A REASSESSMENT OF THE HYBRID APPROACH TO THE CONSTRUCTION OF REGIONAL INPUT-OUTPUT TABLES

by

STEVEN JAMES BRAND

A thesis submitted to the University of Plymouth

in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

Plymouth Business School

Faculty of Social Science

January 1998

Abstract

This thesis lays the foundation for the creation of an economic database for the counties of Devon and Cornwall in the form of a regional input-output table. The thesis reconsiders the popular hybrid approach to the construction of such tables. In particular, the nonsurvey-to-survey ordering of procedure is questioned. The thesis attempts to restore a more logical preference-order which begins with first-best (survey) estimation methods and extends to second-best (survey-based-nonsurvey) methods. The third-best methods of estimation (pure nonsurvey *i.e.* location quotient) are excluded from the process altogether. The thesis is largely concerned with the development of the second-best method.

The second-best method is derived from an empirical analysis of the nature of nonsurvey estimation error. The analysis is able to reject the Stevens *et al.* (1983) hypothesis that differences in regional and national production functions are insignificant. Nevertheless, the strategy of developing 'trade-only' nonsurvey estimation methods is found to be valid since, whilst the error associated with regional trade misspecification can be reduced within a broad method of estimation, the error attributable to the misspecification of regional production functions remains largely intractable to such an approach. Survey resources *must* therefore be devoted to the specification of these functions.

The second-best methodology extends the Stevens *et al.* (1983) by deriving equations that specify the RAS algorithm and local expenditure propensities for households from empirical data for Scotland. These equations have general application within the new hybrid methodology.

By restoring a more logical preference-order of approach to estimating hybrid regional input-output tables, emphasis is placed on the analytical strength afforded by a good data set, and not on the analytical 'strength' of magic-box mathematics. This should encourage the regional input-output table to be implemented as an *evolving* local economic database, which will improve the general quality of regional analysis and, in the long-run, offer cost-savings in data collection and collation.

Table of Contents

page

Table of Contents *(continued)*

Table of Contents ***(continued)*** *page*

Table of Contents *(continued)* *page*

Table of Contents ***(continued)*** *page*

List of Tables

List of Figures

page

List of Figures *(continued)*

Acknowledgements

I would like to thank my research supervisor, Don King for his time, patience, guidance and encouragement. Thanks go to Professor Peter Gripaios, Dr. Paul Bishop, and Dr. Pamela Whisker for their time and helpful comments.

A special thanks goes the Stephen Hill, Director of the Welsh Economy Research Unit at Cardiff Business School for his support in the latter stage of the programme. Thanks also to Dr. Annette Roberts of WERU.

Thanks go to Dr. Lean, Dr. Williams, Dr. Tucker, and Dave for providing light entertainment.

Finally, a very big thanks goes to my mum.

AUTHOR'S DECLARATION

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award.

Relevant Publications

Brand, S., (1996) Heterogeneity in input-output tables and the need for a new approach

Brand, S., (1997) On the Appropriate Use of Location Quotients in Generating Regional Input-Output Tables: a Comment, *Regional Studies*, **31**, pp791-4

Brand, S., Hill, S., Roberts, A., (1998) *Input-Output Tables for Wales, 1995*, University of Wales Press, Cardiff

Brand, S., (1998) Supply Chains, Material Linkage and Regional Development: A Comment, *Urban Studies*, **35**, forthcoming April

Signed S. Brand

Dated 23 · 1 · 98

CHAPTER 1

INTRODUCTION AND OVERVIEW

1.0 Introduction and Overview

The initial objective of this project was to improve the economic data set for the UK counties of Devon and Cornwall. Counties are really the second tier of UK regional geography, representing on average 1-2% of national GDP. Devon and Cornwall together represent just over 2%. The eleven 'standard regions' form the first tier of regional geography. A standard region averages around 10% of UK GDP. The fact that counties are relatively small and numerous means that the availability and coverage of economic variables from UK published sources is relatively poor. The county statistics that *are* produced are seemingly designed to provide a very broad indication of relative economic prosperity and no more. Disaggregations, *i.e.* into industrial sectors, are rare, as are annual series. So, for example, estimates of county GDP are made biennially, with gross value added in manufacturing being the only form of sectoral disaggregation. The ONS annual publication *Regional Trends* provides the main outlet for such statistics. The only relatively complete series is employees in employment, which is taken from the biennial Census of Employment and is available from the National Online Manpower Information Service (NOMIS). These data are obtainable at the 4-digit level of Standard Industrial Classification and therefore provide some idea of a county's industrial profile. Figures on self-employment are only available from the Census of Population, which is made once every ten years. A certain degree of disaggregation within variables such as gross output, value added, wages, stocks, net capital expenditure, and material input purchases is available for manufacturing sectors from the Annual Census of Production. However, these data come at cost and figures are suppressed if they are considered disclosive -

which, at the county level is a notable hazard. Furthermore, the fact that counties are relatively small means that the errors associated with the construction of statistics from national surveys - in particular sampling error - are a particularly significant concern. However, whilst it could not be denied that the official statistics relating to Devon and Cornwall could be improved in some way - for example 'plugging gaps' in biennial series - it seemed doubtful whether this would form a suitably challenging exercise.

The local Economic Development Agencies (EDA) in the two counties - for example the Training and Enterprise Council (TEC), county and district councils - are, naturally, the principal users of county economic statistics. Local information is required in strategic planning and project appraisal to support funding applications *etceteras*. Given the paucity of official statistics, it is hardly surprising to find that the Economic Development Agencies (EDA) commission the collection of local economic information. This work generally takes the form of 'sector studies' where data is compiled on those sections of the economy considered locally important for example, Gripaios *et. al.* (1991) on exporting sectors, DCDI (1996) on inward investment. Similar information is compiled by those institutions involved in academic research - see for example Bishop (1996) on the defence sector.

One of the most common reasons for generating data of this nature is to make an assessment of the relative importance of the sector in question in terms of its contribution to local employment, incomes, and value added. As the next chapter will illustrate, there are various degrees of sophistication by which this objective can be achieved. Local sector based studies have, however, tended to remain at the lower end of this scale. This has nothing to do with any analyst's lack of technical ability, it is because the higher levels of regional economic analysis require the backing of a detailed regional-specific database - which does not exist in official statistics, and no small, relatively *ad hoc* sector-based study can, individually, provide. But why can't these individual sector studies be brought together to form such a database? Apart from easing the 'data problem' and accessing higher levels of analytical methodology, the creation of such a framework would almost certainly improve the cost efficiency of future data collection. How many sector studies have been lost, used once and forgotten, or unnecessarily repeated over time? The net benefit of a more co-ordinated and rationalised approach to data collection in the counties would seem to be significant.

One approach which offers a coherent framework for the collation of detailed economic data, and at the same time offers a moderate degree of analytical sophistication is the regional input-output table. There has been one previous attempt to build an input-output table of Cornwall and that was for 1984 (Johns and Leat, 1986). The absence of updated tables indicates that the attempt failed to impose itself on the nature of Cornish data collection in subsequent years. However, this is perhaps unsurprising. Somewhat mysteriously, the authors were commissioned from the Scottish Agricultural College in Aberdeen - which is about as geographically distant from Cornwall as one can possibly get whilst remaining in the UK. The project incorporated very little 'sector study' data, generated even less original survey material, and instead chose to rely upon a set of standard procedures which scale down the national input-output table to 'regional dimensions'. In short, the study attempted to access higher levels of analytical sophistication without the backing of a good regional-specific database. And so by failing to emphasise the importance of county-specific data in the Cornish input-output table, the model was never recognised as a vehicle for steering the collection and assembly of local economic information. One would suspect however that, whilst the model was never updated, its analytical function probably remained in active use well beyond 1984.

The need for a coherent approach to local data collection and analysis has rarely been so great. Competition amongst areas to gain European and National Lottery funding of local projects has undoubtedly put pressure on the need to demonstrate accountability in funding applications. The significant increase in Foreign Direct Investment (FDI) that occurred in the UK during the 1980s and 1990s (Hill and Munday, 1994) has led to concerns over its impact upon, and role within, the regional economy (Turok, 1993; Roberts, 1996). The success of the Labour party in the 1997 UK general election represents a move to a more regionally-minded central administration. And of course, regional economic disparity remains a feature of the UK economy (see Johnston *et al.*, 1996; Martin, 1997 for recent evidence).

The research objective was therefore to pursue the regional input-output specification as a means to improving the economic database in Devon and Cornwall. A more complete justification for this choice is given in the next chapter. However, in the initial stages of the programme it was not clear whether the research would involve the construction of input-output tables for the counties in question, or whether a more abstract path, which sought instead to lay the foundations for the construction of tables at some future point, was more appropriate. In the event, the latter route was chosen, mainly because the resources required for the construction of respectable county input-output tables were beyond the scope of the project. A Johns and Leat-type study would be unlikely to have the desired impact on changing the method of data collection and analysis in the counties. The research objective was therefore to be achieved by improving the procedures by which regional input-output tables are constructed. This improvement could be both conceptual: improving the way the input-output model is perceived; and practical: reducing the errors associated with model estimation.

The thesis is structured as follows. Chapter 2 provides an introduction to the input-output model and provides a more complete justification for its selection as a means to achieving the research objective. Chapter 3 reviews the various methods of generating regional input-output tables. Chapter 4 provides a critical assessment of these procedures and identifies the weaknesses in their foundations. This chapter provides the basis for the development of a new set of procedures which seek to 're-market' the regional input-output table, thereby improving the quality of the information contained within. Improvements in the procedures will, in part, be established by a process of empirical testing. Chapter 5 therefore considers the development of the procedures from an *a priori* perspective. This generates a set of testable hypotheses. Chapter 6 introduces the main tools and methods that will be used to carry out the analysis implied by Chapter 5. Chapter 7 provides a justification for the selection of the empirical data set and gives details of the transformations that were required prior to the analysis. Chapter 8 presents the results of the analytical exercise, and this forms the basis for the conclusions and recommendations for future research that are contained within Chapter 9.

CHAPTER 2

JUSTIFICATION FOR THE SELECTION OF THE REGIONAL INPUT-OUTPUT SPECIFICATION

2.0 Justification for the Selection of the Regional Input-Output Specification

2.1 Introduction

The intention of this chapter is to provide a justification for the selection of the regional input-output model as the preferred specification for any subsequent project in Devon and Cornwall. The chapter also serves as an introduction to the basic concepts and terminology that characterise input-output analysis. The strengths of the input-output specification as a framework for data collection and analysis are highlighted by considering some of the less sophisticated analytical alternatives. The weaknesses of the input-output model exist mainly within its analytical function. However, it is shown that a basic regional input-output data set can be developed to achieve a higher level of analytical sophistication. It should be noted that there are many applications and extensions of the basic input-output framework which are not explored here - the multi-regional model is a prime example. Miller and Blair (1985) and Richardson (1985) review some of the alternative angles.

2.2 The Basic Regional Input-Output Specification

The input-output model is formally associated with Leontief (1936), although as early as the 18th Century, the French economist Quesnay describes what are essentially input-output multipliers in his '*Tableau Economique*'. The general equilibrium framework of Walras (1874) provides a theoretical basis for the Leontief model.

