I),ROW(I),TRANS(1),I=1,5)
   DO IS=1,5
      IF(DUM(I).EQ.'**') GO TO 5
      IF(ROW(I).GT.N.OR.ROW(I).LT.1) GO TO 100
      READ(10,15)(DUMMY,0,(NAME(0,1),I=1,4)
      DATA(P,O)=TRANS(I)
   CONTINUE
   GO TO 10
READ INDUSTRY OUTPUTS
READ(CR,1006)(DATA(I,NN),I=1,N)
IF(N2)55,55,25
WRITE INTERINDUSTRY TABLE TO TAPE
DO 30 J=1,125
WRITE(7)(DATA(I,J),I=1,110)
CONTINUE
DO 35 J=1,4
WRITE(7)(NAME(I,J),I=1,124)
CONTINUE
GO TO 55
READ INTERINDUSTRY TABLE FROM TAPE
READ(7)(DATA(I,J),I=1,110)
CONTINUE
DO 50 J=1,4
READ(7)(NAME(I,J),I=1,124)
CONTINUE
IF(N3)65,60,65
PRINT INTERINDUSTRY TABLE
PRINT INPUT-OUTPUT MATRIX
DO 70 J=1,N
DUMMY=DATA(J,NN)
DO 70 I=1,N
DATA(I,J)=DATA(J,J)/DUMMY
CONTINUE
IF(N4)80,75,80
PRINT INPUT-OUTPUT MATRIX
KEY=2
CALL PRINT
CALCULATE INPUT-OUTPUT MATRIX
DO 70 J=1,N
DUMMY=DATA(J,NN)
DO 70 I=1,N
DATA(I,J)=DATA(I,J)/DUMMY
CONTINUE
IF(N4)80,75,80
PRINT SIMILARITY MATRIX AND AGGREGATE INDUSTRIES INTO SECTORS
KEY=0
GO TO 92
CALL CLUS1
GO TO 92
CALL CLUS2
IF(KEY.EQ.9)GO TO 105
IF(N6.NE.0.AND.N2.NE.0)CALL MATAGG
GO TO 105
WRITE ERROR MESSAGES
WRITE(LP,1007)0,N
GO TO 105
WRITE(LP,1008)HOW(I),N
ENDFILE 7
CALL EXIT
1001 FORMAT(16X)
1002 FORMAT(1G15)
1003 FORMAT(1H1,2X,'NO OF INDUSTRIES=',I5,5X,'NO OF SECTORS=',I5,5X,
1        'CLUSTERING ALGORITHM NO=',I5,2X,'N2=',I5,5X,'N3=',I5,5X,
1        'N4=',I5,5X,'N6=',I5,5X)
1004 FORMAT(A14,5X,4A5)
1005 FORMAT(5A1,14,1X,F7,0,2X))
1006 FORMAT(5F15.0)
1007 FORMAT(2X,'ROW NO ('',10,') EXCEEDS NO OF COLS ('',13,') IN TABLE')
1008 FORMAT(2X,'ROW NO ('',10,') EXCEEDS NO OF ROWS ('',13,') IN TABLE')
END

199