2.21 The Transactions Tables

The feature that distinguishes the input-output model from any less sophisticated mode of economic analysis is its disaggregated representation of the economy. The economy is split into sectors which interact in terms of supply and demand. At the broadest level these relate to intermediate production, the rewards to the factors of production, and the demand for final products - a basic system of circular flow. Within each of these broad categories however, further distinction between sectors is normally made. So, for example, in the productive or intermediate sector of the economy, distinction is made between agriculture, mining, manufacture, and services. Labour income is usually separated from the gross return to the factors of production - import purchases and taxation are recorded in what are termed 'primary inputs'. Consumers, governments, investors form the bulk of the demand for final products - or 'final demand'. The sales of exports and the change in inventories are also classed as final demand. Hence the input-output table is a record of the *transactions* that took place between the identified sectors of the economy in the time period in question. This time period is normally a year. A representation of the *transactions table* is shown in Table 2.1 below.

Table 2.1 The Transactions Table

	Industry Purchases	Final Demands	Totals
Industry Sales			
	x	*f*	X
			(output)
Primary Inputs	*i*		I
Totals	X *(input)*	F	Σ

As the table shows, the columns of the matrix record purchases of goods and services - those by industry and final demanders. The rows of the matrix record sales of goods and services - those by industry, and, for example the 'sale' of labour services. The model therefore has a basic 'accounting feel' to it, with each purchase simultaneously represented as a sale. Note that the purchase of an industry's total inputs (column sums) equals the sum of its total outputs (rows)[1].

The transactions table is commonly referred to as the *use matrix*, because its columns show what each sector used in the period. The *combined use matrix* shows the purchase of all inputs irrespective of their origin. So, for example, if a UK industry purchased steel from both Sheffield and the Far East, the total value of the transaction would appear in the intermediate section (*x*) of the combined use matrix. The *domestic use matrix* however would provide distinction between these two purchases. The domestic use matrix records purchases made from national sources. Hence the purchase of steel from Sheffield would be recorded in the intermediate use matrix, the purchase from

[1] There are a number of alternative accounting systems. The symmetrical relationship between total input-output described here is the most simplistic. Chapter 7 considers a more complex representation.

the Far East would be recorded in primary inputs (*i*) as an import from overseas. The domestic transactions table shows these imports as a single row entry, even though each industry's import purchase will consist of a mixture of products. The *imports use matrix* however provides a disaggregation of this import purchase. One could of course produce a similar disaggregation showing the sectoral destination of exports.

In the regional input-output table, further distinction is normally made between the transactions that occur between local sectors and those that are exports and imports to/from other regions within the nation. Therefore, if the regional input-output table represented South Yorkshire, the purchase of steel from Sheffield would be classed within the *regional use matrix*. If the regional table represented Devon, the purchase of steel would be classed within primary input as an import from the 'rest of the UK'. The disaggregation of this import row would be referred to as the *regional imports use matrix*. This definition of terminology will be used throughout the thesis.

Therefore, the regional input-output table is essentially a collection of local information - a regional economic database which stands alone as a valuable source of reference. A range of useful economic statistics can be derived from its components with simple manipulation: GDP in each sector, annual average wage per employee in each sector, import propensities *etcetera*. However, the specification has a more sophisticated analytical function.

2.22 The Leontief Analytical Function

It should be clear from the preceding discussion of the basic accounting framework that *structural dependence* between the defined sectors of the economy is a defining feature of the input-output model. *Circularity* within the input-output system was also hinted at. These two concepts form the basis for the derivation of multipliers from the Leontief model.

The notion of the multiplier is most readily associated with the work of Keynes (1936). The calculation of the multiplier is an attempt to quantify the extent to which an exogenous increase in demand (the 'multiplicand') stimulates increases in demand within the endogenous economic system.

Developing upon Table 2.1, the input-output system can be represented as a system of linear equations

$$\begin{array}{ccccc} x_{11} + & \cdots & +x_{1j} & +f_1 & = X_1 \\ \vdots & & \vdots & \vdots & \vdots \\ x_{i1} + & \cdots & +x_{ij} & +f_i & = X_i \end{array} \qquad [2.1]$$

For each industry j, its input purchases x_{ij} can be represented as shares of total industry input - the column sum X_j:

$$a_{ij} = x_{ij} / X_j \qquad [2.2]$$

Each a_{ij} express the amount of i directly required to produce a single 'unit' of industry j's output. The a_{ij} are termed the direct requirement coefficients - which is often shortened to 'direct coefficients'. Together, the a_{ij} in any column describe the input composition of a single 'unit' of j - each j column of a_{ij} can therefore be interpreted as the jth industry's production function. Equation [2.2] can be rearranged so that:

$$x_{ij} = a_{ij} X_j \qquad [2.3]$$

which can be substituted into the system of equations represented by [2.1]:

$$\begin{matrix} a_{11}X_1 + & \cdots & +a_{1j}X_j & +f_1 & = X_1 \\ \vdots & & \vdots & \vdots & \vdots \\ a_{i1}X_1 + & \cdots & +a_{ij}X_j & +f_i & = X_i \end{matrix} \qquad [2.4]$$

The system of equations in [2.4] is more conveniently expressed in matrix algebra:

$$\mathbf{AX} + \mathbf{f} = \mathbf{X} \qquad [2.5]$$

where

A $[a_{ij}]$ is the matrix of direct coefficients;

X is the vector of gross industry outputs (=inputs);

f is the vector of final demands;

Equation [2.5] can be rearranged to express **X** in terms of **f**:

$$\mathbf{X} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{f} \qquad [2.6]$$

where

I is the identity matrix.

Thus for any exogenous change in demand (**f**) the effect upon output can be computed. The matrix $(\mathbf{I}\text{-}\mathbf{A})^{-1}$ is referred to as the *Leontief inverse solution*. Each element of the Leontief inverse shows the direct and indirect demands for the *i*th industry's output resulting from a unit increase in the *j*th industry's final demand. The row sums of the Leontief inverse therefore express the increase in the *i*th sector's output arising from a unit increase in demand for all *j* industries. The column sums of the matrix express the increase in the output of the economy arising from a unit increase in the *j*th sector's final demand. These column sums are the conventional *Type I* output multipliers.

The Leontief solution can be more clearly understood by representing [2.6] as a power series expansion, that is:

$$(\mathbf{I} - \mathbf{A})^{-1} \approx \mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \mathbf{A}^3 + \cdots + \mathbf{A}^n \qquad [2.7]$$

Thus the Type I multiplication process is made up of the *initial* effect, **I** (unity, the multiplicand), the *first-round* effect **A**, and the *industrial support* effect which is the sum of all subsequent *n* rounds of 'knock on' demand. It should be clear that since each a_{ij} must be non-negative and less than unity, as $n \to \infty$ then $\mathbf{A}^n \to \mathbf{0}$. The power series expansion forms the basis of the disaggregated multiplier framework (West and Jensen, 1980) where the separate effects are identified within the inverse solution by subtraction.

There are two key assumptions behind the Leontief analysis. The first is that the production functions exhibit constant returns to scale. That is to say that, if the production of one unit of j requires a_{ij} of the ith input, then the production of n units of j requires na_{ij} of i. This also implies that inputs are purchased in fixed proportions (*i.e.* there is no input substitution). The second assumption is that the defining sectors of the input-output tables are homogeneous. Hence, each column of coefficients must describe the production of a single and unique product. Clearly, if the a_{ij} relate to a number of commodities with quite different input purchasing patterns, the jth multiplier loses definition and becomes difficult to interpret.

The basic output analysis can be extended to consider the effect on income and employment, or indeed any other sectoral variable. The approach is a simple one since it assumes that there is a direct linear relationship between output and the variable in question. So, in calculating the impact upon incomes for example, the vector of income per unit of output is calculated $\mathbf{y}$ $[y_j]$:

$$y_j = \psi_j / X_j \qquad [2.8]$$

where

ψ is labour income

The vector of effects upon income for a unit increase in output in the jth industry, $\mathbf{k}$ $[k_j]$ is then given by:

$$\mathbf{k} = \mathbf{y}(\mathbf{I} - \mathbf{A})^{-1} \qquad [2.9]$$

The Type I income multiplier, which expresses the change in the income of the economy for a unit (£) increase in *income* in the jth industry is therefore given by k_j / y_j. The disaggregated framework follows the same principle as before:

$$\mathbf{k} \approx \mathbf{y} + \mathbf{yA} + \mathbf{yA}^2 + \mathbf{yA}^3 + \cdots + \mathbf{yA}^n \qquad [2.10]$$

Multipliers can therefore be estimated for any variable which can be related to output in the manner implied by [2.8].

Type I multipliers are a description of interdependency between productive sectors alone. However, one important agent within the economic system is clearly the employee/consumer. The Type II multiplier process takes into account the effect of earned labour income being spent within the economy. The approach to this is, again, straightforward. The labour row and consumer column are simply endogenised as another 'industry'. Hence the **A** matrix is extended by one row and column to include the direct coefficients of labour income (y_j) and consumers' expenditure *i.e.* c_i / C where C [c_i] is consumers' expenditure. The solution is the same as that described by [2.6] and [2.7], although the **A** matrix is of course larger. The calculation of Type II multipliers follows exactly the same principles as for the Type I process, and *consumer induced* effects can be isolated within the solution by subtraction:

$$\mathbf{Y} = \begin{bmatrix} (\mathbf{I}-\mathbf{A}) & -\mathbf{c} \\ -\mathbf{l} & 1 \end{bmatrix}^{-1} - \begin{bmatrix} (\mathbf{I}-\mathbf{A}) & 0 \\ 0 & 1 \end{bmatrix}^{-1} \qquad [2.11]$$

where

$\mathbf{c}$ is the vector of consumers expenditure coefficients

$\mathbf{l}$ is the vector of labour income coefficients.

The analytical function of the input-output table illustrated here represents only the simple case. The system of equations given by [2.4] can be rearranged into a solution which makes any combination of variables endogenous and exogenous. So, for example, the purchases and sales of an industry could become an exogenous component, thus allowing an assessment of its local economic 'significance' (see Miller and Blair, 1985 for a review; also West, 1993). The specification provides a basis for analysing economic structure, and moreover, for analysing changes in economic structure. Some of these functions are considered later in this chapter and in Chapter 4.

2.3 The Strengths of the Regional Input-Output Model

The strengths of the input-output model as an approach to regional analysis can be revealed by considering the less sophisticated alternatives. Three such specifications are reviewed: shift-share analysis; economic base multipliers; and Keynesian income multipliers.

2.31 The Shift-Share System

Firstly, it should be stressed that the shift-share system is more of a statistical tool than a regional economic model. It is a simple and rather crude way of categorising and analysing the components of regional economic growth between time periods. Bishop and Simpson (1972) provide an account of the technique. Three components of growth are identified:

(*i*) National Growth

This is the growth in the region that would have occurred, had the region grown at the same rate as the nation during the period in question.

(*ii*) Structural Growth

The structural component captures the extent to which the regional economy specialises in relatively fast- or slow-growing national industries.

(*iii*) Differential Growth

This can be defined as the extent to which the regional economy has grown at a faster or slower rate than the nation.

Empirical studies have generally focused upon the analysis of employment, which is assumed to reflect changes in regional economic welfare. There are a number of ways of calculating the three growth components. The 'National Growth Rate Standardisation' is summarised in equations [2.12] to [2.15] below.

$$n_{i,t} = \left(e_t^n / e_{t-1}^n\right) e_{i,t-1}^r \qquad [2.12]$$

$$s_{i,t} = \left(e_{i,t}^n / e_{i,t-1}^n - e_t^n / e_{t-1}^n\right) e_{i,t-1}^r \qquad [2.13]$$

$$d_{i,t} = \left(e_{i,t}^r / e_{i,t-1}^r - e_{i,t}^n / e_{i,t-1}^n\right) e_{i,t-1}^r \qquad [2.14]$$

$$e_{i,t}^r = n_{i,t} + s_{i,t} + d_{i,t} \qquad [2.15]$$

where

n, s, d are, respectively, the national, structural, and differential growth components;

e is total employment;

e_i is employment in the ith industry;

r, n denotes the region and nation respectively;

t denotes the time period;

Interpreting t-1 as the present, and t as a future period, the technique can be used as a tool for forecasting regional employment change. It should be clear from the above formulae that both the national and structural growth components can be derived from national employment forecasts. Keurre and Weller (1989) therefore consider how one might forecast the differential component. Neoclassical assumptions predict a zero differential over time as regional growth rates converge. The cumulative causation model however suggests that agglomeration economies may be achieved as the regional industry grows and therefore the differential component may remain non-zero over time. Keurre and Weller suggest that, since economic theory can provide no definite indication as to the likely behaviour of the differential component over time, time series forecasting techniques should be employed.

2.32 Economic Base Multipliers

The principles of economic base analysis can be traced back to Aurousseau (1921). The economy is said to consist of two sectors: the export or 'basic' sector, and the service or 'non-basic' sector. Regional growth is said to be wholly export-driven. Regional output (Y) is therefore represented as the sum of basic (b) plus non-basic (n) output:

$$Y = b + n \qquad [2.16]$$

In addition, the level of non-basic output required by the region is said to be proportional to total output:

$$n = eY \qquad [2.17]$$

Substituting [2.17] into [2.16] and rearrange gives the expression:

$$Y = b/(1 - e) \qquad [2.18]$$

and therefore the economic base multiplier, defined as the change in total regional output arising from an increase in exports, d*Y*/d*b*, is equal to 1/(1-*e*). Income and employment base multipliers can also be derived (see, for example, Wagstaff, 1973). The usual approach to estimating the base multiplier is to estimate total regional output and, in particular, the output of the exporting sector - with *Y* and *b* known, *e* can be inferred from [2.18].

2.33 The Keynesian Regional Income Multiplier

The Keynesian regional income multiplier can be derived by specifying a simple macroeconomic model:

$$Y = C + I + G + (X - M) \qquad [2.19]$$

$$C = cY_d \qquad [2.20]$$

$$M = mY \qquad [2.21]$$

$$Y_d = (1 - t)Y \qquad [2.22]$$

where

Y is regional income;

C is consumption;

I is autonomous investment;

G is autonomous government expenditure;

X is exports;

M is imports;

Y_d is disposable income;

c, m, t are the marginal rates of consumption, importation, and taxation.

Investment, government expenditure, and exports are assumed exogenous. Substituting [2.20] to [2.22] into [2.19] and rearranging gives:

$$Y = \frac{(I + G + X)}{(1 - [c - m][1 - t])} \qquad [2.23]$$

And hence the Keynesian income multiplier, the change in income arising from an increase in exogenous demand, $dY/d(I+G+X)$ is equal to $1/(1-[c-m][1-t])$. There are a number of variations on [2.23] - for example Brown *et al.*, 1967; Steele, 1972; Black, 1981, however, essentially the Keynesian multiplier varies inversely with the propensity to withdraw from the economic system.

These three alternative modes of analysis are clearly quick and easy to implement, and this is their principal attraction. Simplicity however is the root of most criticism. The shift-share system represents a cheerful approach to the analysis of employment change, but is clearly just too limited to offer anything more. The possession of *input-output tables* for a given region at two points in time clearly opens up a vastly superior range of structural decomposition analyses (see, for example, Dewhurst 1993; Madden *et al.*, 1996). The estimation of the economic base is notoriously problematic. The most usual method is to assume relative regional specialisation identifies exporting sectors, but the empirical pedigree of such techniques has been very poor (see, for example, Isserman 1977; Norcliffe, 1983). The Keynesian approach has maintained greater respect, but this is largely due to the fact that studies have focused upon assembling good primary data for the estimation of the multiplicand, recognising that the Keynesian multiplier specification is just too aggregate an approach to warrant major attention (*i.e.* Wilson, 1968; Armstrong, 1993; Bishop, 1996). It is the fact that the input-output framework takes explicit account of the pattern and strength of local linkages that makes it a superior tool for economic analysis.

But relative analytical sophistication is really just one of the rewards to what is perceived as the main justification for promoting the input-output approach in Devon and Cornwall. The real advantage of the input-output approach over these alternatives is that it is represents a framework for the creation of a regional economic database. Instead of the local data set being a series of separate sector-based Keynesian studies, the input-output approach should integrate these studies within a framework which provides a general strategy for the collection, assembly, and use of local economic information in future. As such it should result in a much more efficient and effective approach to local data collection and analysis.

The input-output approach is of course not without its weaknesses, and these short-comings relate mainly to its analytical function. However, as the next section illustrates, higher levels of sophistication are, in general, accessed by building on and around the basic input-output data set. The input-output table is therefore very much a foundation stone for more advanced modes of analysis.

2.4 The Weaknesses of The Regional Input-Output Model

The basic Leontief input-output specification is far from perfect. This section focuses upon two essential areas where the basic input-output specification may be considered relatively 'weak'. The first of these relates to its representation of the various players within the economic system. The basic Leontief approach is very much an industry-orientated representation. The activities of other economic agents, in particular householders, may however be of greater importance and relevance within a regional analysis. The second weakness of the input-output specification relates to the fact that it is a strictly static representation of the economy. The model's assumption of linear production functions, with fixed technical coefficients and factor prices eliminate the possibility for substitution amongst inputs over time (*i.e.* structural change). This lack of dynamism weakens the model's potential use for impact analysis and economic forecasting.

However, research has provided a number of solutions to these inherent deficiencies, and the following section considers some of these. The first section briefly illustrates how the basic input-output framework can be extended to take a more explicit account of economic-demographic factors. The second section considers, again only briefly, the development of more a dynamic analytical framework.

2.41 The Household-Extended Regional Input-Output Framework

The basic Leontief model has its principal focus upon the productive sectors of the economy. The specification provides only a minimal representation of the labour market and the role of consumers within the economic system. However, there are a number of reasons for believing that population should be given much greater attention within the specification. Firstly, as it will be revealed in later chapters, the household-income coefficients have been shown to be of particular importance within the process of multiplication (*i.e.* Garhart, 1985; Hewings, 1986). These results extend from the fact that labour is generally the largest single item of expenditure within the production function; householders simultaneously exhibit the most complete range of product demand of any purchasing sector. Secondly, including labour-householders as a single sector undoubtedly offends the homogeneity assumption which lies at the root of a 'sensible' Leontief solution matrix. The single sector representation undoubtedly conceals a number of groups with identifiably different labour returns and patterns of consumption. Thirdly, and a point which is related to the last, the Leontief model fails to deal with issues such as migration, unemployment and transfer payments in any satisfactory way, and yet these issues will undoubtedly be of particular relevance within any regional analysis. There is therefore a very strong case for giving representation to demographic features within the regional input-output model.

The household extension of the simple input-output framework is most readily associated with the work of Batey and Madden (for example Batey and

Madden, 1981; Madden and Batey, 1986; Batey, 1991; Batey *et al.*, 1993; Madden *et al.*, 1996).

Although there are various formulations and re-formulations of the household-extended model, the block matrix representation of [2.24] provides an illustration of the model when a simple distinction is made between employed and unemployed workers

$$\begin{bmatrix} (\mathbf{I}-\mathbf{A}) & -\mathbf{c}^e & -\mathbf{c}^u \\ -\mathbf{l} & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{X} \\ e \\ u \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ 0 \\ L \end{bmatrix} \qquad [2.24]$$

where

L, e, u are scalars representing total labour supply, employment and unemployed workers respectively;

$\mathbf{l}$ a row vector expressing employment per unit of output;

$\mathbf{c}^e$ a column vector expressing consumption expenditures per employed worker;

$\mathbf{c}^u$ a column vector expressing consumption expenditures per unemployed worker.

The block at the far left of [2.23] provides the basis for the computation of Type IV multipliers[2]. Note that in [2.23], the usual link between income-expenditure is replaced by a relationship between employment generation and average spending per employee. So, increases in output lead to increases in labour demand; the consumption of employed households rises, that of unemployed households falls as the new jobs are taken up. Consequently Type IV multipliers are generally lower than those calculated under a Type II process. Development of [2.23] seeks further to improve the functional link between labour reward and spending. Recognising that expenditure data relates to *households*, whilst the labour-generation process is concerned with *individuals*, Madden and Batey (1986) offer a function which attempts to provide a mapping between the two. Batey (1991) considers further development to take into account: variation in household size; the transition between economic activity and inactivity; in-migration; the role of variation in sectoral wage rates on household consumption; and transfer payments within the context of variable household size.

2.42 Introducing a More Dynamic Analytical Framework

The static nature of the input-output framework is of obvious concern when one comes to consider its function within impact analysis and economic forecasting. Leontief (1970) introduces a dynamic framework which requires the estimation of a matrix of 'capital coefficients', **V** $[v_{ij}]$ - each element representing the stock of i used per unit of j during the given time period. Equation [2.6] becomes:

---

[2] The Type III framework is developed by Miernyk *et al.* (1967) and distinguishes between the spending patterns of new and indigenous households.

$$\mathbf{X}^t = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{f}^t + (\mathbf{I} - \mathbf{A})^{-1}\mathbf{V}\Delta\mathbf{K} \qquad [2.25]$$

where

t denotes time period;

$\Delta\mathbf{K}$ measures the change in capacity between time periods.

Another approach to a more dynamic regional modelling framework been the development of regional econometric input-output models (REIM). The REIM incorporates the input-output framework within a wider system of econometric equations which explain, for example, labour, income, household consumption, demographic aspects, and prices over time, thus enabling a disaggregated temporal analysis of the regional economy (see Dewhurst and West, 1991; West, 1994). In the UK, the Cambridge Econometrics multi-regional model (Cambridge Econometrics, 1993) is essentially a set of 'regionalised' national input-output information which is driven by econometric forecasting equations. Israilevich *et al.* (1996) show that using 'regionalised' input-output information (*i.e.* information that is not based on direct regional observation) is a significant source of forecast error. This result highlights the fundamental importance of a good regional input-output *database*.

The other main approach to dynamic modelling has been the development of the computable general equilibrium (CGE) framework. CGE models are characterised by their treatment of the supply side of the economy (although West, 1994 notes the 'blurring' of REIM and CGE specifications). It should be clear that the Leontief specification is entirely demand driven, with relative price changes eliminated from the simulation. The irony is that, whilst the treatment of supply within the input-output model amounts to a long-run assumption, such models are most frequently used within the context of short-run analyses (Harrigan *et al.*, 1991a). Hence the CGE model is a supply-constrained specification, which allows fluctuation in factor prices, and hence substitution between inputs (although usually only between intermediate purchases and value added). The CGE model is generally specified from a social accounts matrix (SAM), which covers a much broader spectrum of economic activity than the production-based accounts of the input-output framework (see Round, 1986). The data requirements of the CGE specification expand rapidly with sectoral disaggregation and thus such models tend to differentiate between a select number of commodities. Harrigan *et al.* (1991b) present a CGE model for Scotland which is defined across four transactions groups - households, non-household personal sector, corporations, and government; three commodities/activities - manufacturing, traded non-manufacturing and non-traded; and two factors of production - labour and capital. McGregor *et al.* (1996) provide some simulation results produced by the model.

2.5 Conclusion

In conclusion therefore, the input-output model can be justified on the grounds that it represents a solid framework for the formation of a regional economic database. The absence of a cohesive strategy to the collection, collation, and analysis of the regional data set seems to be one of the most obvious reasons why a 'regional data problem' exists. In providing this strategy, the regional input-output framework should offer cost and efficiency savings in the process of local economic analysis.

The regional input-output specification can also be justified from an analytical perspective. The model's analytical function has been shown to be superior to the shift-share system, economic base, and Keynesian multiplier specifications. It should be relatively clear that, in terms of the assumptions and principals underlying input-output, economic base, and Keynesian multiplier approaches, there is little to distinguish between the three. The input-output model's superiority is afforded by the fact that it is specified upon a vastly superior set of parameters - *i.e.* the observed patterns of supply and demand within the economic system.

The weakness of the input-output table have been explored in terms of its unsatisfactory treatment of labour incomes, household expenditures, and demographic factors; and the general lack of dynamism within the basic Leontief specification. The household-extended input-output approach has sought to address the former issue. Dynamic solutions have included the development of regional econometric input-output tables and CGE specifications. The fact that these solutions build on and around the input-output framework signifies the importance of the initial database.

Within the context of Devon and Cornwall, the development of REIM and CGE specifications would have to represent long term objectives. In terms of the REIM, there are severe limits on the time series data available at the county level; even if the collection of such data began with immediate effect, the number of observations required for a sensible time series analysis would only accrue in the long run. The CGE specification requires careful consideration at the county level: West (*op. cit.*) questions the applicability of a general equilibrium framework for relatively small, open economies.

However, there would seem little to stop the development of demographic features within any input-output specification for Devon and Cornwall - most of the extra information could probably be found, at least in some form, within the Census of Population and Family Expenditure Survey. The household-extension represents an important and necessary addition to the regional input-output table. It would seem an appropriate step in developing an input-output database for Devon and Cornwall. However, the next chapter turns to consider the methods of generating the basic input-output information set more closely.

CHAPTER 3

METHODS OF GENERATING DATA

FOR USE IN

REGIONAL INPUT-OUTPUT TABLES

3.0 Methods of Generating Data for use in Regional Input-Output Tables

3.1 Introduction

Chapter 2 introduced the basic framework of the regional input-output model. In Chapter 7, a more explicit illustration of the input-output accounting framework will be presented. The focus of this section however is upon exploring the issue of data generation within the context of the regional input-output model. From the previous chapter it should be clear that the regional input-output model is a data-intensive specification and therefore, the issue of data generation is central to the area of regional input-output research. The section describes the two broad approaches to data generation: survey and nonsurvey, and considers the principal advantages and disadvantages of each. The chapter concludes with an introduction to hybrid methodologies which combine elements of both the survey and nonsurvey genre. Critical discussion of individual methods or studies within each broad area of approach is kept to a minimum. The aim is to make way for a deeper critical and theoretical analysis of specific methodologies in the following chapter. However, the first section of the chapter explores a concept which is fundamental to the process of data generation for input-output models, and that is the notion of accuracy.

3.2 The Concept of Accuracy within Input-Output Models

Jensen (1980) defines two principal views of accuracy within the context of input-output models. The first is 'A'-type accuracy which is an expression of the degree to which the estimated input-output model reflects the 'true' unobservable input-output model. Jensen's 'B'-type accuracy is concerned with the degree to which the input-output model correctly specifies the true function of the given real economy.

This latter form of error is extremely difficult to observe because, for example, the simultaneity of diverse events in any real economy make it impossible to isolate the effect of any given impact from which some degree of assessment could be made. McNicholl's (1982) *ex post* analysis of Shetland is perhaps the only study of its kind to have been completed in the UK. Other possible angles on 'B'-type accuracy involve relaxing the assumptions of the input-output model, for example, through a CGE framework, and making a comparative analysis of the simulation results (*i.e.* Harrigan *et al.*, 1991b; Conway, 1991; West, 1994).

The intractability of 'B'-type accuracy has meant that, relatively speaking, its existence has been ignored. The principal focus of research has been on aspects of 'A'-type accuracy. So, for example, authors have tested the performance of a given data simulation technique on the basis of the assumption that the estimated survey table which is in their possession represents the 'true' set of observations (*i.e* Czamanski and Malizia, 1969). One of the key distinctions concerning 'A'-type accuracy lies between 'partitive' and 'holistic' accuracy.

Partitive accuracy concerns a *cell-by-cell* assessment of the input-output framework. If a model is to be considered accurate in the partitive sense its individual entries form a correct accounting statement of the transactions that took place in the given economy in the given period of time.

Holistic accuracy on the other hand demands much less in terms of the exactness of the data set. For a model to be considered accurate in the holistic sense, then it should succeed in capturing the main features of the economy in question. To achieve this, there must exist some identifiable order of importance amongst the elements of the input-output model for the given economy.

These distinctions are fundamental to the issues that are covered in the remainer of this chapter, and indeed, the remainder of the thesis.

3.3 The Survey Approach

The first, and most obvious method of constructing a regional input-output table is by survey. The defining characteristics of a 'survey model' are not altogether clear (Hewings and Jensen, 1986). In the purest sense, the survey model would be formed in the partitive sense, by observing the purchasing and selling decisions of each and every economic agent: industries, householders, government, and all other defined sectors of the regional economy. As Miernyk (1976 p.53) comments however:

'...such an ideal probably is unattainable. Judgement enters into the construction of any input-output table at many stages.'.

In practice the survey model will be based upon a sample of agents and as such will be vulnerable to the usual problems of, for example, sampling and non-response biases which accompany such exercises (see Bulmer-Thomas, 1982 Chapter 3). Hence, all input-output models can only be classed as 'survey-based' which, purely in terms of the general accuracy of the information set, must imply a degree of inferiority with respect to the *'ideal'*. It is therefore true to say that the survey approach *attempts* to deliver partitive accuracy, but its ability to do so is invariably constrained to the holistic level. But what distinguishes the 'survey model' from other modelling approaches? It would seem that the key feature of the survey model is that its framework, in terms of what it describes, is *constructed*, and this process is not constrained by the framework of any other model, but only by the limitations of the primary data set. Since regions tend to be idiosyncratic, the important advantages of such freedom are obvious. Morrison (1973) for example is able to define no less than fourteen retail sectors in his study of Peterborough. Any alternative approach cannot compete, as he comments:

'It is ... particularly disappointing to see all distributive trades allocated to one row and one column in the ... United Kingdom input-output table...'

Furthermore, national tables, which are the starting point for all other construction methods, are produced at what can be distant intervals. In the UK for example, at the start of 1994, the only full set of published national tables related to 1984. Constructing the regional model by means of survey clearly avoids this constraint. Hence the survey approach allows the model builder to define a scheme of design which is appropriate to the economy under study.

However, the relative accuracy and flexibility of the survey approach comes at a cost, and indeed the resources required to construct a survey table are commonly viewed as *'prohibitively expensive'* (Flegg *et al.*, 1995). Richardson (1985 p.630) notes the *'demise of the survey-based model'* in the United States and observes that the US regional survey tables that do exist are for small areas. The same has been true in the UK, for example: McNicholl (1976) studies Shetland; Morrison and Smith (1977) study the town of Peterborough; McDowall (1973) surveys Sutherland; Proops *et al.* (1981) compile a table for North Staffordshire. The only large region to have any history of survey-based input-output is Scotland (for example Fraser of Allender, 1978; Scottish Office, 1994). Scotland is - at the time of writing - distinct from most other regions of the UK in that it has some status as a country and an additional 'layer' of government in the form of the Scottish Office. These factors may contribute significantly to the ability to raise resources; to gain access to individual Census returns and 'unpublished' statistics; to gain the support of industry, thus positively influencing the rate and quality of response. The reason behind the general bias towards the small area model however is obvious: the smaller the area, the smaller the population; complete identification and coverage of economic agents is therefore both relatively cheap and easy. So for example, Morrison (1973) is able to construct a sample which equates to 50% employment coverage, with some sectors completely represented.

Most studies are based upon postal survey[1]. Therefore, having identified the population, scheme of disaggregation and constructed a relevant sample, the model builder must design an appropriate questionnaire. A poorly or ambiguously designed form can clearly lower the rate and quality of responses irrespective of the context of the survey. Within the context of input-output analysis however, the disaggregated nature of the study means that complex information is required; questionnaire design is therefore critically important. Typically, the survey form requests information on five main categories: employment, wages and salaries; operating purchases; capital expenditure; sales; and the source and destination of all transactions. Generally, rather than asking the respondent to specify purchases from a full list of commodities, a selection of 'common' purchases such as electricity are specifically requested with the respondent encouraged to select and define other purchases from a given list. The objective is to simplify the look and 'feel' of the questionnaire and increase the probability of a return. Usually there is some 'tailoring' of questionnaires to specific

[1] The discussion of this paragraph is based principally upon Hill and Roberts (1996), Roberts (1996), and the Scottish Office (1994); each of these texts provides examples of survey questionnaires.

industry groups. So, for example, a manufacturing questionnaire may include a detailed list of material purchases; a retailing questionnaire will ask for information on purchases for resale. Respondents are typically asked to place a monetary value on the total purchase or sale of a commodity and then asked to estimate the source or destination as a percentage. This of course results in 'rounding' which erodes the quality of the information, but as Miernyk (1976) points out, many respondents probably do not *know* the exact source and destination of their trading activities, and hence rough estimates are the best that can be hoped for. Inevitably, ambiguities and misinterpretations will arise which mean that a proportion of responses are unusable (for example, in the 1995 Welsh Input-Output survey returns some respondents treat the questionnaire as a 'tick-box' exercise: the information is almost entirely useless). Most surveys improve their response rate and the quality of their data by incorporating feedback from a pilot study into the final questionnaire design, by sending reminder letters to initial non-respondents (this may enable a test of response bias, Oppenheim, 1992 p.34), by clarifying phone calls, and by in-depth interviews with responding firms or industry 'experts'. Other techniques, such as timing the survey to coincide with statutory questionnaires that make similar informational requests (*i.e.* Census of Production *etceteras*) have been employed to encourage and improve responses. Once a satisfactory sample of responses as been obtained, the information is 'grossed up' using published output or employment figures as control totals. From here, items such as distributors' trading margins have to be removed. However, even at the national level, little is known about this item (ONS, 1995).

The most radical approach to survey is where respondents are *given* a set of input-output coefficients and asked to adjust them accordingly. Although there exist no formal guidelines, Hewings (1996) comments on the use of such methods:

'We found the response rates and quality of responses to be much higher.'

In a study of metal plants in Chicago, Hewings cross-checks responses against known purchases of electricity. The approach is also reported to have been used as a means of updating Australian regional tables. Whilst the approach has some obvious advantages, the logistical difficulties involved in such an exercise, for example in ensuring the correct set of coefficients go out to each firm, would seem to limit its general applicability.

A more serious data generation problem relates to the vector of household expenditures. Clearly, whilst industries may have some record and knowledge of the value source of their purchases in the year of study, householders are likely to possess little or any. In all but the most extreme of study areas, households will vastly outnumber industries. The cost and logistical difficulties involved in obtaining a representative sample of consumers' purchasing decisions may therefore be truly prohibitive. McNicholl takes a 10% sample of Shetland households, but this represents a virtually unique example. The true gravity of the problem has only become apparent in the light of a succession of studies which have shown that the results of impact analysis are most sensitive to changes, or errors, within the vector of household expenditures, and are therefore the most 'important' to estimate correctly within the paradigm of holistic accuracy (see, for example, Hewings and Romanos, 1981; Garhart, 1985).

One of the areas to attract considerable attention in the literature of the mid-seventies is that of 'reconciliation'. As mentioned above, the general survey method seeks to obtain information on both sales and purchases of industry. In all but the most extraordinary of cases, this process will generate two estimates of the value of a single cell in the transactions matrix: one based upon sales information, the other on purchasing data. Hence Jensen and McGuarr (1977 p.328) define the reconciliation problem as:

'...the derivation from these dual estimates of single-valued estimates, which are consistent with row and column totals.'

In many cases, the model builder will have little or no knowledge as to the relative accuracy of either estimate, and perhaps only limited information as to the implications of any reconciliation decision upon the model's impact-analysis facility. Various approaches to reconciliation are considered, ranging from simple judgement (Borque *et al.*, 1967), methods of averaging, and constrained optimisation techniques such as RAS (see the later section in this chapter). Whilst Jensen and McGuarr's experimental results support the use of constrained optimisation methodologies, perhaps more significantly for the survey analyst, they reveal that constructing a table purely by purchases-only data produces results which are *'more satisfactory'* than the alternative sales-only data set.

Gerking (1976; 1979) addresses the problem of reconciliation from a probabilistic perspective. He suggests that, given data on purchases from a sample of firms within a sector, input-output coefficients could and should be estimated econometrically, rather than the more traditional approaches which take a simple average of these observations, or rely upon subjective judgement. Gerking's approach fuelled much debate: Hanseman and Gustafson (1981) and Hanseman (1982) challenge the specific estimator suggested by Gerking. Miernyk (1976), Brown and Garrianti (1979) express concern over the suppression of judgement - a point which Jensen (1980) notes is a conflict between partitive and holistic accuracy. Jackson's (1986) 'full-distribution' approach treats each sample coefficient as deriving from the population of coefficients for a firm within that sector. For purely illustrative purposes, Jackson treats the sectors of a highly disaggregated US national model as observations of individual firms. Hence every cell at a higher order of aggregation has an underlying, and known, frequency distribution which Jackson uses to generate interval estimates for gross output and multipliers.

Finally, it may not be possible to publish, or indeed operate with information at the desired level of disaggregation for reasons of disclosure. Thus it may be necessary to aggregate sectors, which will inevitably result in some information loss. The principals under which this is achieved are relatively straightforward, and seek to minimise information loss: 'key' sectors may be identified by some relatively crude measure (see Rasmussen, 1957; Chenery and Watanabe, 1958 for example) or by more sophisticated means (see for example West, 1982) and the aggregation should, wherever possible, avoid disturbing these elements (see Hewings, 1974).

Therefore it is clear that the survey model builder armed with even a fairly complete set of returns faces a set of problems which act to undermine the partitive accuracy of the entries within the final input-output table. Initially, the survey model builder is reliant upon the judgement of the respondent, and in turn he or she will have to call upon other forms of subjective assessment: this error generating process is inevitable. At the same time, constructing a regional model from 'first principles' is a resource intensive exercise. Given the imperfections that are inherent within input-output data and the high cost of collation, analysts have sought to develop techniques which avoid the costs of model construction whilst providing an intersectoral framework for assessing regional impacts. These so-called nonsurvey techniques are presented in the following section.

3.4 Nonsurvey Methods

The extreme alternative to the survey approach is to apply nonsurvey or 'bottom down' methodologies. The defining characteristics of nonsurvey methodology are that they make use of a national input-output framework and seek to estimate from this a matrix of regional purchasing coefficients using published information (Smith and Morrison, 1974; Miller and Blair, 1985; Hewings and Jensen, 1986). The genre of nonsurvey techniques therefore aim to 'scale down' the national input-output model to regional dimensions using what are essentially mechanical procedures. The emphasis is upon producing a table quickly and cheaply, whilst maintaining a general level of accuracy; the regional data requirements for these procedures are necessarily low. Generally, the minimum regional data requirements are considered to be a national input-output framework and regional and national employment figures disaggregated to a reasonable[2] level.

A wide range of nonsurvey techniques have been developed and it is neither possible nor desirable to consider every variation. Moreover, whilst the defining characteristics of a nonsurvey technique are clear, variation in the availability of published regional data sets between countries has meant that authors have differed in their identification of the specific group of techniques which make up the nonsurvey set. The following section therefore presents the main techniques of nonsurvey modelling as defined from the perspective of published UK regional data. Notable absentees from this set are the 'short-cut' multiplier approach (*i.e.* Burford and Katz, 1977) and RAS (*i.e.* Stone, 1961) which have been classified as hybrid procedures.

Furthermore, what follows pays particular attention to the most recent UK developments within the nonsurvey field.

3.41 'Classic' Nonsurvey Methods

The following section presents methodologies and assumptions which are evident in the early nonsurvey studies (*i.e.* Schaffer and Chu, 1969) and therefore may be viewed as the 'classic' or 'original' set. This set includes the assumption of national technology, the commodity balance approach, simple and cross-industry location quotients.

3.411 The Use of National Technical Coefficients

National technical coefficients - the proportions of industry purchases sourced nationally - lie at the heart of nonsurvey methodology. In principal, they provide a skeleton for the regional model which would otherwise have to be constructed by resource-intensive survey work. Qualification for such an approach can be found within the notion of Fundamental Economic Structure (FES) (*i.e.* Simpson and Tsukui, 1965) which claims that the basic forms of production are relatively similar across developed economies. Hence for any given nation, one may expect to observe economic structure which is fundamental to the nation within any of its defined spatial subsets. The starting point for all nonsurvey methodologies is therefore:

---

[2] The interpretation of 'reasonable' may be a function of a number of factors *i.e.* regional size, detail within the national model *etceteras*. Consistency must of course be achievable between employment and input-output definitions.

$\mathbf{A}^r \approx \mathbf{A}^n$ [3.1]

where

A denotes the estimated matrix of technical coefficients;

r, n denote the region and nation respectively.

Hewings and Jensen (1986 p.310) question the logic of [3.1] as the initial framework for the nonsurvey regional model since it implies that the propensities to import inputs from overseas are equalised across space. Since there is no economic reason why this should hold, they suggest that the coefficients of the combined use matrix, which describe the production functions in full, are a more appropriate starting point for the nonsurvey estimation process.

Smith and Morrison (1974 p.22) comment:

'In attempting to adapt national coefficients at the regional level it is essential either to make some allowance for... [variations] *... or to make the assumption that they are insignificant in their effect.'*

As it will become clear, most nonsurvey techniques work from the assumption that differences in technology are insignificant and assume that [3.1] holds as an identity. Hence nonsurvey techniques extend the notion of 'similarity' implied by the FES literature to claim that national technical coefficients are *invariant* across *any* spatial subset. The validity of assumption [3.1] is strongly challenged in the next chapter.

Clearly, the nonsurvey regional model is constrained at the maximum to the defined sectors of the national input-output framework. This scheme may well be inferior to the desired representation (see Morrison, 1973 above). Hence, the flexibility of the survey approach is generally sacrificed by nonsurvey procedure.

3.412 Unadjusted National Coefficients

One additional extension to identity [3.1] is found in the earliest applications of nonsurvey methodology. In identity [3.2] below, the unadjusted national technical coefficients are used to proxy the regional purchasing coefficients (see for example Moses, 1955; Moore and Peterson, 1955):

$$\mathbf{R} \equiv \mathbf{A}^{n} \qquad [3.2]$$

where

R is the estimated matrix of regional purchasing coefficients

One would suppose that the conditions under which [3.2] holds would be considerably more severe than in [3.1] (Chapter 5 considers the case where [3.2] holds).

3.413 Commodity-Balance or Supply-Demand Pool Approaches

It is commonly accepted that as regional size diminishes, the range of local commodities supplied becomes smaller, and hence the propensity to trade inter-regionally increases. Identity [3.2] tends to suggest that interregional trade does not take place, or at least, as a weaker equality, that imports are non-competitive (*i.e.* there is no competing local supplier). Discomfort with [3.2] is reflected in the number of early studies which attempted to make some adjustment to the national technical coefficients in order to account for competitive imports. In particular, studies by Isard and Keunne (1953) and Miller (1957) attempted to construct a local consumption propensity by:

$$p_i = \frac{X_i - E_i}{X_i - E_i + M_i} \qquad [3.3]$$

where

X, E, M are gross output, total exports, total imports of commodity i respectively.

The matrix of regional purchasing coefficients is then estimated by applying these propensities across the rows of the national technical coefficients matrix:

$$\mathbf{R} = \hat{\mathbf{p}}\mathbf{A} \qquad [3.4]$$

where

$\mathbf{p}$ is a vector formed from the i elements estimated by [3.3].

The approach is essentially based upon the notion of commodity balances (Isard, 1953). There is some debate as to whether this approach can be classed as purely nonsurvey, since data on commodity trade flows between regions are not normally available from published statistics and would therefore have to be generated by survey. However, the approach is included here because it holds a position as one of the forerunners of nonsurvey methodology. Two points of interest arise from Miller's and Isard and Keunne's use of the technique. The first is that the proportions in [3.4] were estimated using a variety of sources which included not only published trade data, but *'judgements of informed persons within the area'* (Isard and Keunne, *op. cit.*, p.296). Secondly, an assumption of a minimum import propensity of 5% was employed to reflect the fact that all commodities are imported to some extent. The use of 'informed judgement' and 'common sense' assumptions are important elements which, it will be argued in the next chapter, have been suppressed in recent developments.

Schaffer and Chu (1969) present a technique which essentially formalises the commodity balances of equations [3.3, 3.4] as a purely nonsurvey methodology. In the first step of what they term as the 'supply and demand pool' (SDP) approach, local demand for each commodity, d_i is assessed by:

$$d_i = \sum_j X_j^r a_{ij} + Y_i \qquad [3.5]$$

where

Y_i is the region's share of national final demand for i (excluding of course exports).

Local supply of i is regional output, X_i. The balance between local supply and demand for i, b_i, is therefore given by:

$$b_i = X_i^r - d_i \qquad [3.6]$$

When the balance is positive, local supply of i is sufficient to meet local demand and the estimated average propensity to purchase i from local supply is taken as unity.

When the balance is negative however, the average propensity to purchase locally is given as the ratio of supply to demand. The estimated regional purchasing coefficients r_{ij} are therefore given by:

$$r_{ij} = \frac{X_i^r}{d_i} a_{ij} \quad , \quad \frac{X_i^r}{d_i} \leq 1 \qquad [3.7]$$

Kokat (1966) provides a similar procedure. Nevin, Round and Roe (1966) provide a practical example of the commodity balance approach in their table for Wales. The calculation of import propensities and the treatment of final demands can be found in a general section below.

Three points however are of importance. Firstly, national technical relationships are assumed to hold at the regional level. Secondly, all *j* local demanding industries are assumed to import *i* at the same rate. Self-sufficiency is assumed when local supply outstrips local demand.

3.414 Location Quotient Techniques

The second, and perhaps most popular, group of nonsurvey methods to have been used for the generation of regional input-output tables are based upon the location quotient. Location quotients have long been used as an indicator of export orientated production activity (see Isserman, 1977; Mayer and Pleeter, 1975 for a theoretical exposition). However, their formal use in the generation of regional input-output tables can be traced to Schaffer and Chu (*op. cit.*), although Round (1983) notes that these techniques were in practical use some time before this date.

3.4141 The Simple Location Quotient (SLQ)

As its title suggests, the SLQ is the simplest of location quotient formulae and indeed, it is true to say that all subsequent variations are a development upon its form. The SLQ for a given regional industry *i* is given as the *i*'s share of regional output divided by the share of national output held by the national industry *i*:

$$q_i = \frac{X_i^r / X^r}{X_i^n / X^n} \equiv \frac{X_i^r / X_i^n}{X^r / X^n} \qquad [3.8]$$

where

i signifies a particular industry, the sum of all industries otherwise

When q_i is greater than unity, the region is said to be relatively specialised in the production of *i* and is therefore a net exporter of the commodity; conversely, when q_i is less than unity, the region is identified as relatively unspecialised in the production of *i* and is therefore a net importer. A quotient of unity implies exact self-sufficiency within the production of *i*.

The conditions under which q_i correctly specifies the trade orientation of *i* are set out by Isard (1960). The methodology mainly requires equality between regional and national production techniques and consumption tastes. Cross-hauling (simultaneous import-export flow) of any given product is also eliminated. Mayer and Pleeter's general equilibrium analysis contests the assertion that average incomes have to be equalised across regions.

Where disaggregated output data are not available, proxy measures are utilised (see West, 1980a). Employment data are the usual substitutes, although value added or income are credible alternatives. The use of such proxy measures necessarily builds upon the assumptions of the output-based quotient. For example, use of employment data would require an assumed equality between regional and national labour productivities.

The assumptions upon which location quotient analysis is founded are generally viewed as 'restrictive'. Empirical analysis by Greytak (1969) and Isserman (1977; 1980) has found that the quotient does not identify trade orientation well. Within the context of their use in input-output analysis, Round (1983, p.197) comments:

'Unfortunately, [the conditions] *tend to assume away the very regional differences a regional input-output model is designed to highlight.'*

The method by which the simple location quotient is used to generate regional tables is really an extension of the original export base utility. The methodology is as follows. Firstly, by assumption, the framework of national technical coefficients is the starting point for the estimation process. The location quotients for each selling sector *i* are calculated and constrained to unity. The quotients are then multiplied across the rows of the national **A** matrix to form the estimated regional purchasing coefficients. Hence:

$$\mathbf{R} = \hat{\mathbf{q}}\mathbf{A} \qquad [3.9]$$

where

q is the i dimensional vector of location quotients, $q_i \leq 1$

The quotient is essentially a measure of the propensity to purchase locally and, as such, has strong ties with the commodity balance approach of equations [3.3, 3.4]. Indeed, the notable features of the SLQ are the same as for the SDP: that it works from national technology; it assumes a constant import propensity for each commodity; local specialisation in production entails a zero import propensity.

3.4142 The Cross-Industry Location Quotient (CILQ)

The Cross-Industry Location Quotient (CILQ) is often described as a refinement to ordinary SLQ methodology (for example Flegg *et al.*, 1995). The originator of the formula is somewhat unclear, although Schaffer and Chu cite Levin as one candidate. The principal argument against the SLQ is that, whilst it takes account of the relative size of the supplying sector i in the computation of the propensity to purchase locally, the SLQ fails to include any measure of the relative size of the demanding sector. The suggested formula for the CILQ between trading sectors $i\,j$, q_{ij}, is therefore:

$$q_{ij} = \frac{X_i^r / X_i^n}{X_j^r / X_j^n} \qquad [3.10]$$

Once again, whilst gross output is the preferred data set, other measures such as employment are common alternatives.

It can be shown that the CILQ is equivalent to the ratio of location quotients for i and j:

$$q_{ij} = \frac{q_i}{q_j} = \frac{(X_i^r / X_i^n)/(X^r / X^n)}{(X_j^r / X_j^n)/(X^r / X^n)} = \frac{X_i^r / X_i^n}{X_j^r / X_j^n} \qquad [3.11]$$

Hence, when the purchasing sector is relatively more specialised (or 'larger') than the selling sector, q_{ij} is less than unity and j will make some import purchases of i. When the selling sector is relatively more specialised than the purchasing sector q_{ij} is greater than unity and it is assumed that the selling sector can accommodate the local requirement. The application of the CILQ is principally the same as the SLQ with the national technical coefficients reduced by the propensity to purchase locally, which is constrained to values less than or equal to unity:

$$r_{ij} = a_{ij} q_{ij} \quad , q_{ij} \leq 1 \qquad [3.12]$$

It should be clear that since for any given row i, since the CILQ is variable, it can theoretically account for cross-hauling.

Despite a relatively poor empirical pedigree (see the following chapter) the CILQ remains one of the most popular nonsurvey techniques and is in evidence in a number of studies, see for example Johns and Leat, 1986; Garhart *et al.*, 1996.

Thus to summarise, the CILQ works from national technology. It differs from the SLQ in that it accounts for the relative size of the purchasing sector, and therefore does not assume a constant propensity to import across a given row. These points are applicable to the variations on the CILQ that follow.

3.42 Nonsurvey Methods: Developments on the Classic Approaches

A strong body of evidence has formed which has shown that the 'classic' set of nonsurvey methods do not provide acceptable simulations of regional input-output data (*i.e.* Schaffer and Chu, 1969). This evidence will be reviewed in the next chapter. However, on the basis of this evidence a number of suggestions aimed at improving the estimation performance of classic approaches have been made. This section presents the principal contributions.

3.421 National Technical Coefficients

Shen (1960) suggests that one way in which differences in technical coefficients may arise is through differences in regional and national industry mix. Starting with a highly disaggregated table for the United States, Shen demonstrates that by applying regional weights (*i.e.* gross output, although Shen uses value added) to national technical relationships, regional industry mix is reflected in the technical matrix at a higher order of aggregation. Shen's methodology however does not represent a physical adjustment process since the underlying national technical relationships of course remain unaltered.

Round (1972; 1978) suggests a mono-proportional column adjustment to the technical coefficients based upon differences in the proportion of value-added accounted for in gross regional and national outputs of the form:

$$s_j = \frac{\left(1-\alpha_j^r\right)}{\left(1-\alpha_j^n\right)} \qquad [3.13]$$

where

α_j represents the proportion of value-added in gross output for sector j.

$$\mathbf{A}^r = \mathbf{A}^n \hat{\mathbf{s}} \qquad [3.14]$$

Estimates of value-added in UK regions are available for at least some sectors (*i.e.* from the Annual Census of Production) and hence Round's technique can be classified as nonsurvey. However, there is little evidence to suggest that Round's technique has been applied in practice.

3.422 Adjustments to the SLQ

Tiebout (1967) suggests a modified SLQ known as the Purchases-Only Location Quotient (POLQ). Instead of total regional and national gross output in equation [3.8], Tiebout suggested that only the industries which made purchases from i be included. The application of POLQ is then precisely the same as for the SLQ. The extent to which this difference in approach translates into a significant difference in the value of the quotient is questionable. Indeed Smith and Morrison (1974, Table A2) show the difference to be virtually negligible.

West (1980) suggests accounting for differences in per capita consumption levels between region and nation for each commodity, and this is also reflected with respect to the quotient's more general export-base use in Norcliffe (1983).

3.423 Smith and Morrison : Principal Diagonal Adjusted CILQ

One particular result often quoted in reference to the CILQ is that when $i=j$, $q_{ij}=1$ (see Smith and Morrison for example). Hence the CILQ predicts self-sufficiency along the principal diagonal of any regional input-output matrix (the intra-industry flows). In the light of countering empirical evidence, Smith and Morrison suggest two variations on the CILQ aimed at improving its simulation performance along the principal diagonal. The first suggestion is to apply the (unitary constrained) SLQ for $i=j$ transactions. The second and more severe suggestion is to replace the principal diagonal of the estimated regional transactions table with zeros - hence all flows within industry groups are traded. This last suggestion occurs in response to the small and open economy of their study area (Peterborough, England), and hence is not generally appropriate.

3.424 Round's Location Quotient (RLQ)

One popular variation on the CILQ is due to Round (1978). Round suggested that the propensity to consume locally depended upon three principal elements: the relative size of the selling sector; the relative size of the purchasing sector; the relative size of the region. He suggested that a location quotient formula should capture these three effects. Clearly, from equations [3.8, 3.10], both the SLQ and CILQ are a function of the size of the selling sector i; only the CILQ is a function of the size of the purchasing sector j; but only the SLQ is a function of total regional output (regional 'size'). Hence, Round suggested the following adjustment to the CILQ formula:

$$_{R}q_{ij} = \frac{q_i}{\left[\log_2(1+q_j)\right]} \qquad [3.15]$$

The logarithmic denominator ensures that the variables, total regional and national output, do not cancel out of the normal cross-industry equation (see equation [3.11]); hence Round's formula is a function of relative regional size. Round's formula can be found in evidence in a number of studies, for example Batey *et al.* (1993).

3.425 Flegg's Location Quotient (FLQ)

One of the most recent innovations in location quotient methodology comes from Flegg *et al.* (1995). The motivation behind Flegg *et al.*'s development is their belief that Round's semi-logarithmic quotient accounts for regional size in a manner that is 'counterintuitive'. They attempt to demonstrate that, as regional size rises relative to the nation, the RLQ between two given industries falls, implying a more 'open' economy. Flegg *et al.* maintain that the opposite effect should be observed and hence supply the following modification, known as the ELQ:

$$_{E}q_{ij} = \frac{[\log_2(1+q_i)]}{q_j} \qquad [3.16]$$

However, the authors then demonstrate that the ELQ is *'a theoretically unappealing adjustment formula* vis-à-vis *the CILQ'*. Hence they offer a second formula, the FLQ:

$$_{F}q_{ij} = q_{ij}\left(\frac{[E^r/E^n]}{[\log_2\{1+E^r/E^n\}]}\right)^{\beta} \qquad [3.17]$$

where

E denotes total employment[3];

β is a parameter to be estimated.

The impressive-looking addition to the CILQ is a scale parameter which is designed to capture the effect of regional size: the smaller the region, the more open the economy,

[3] Employment is Flegg *et al.*'s choice of exposition, although no reason is given for this. Output would, presumably, be the preferred data.

hence the smaller the scale parameter. This parameter is applied to every *ij* pair of cross-industry quotients, which are then subject to the usual unitary constraint. Determining a value for the power parameter β is left to the analyst's discretion. However the authors suggest a value of between 1 and 5, with 5 applying to very small areas. A full critique of the FLQ can be found in the next chapter.

3.426 Lui, Grainger and Jaffrey

Lui *et al.*(1995) in another recent paper suggest that there are three important sources of difference between region and nation that should be quantified in the location quotient: productivity, industry mix, and regional size. The authors contend (p. 9) that:

'none of the location quotient based measurements capture the difference in productivity between region and the nation for the given sector'

From this statement one can only assume that the authors are under the impression that location quotients are constructed exclusively from employment data. Whilst a critical analysis of quotient methodology is reserved for the next chapter, the naiveté of Lui *et al.*'s analysis is more appropriately exposed here. They suggest deriving regional transactions by use of the LLQ, which they give as folllows:

$$x_{ij}^r = x_{ij}^n \cdot Y_i^r / Y_i^n \cdot q_{ij} \qquad [3.18]$$

where

x_{ij} is a transaction;

Y is GDP;

q_{ij} is the employment based CILQ.

The ratio of GDPs is intended to capture productivity differences. Surely this wouldn't be necessary if the quotients were constructed using GDP and not employment data? Furthermore, consider the derivation of regional purchasing coefficients from [3.18]. Gross regional output for each sector is presumably estimated by:

$$X_j^r = X_j^n \cdot Y_j^r / Y_j^n \qquad [3.19]$$

Dividing equation [3.18] through by [3.19] gives:

$$\begin{aligned} r_{ij} &= \frac{x_{ij}^n \cdot Y_i^r / Y_i^n \cdot q_{ij}}{X_j^n \cdot Y_j^r / Y_j^n} \\ &= a_{ij}^n \cdot q_{ij}^{*} \cdot q_{ij} \end{aligned} \qquad [3.20]$$

The quotient marked with an asterisk is unconstrained and, as such, may generate column sums of regional purchasing coefficients which are greater than unity. An examination of the comparative multipliers in Tables A1-A3 (p.21-3) suggests serious flaws in their approach.

3.427 Regional Purchase Propensities (RPP)

One of the more sensible recent suggestions is due to Stevens *et al.* (1983, 1989). The work derives from Stevens and Trainer (1976, 1980) and Park *et al.* (1981) who provide evidence to suggest that errors in import propensities are 'more significant' than errors in technical coefficients. The basis of these conclusions is severely challenged in the next chapter. However, it is on these grounds that Stevens *et al.* set about building an equation which estimates an observed set of local purchasing propensities. The authors label these purchase propensities *'Regional Purchase Coefficients'* (RPCs) - 'coefficient' being a rather unwise choice of terminology. To avoid confusion, the Stevens-type approach is referred to here as RPP (Regional Purchasing Propensity). However, the function of these RPPs is precisely the same as location quotient or commodity balance-type methodologies; equation [3.3] above equates to the RPP. The equation is derived from location theory, the basic premise being that local demand for local output relative to imports should be a function of relative delivered costs between the region and the nation. In Stevens *et al.* (1983) the following proxy-measures for relative costs are included in the equation:

$$ {}_{S}q_i = f(w_i, e_i, v_i, q_i, a) \qquad [3.21]$$

where

w is relative average annual wages per worker;

e is the regional employment share in i;

v is the weight-value ratio;

q is the employment location quotient, with manufacturing employment as the denominator;

a is relative area.

In Stevens *et al.* (1989) the number of continuous variables is reduced to two: the average US weight-value ratio for the commodity and the industry's share of national gross output. The basic principle is that, with relatively easily obtainable data, regional purchasing propensities can be generated through the estimating equation. Details of the functional form of the equation can be found in Chapters 5 and 8.

Important points to reiterate are: that the methodology is based upon theoretical and empirical evidence which suggests errors in technical coefficients are of little significance, and thus allows the use of the national technology matrix at the regional level; a single RPP is estimated for each commodity which is applied across the row of the national technology matrix.

3.43 Completion Procedures for Nonsurvey Estimators

What follows is an outline of nonsurvey procedure which proceeds the estimation of inter-industry coefficients by any of the above methods. These procedures develop a 'full' regional transactions table from the estimated purchase coefficients. Variations on these procedures undoubtedly exist although what follows captures their general flavour. The Johns and Leat (1986) study provides a reasonable practical example.

Often, it is the case that even the most basic of regional data requirements, gross output, is not available from published statistics. Under such circumstances, regional shares of national gross outputs are allocated according to some proxy measure, *i.e.* employment, under the assumption of equal labour productivity (see also equation

[3.19]). Where regional-specific data on factor returns (*i.e.* labour income, profits) are not available, most nonsurvey techniques assume that national coefficients apply. In the case of final demand vectors, for consumers, government, and investors, the basic nonsurvey procedure is to assume that the national coefficients are subject to a regional import propensity. This is generally taken as the SLQ for commodity i, which is applied in precisely the same way as for the estimation of regional purchasing coefficients[4]. Again, the nonsurvey approach to estimating gross values for these vectors is through some proxy variable's regional share.

The imports use matrix is given as the difference between the matrix of national technical coefficients less the estimated regional purchasing coefficients

$$ {}^{t}\mathbf{M}^{r} = \left({}^{c}\mathbf{A}^{n} - {}^{c}\mathbf{R} \right) \hat{\mathbf{X}}^{r} \qquad [3.22] $$

where

t, c denote transactions and coefficients matrices respectively. Where necessary, these superscripts will be used to distinguish such matrices from now on.

Exports are generally derived as a residual column, being calculated as the difference between the gross output of sector i and the sum of local demand for local production of i:

$$ E_i^r = X_i^r - \sum_j r_{ij} X_j^r - Y_i^r \qquad [3.23] $$

[4] Precisely why proponents of the cross-industry location quotient do not in general extend the principle of measuring relative supply and demand to the estimation of final demand vectors is not apparent, although see Garhart *et al.* (1996),

where

Y are other final demands.

This procedure is aimed at 'balancing' the table (ensuring row and column sums match). Generally, nonsurvey procedures provide no guarantee that the export residual will be nonnegative. Under such circumstances, setting negative elements of the exports column to zero and applying some proportional adjustment technique is a common balancing procedure (see following section on iterative procedures, in particular McMenaman and Haring, 1974).

3.44 Nonsurvey Methods : Summary and Conclusions

Nonsurvey approaches to generating regional input-output tables have been developed because the costs of the survey model are perceived to be extremely high. The characteristics which define the nonsurvey approach are that they estimate a regional model from the framework of a national input-output table, making adjustments on the basis of published regional data. In doing so, the advantages of the survey approach - in particular their accuracy and flexibility - are partially sacrificed. Indeed a body of evidence has formed which suggests that the errors associated with the application of 'classic' nonsurvey techniques are unacceptably high. This has provoked two responses.

On the one hand, researchers have sought to improve nonsurvey methods. For the mainpart, this research has been centred upon improving the function of regional trade estimators. Implicit within this action is the assumption that differences in regional and national technology do not occur to any significant degree. This assumption has attracted theoretical and empirical support.

The second response has been to develop a set of procedures which, in general, supplement nonsurvey procedures with an optimal level of survey-based information. These 'hybrid' procedures are now given consideration.

3.5 The Hybrid Approach

Lahr (1993) states with respect to the hybrid approach to regional modelling:

'...such models combine nonsurvey techniques...with superior data, which are obtained from experts, surveys and other reliable sources...'

In the context of the preceding discussion on survey and nonsurvey approaches, Lahr's definition suggests that the hybrid approach has been in use, in one form or other, since the earliest attempts to build regional input-output accounts were made. However, it is not really until the work of Schaffer (1976) and Jensen *et al.* (1979) that any formal declaration of hybrid procedure is made. Experiments as early as Evans (1954) and later, for example, by Jensen and West (1980) suggested that there should be some order of importance placed upon the estimation of the elements within the input-output table. Indeed, Jensen and West's experiments indicated that around 50% of the coefficients in an input-output table could be 'deleted' before the model's analytical function became corrupted to any significant degree. Jensen's (1980) concept of holistic accuracy within the context of the input-output table has been a key rationale for the development of hybrid procedures. Thus, broadly speaking, the hybrid procedure seeks to utilise the ready-made framework of the nonsurvey model, identify its 'important components', and subsequently target the limited pool of resources to their estimation. As Hewings and Jensen (1986 p.313) comment, this process

'...seeks to capture the advantage of the presumed higher level of accuracy of the survey method and some of the economy and speed of the nonsurvey approach.'

Therefore, probably *the* central question concerning hybrid methodology has been, and still is: *'how are important model elements identified?'*

The following section investigates how 'important' model components might be identified. A brief guide to the steps typically involved in the construction of a hybrid model follows. This is proceeded by a presentation of two methods of generating hybrid regional data - 'short-cut' multipliers, and RAS-type techniques. The RAS algorithm has wider application within input-output other than the 'regionalisation' of a national table and this is considered. Most authors classify these techniques as nonsurvey. In the context of published UK regional data however, the informational requirements would have to be satisfied by survey - hence their classification within hybrid methodology.

3.51 The Identification of Important Model Components

There have been essentially two approaches to the identification of important components within the context of the regional input-output model. The first is the 'key sector' approach which utilises relatively primitive means to identify those sectors with above average linkages. The second is the more intrusive method of computer simulation where, generally, the elements of either survey-based or randomly generated models are subjected to change, the effects of which are then observed.

3.511 The Key Sector Approach

Hirschman's (1958) seminal text provides definition on linkages. Backward linkages are those which relate to purchasing activity; forward linkages are those which relate to the activity of selling. Thus Hirschman defines the 'key sector' as one which possessess above average backward and forward linkages, and hence can be expected to be associated with growth in the regional economy.

The simplest form of key sector identification is provided by Chenery and Watanabe (1958) who focus upon the intermediate row and column sums of transactions expressed as a proportion of the respective row and column totals. This measures direct linkages. Rasmussen (1957) takes into consideration direct and indirect effects by taking the average of row and column elements in the Leontief inverse (open or closed to households). Rasmussen normalises these measures by indexing them to the average value in the Leontief inverse, thus enabling some comparison between sectors. Hence the *'index of the power of dispersion'* (p. 134) - the measure of backward linkage U_j is given by:

$$U_j = \bar{b}_j / \bar{b}$$ [3.24]

where

b denotes an element in the Leontief inverse;

- denotes the mean value in the matrix, *j* being the column mean;

The *'index of the sensitivity of dispersion'* - the measure of forward linkage U_i is given by:

$$U_i = \bar{b}_i / \bar{b}$$ [3.25]

where

- denotes the mean value in the inverse matrix, *i* being the row mean;

Thus a key sector is generally identified by U_j , $U_i > 1$. Rassmussen further complements these indices by considering the coefficient of variation (the standard deviation of any row/column divided by its mean). Thus a high coefficient of variation indicates the spread of effect is uneven, and limited to relatively few elements. A low coefficient of variation indicates a more even spread - Rasmussen cites a high backward linkage with even spread as identifying an important sector. Hazari (1970), Jones (1976), Beyers (1976) and McGilvray (1977) note the problems inherent in the use of input-output coefficients. In particular, the use of direct requirements coefficients provides an inappropriate measure of forward linkage, hence the supply driven inverse is the recommended base for their calculation (*i.e.* from the transposed transactions table, Ghosh, 1958). Some form of weighted (*i.e.* output) measure of linkage is also suggested to facilitate distinction between 'large' and 'small' sectors (see also for example Schultz, 1976; Szyrmer, 1992).

3.512 Simulation Approaches

The simulation approach to the identification of important elements in the input-output model has risen principally through a consideration of the role and nature of error transmission within matrices. In later papers, the focus is extended to the concepts of sensitivity and coefficient importance. Evans (1954) and Quandt (1958; 1959) however form the early exploratory studies. Evans shows that small uniform errors introduced to a single row of coefficients have a minor effect on the solution values of gross output. Quandt's early paper derives confidence intervals for the solution values of gross output for a hypothetical two sector model. Concern with error transmission is reflected in later papers, for example the reconciliation debate (particularly Jackson's, 1991 full distribution approach). West (1986), under an assumption of normality in the distribution of direct coefficients, derives the probability density function of multipliers, from which standard errors and confidence limits are obtained.

Jilek (1971) however is one of the first to formally recognise that the process of model construction could be made more efficient by focusing on the 'most important' coefficients of the given model. The line of research, which is later developed by Schintke and Stäglin (1988), centres upon the notion of 'tolerable limits'. Schintke and Stäglin seek the level of error that can be introduced to a coefficient before the error in the gross output solution reaches some predetermined critical level.

One of the most significant practical contributions is that of Sherman and Morrison (1950). The authors present a formula which, from the introduction of a given change in a single element of a matrix of direct coefficients, enables the elements of its inverse to be computed without requiring the inversion of the perturbed matrix. Following Sherman and Morrison and the exposition by Sonis and Hewings (1989), given an additive change, *e*, in an input-output coefficient, *a*, at location *ij*, the Sherman-Morrison formula for computing the elements of the inverse matrix is

$$b_{kl}^{*} = b_{kl} + \frac{b_{ki} b_{jl} e}{1 - b_{ji} e} \qquad [3.26]$$

where

b are the elements of the Leontief inverse matrix

$*$ denotes an element in the Leontief inverse corresponding to the changed matrix of direct coefficients.

The Sherman-Morrison formula is the basis for the Bullard and Sebald (1977) study which showed, for a model of the United States, that the analytical function of the model was sensitive only to a limited number of direct coefficients. Hewings and Romanos (1981) follow Bullard and Sebald's lines at the regional level and draw similar conclusions. Most significantly, they find that around half the direct coefficients which are identified as 'inverse important' are located in the vectors of household income-expenditure. At around the same time, Jensen and West (1980) were conducting their well known experiments which revealed that about half the direct coefficients could be eliminated from the regional model before its analytical function became corrupted, and moreover, that the larger coefficients were instrumental in preserving this function. Hence, West (1981; 1982) extends the Sherman-Morrison formula to consider the effect of change on column multipliers. For a single coefficient change West demonstrates that, for a proportional change e to a occuring at ij, the effect upon the j's column multiplier is a function not only of the size of j's column multiplier, but also of the column multiplier relating to row i. Thus, he concludes that size and *location* within the matrix determine the inverse importance of a coefficient. In general, given the proportional change e in a occurring at ij, the change in the kth output multiplier is given by:

$$M_k^* = M_k + \frac{b_{jk}e}{1 - b_{ji}e} \cdot M_i \qquad [3.27]$$

Proportional changes in the kth multiplier are clearly derived by dividing [3.27] through by M_k. West also suggests calculation of the sum of changes in all k multipliers, and the average of these. This enables a ranking of coefficients according to their relative effect on the inverse for a given proportional change. West extends the analysis to income and employment multipliers, and furthermore derives a formula which approximates simultaneous changes in coefficients. Xu and Madden (1991) consider the implications for important coefficient selection for: different importance functions; variations in the degree of perturbation; the use of absolute versus relative measures of importance. The authors' main conclusions are that an importance function which measures the change in direct coefficients upon the sum of gross outputs is preferable to other alternatives (*i.e.* measuring the effect upon the sum of multipliers) because it is more holistic in its account. Jackson (1991) arrives at a similar conclusion in favouring the use of value transactions in the importance function. Sonis and Hewings (1989;1991;1992) define the *'field of influence'* associated with the initial change in the direct coefficient where, for a change at *ij*, the first order field of influence is the matrix produced by the product of the *i*th column and the *j*th row. Their approach, which really only differs from West's (1982) in its exposition, is able to assess simultaneous coefficient change.

Hence, whilst the 'key sector' approach is necessarily broad in its direction, the 'inverse importance' approach is capable of delivering a 'shopping-list' of individual coefficients.

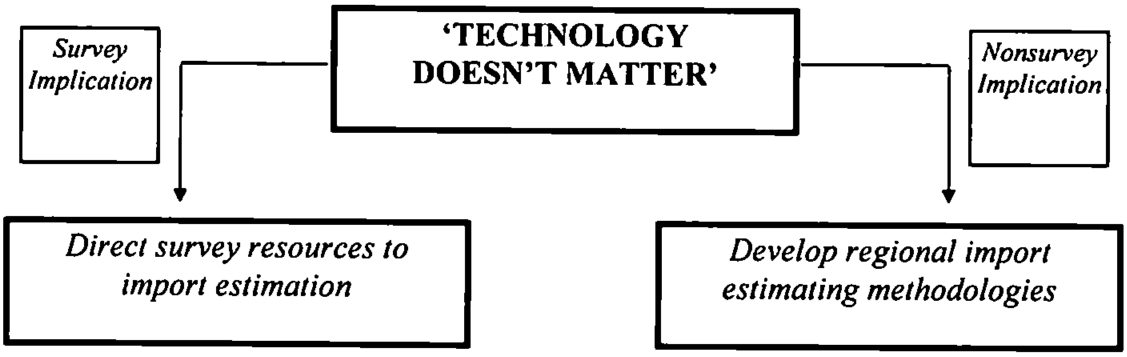
However, it is this exactitude which leads one to question whether the techniques, when applied to the context of a hybrid modelling process at the nonsurvey stage, are capable of providing a reliable guideline for the effective deployment of survey resources. Boomsma and Oosterhaven (1992) note the *Catch-22* situation that is associated with hybrid philosophy:

'Without a regional table one cannot determine the inverse-important cells and without that information one cannot construct a decent regional table.'

Hence, as Xu and Madden (1991) note, the principal use of important coefficient idenification techniques has been within the context of updating models through time (see the section on RAS below), and within structural analyses (*i.e.* Sonis and Hewings, 1992).

Does this therefore mean that the hybrid process is fundamentally flawed? Fortunately, since simulation studies have consistently shown that the household expenditure-income vectors are by far the 'most important' relative to the interindustry matrix (*i.e.* Jensen and West, 1980; Garhart, 1985; Hewings, 1986), 'holistic' guidance is at hand. It should also be clear that the approach of Stevens and Trainer (1976; 1980), Stevens *et al.* (1983; 1989) and Park *et al.* (1981) has a holistic hybrid interpretation. They prescribe directing survey resources towards the estimation of import propensities on the basis of a theoretical and empirical analysis which indicates that the implications of differences in regional and national technical coefficients are relatively insignificant. Evidence to support this claim can also be drawn from the simulations of Conway (1980) who concludes that short-term volatility of import propensities contributes significantly to forecast error. Figure 3.1 below illustrates the implications of the Stevens-hypothesis for the hybrid approach:

Figure 3.1 Hybrid Implications of the Trade-Only Hypothesis



Therefore, whilst researchers have developed a set of tools for assessing coefficient importance, as the next section shows, these are really only relevant in the latter and post-construction phases of development. The issue of importance taken from a more holistic perspective yields a potential line of inquiry: is the estimation of import propensities more important than estimating the 'technology' of the regional production functions?

3.52 Steps Outlining the Hybrid Modelling Process

Jensen *et al.*'s (1979) GRIT (Generation of Regional Input-Output Tables) methodology has probably become the most widely used set of hybrid modelling guidelines, although there are several, broadly similar, alternative frameworks (*i.e.* Schaffer, 1976; Greenstreet, 1989; Boomsma and Oosterhaven, 1992). The original GRIT principles have been updated over time (West; 1990), and formed into computer software (West, 1993). Table 3.1 below is the methodological sequence suggested by West (1990). As one can see, nonsurvey methodology forms the initial Phases I and II of model construction. However, Phase I Step 2, which adjusts for differences in technology is clearly not served well either by the UK regional data set or by the range of available nonsurvey techniques. Phase III entails inserting any existing 'superior' data (perhaps published regional trade, income). As the table shows, it is not until Phase IV that important model components are identified for closer estimation.

It should be fairly clear from this that in GRIT, the nonsurvey stage of estimation has an important role to play in the overall hybrid process: it stands as the model's foundation stone. Logic suggests that improvements in the accuracy of the nonsurvey phase feed through to the final model, and this is the rationale behind the development of nonsurvey methods.

Table 3.1: The GRIT Methodological Sequence: from West (1990)

Step Number	
	PHASE I ADJUSTMENTS TO PARENT TABLE
1	Selection of parent input-output table
2	Adjustment for updating
3	Adjustment for international trade
	PHASE II ADJUSTMENT FOR REGIONAL IMPORTS
4	Calculation of non-competitive imports
5	Calculation of competitive imports
	PHASE III DEFINITION OF REGIONAL SECTORS
6	Insertion of disaggregated superior data
7	Aggregation of sectors
8	Insertion of aggregated superior data
	PHASE IV DERIVATION OF PROTOTYPE TABLE
9	Derivation of initial transactions values
10	Manual or iterative adjustments to derive prototype table; consistency checks, analysis of sensitivity and coefficient significance
11	Derivation of inverses and multipliers for prototype table
	PHASE V DERIVATION OF FINAL TRANSACTIONS TABLE
12	Final superior data insertions and other adjustments
13	Derivation of final transactions table
14	Calculation of inverses and multipliers for final table

3.53 Short-Cut Regional Input-Output Multipliers

One of the most radical and contentious methods of generating hybrid input-output information is the short-cut multiplier approach. The approach centres upon generating aggregate column output multipliers, and therefore avoids the generation of an i x j matrix of direct requirement coefficients. Indeed, the basic approach reduces the estimation of the output multiplier for a given sector to the specification of just two parameters. The line of research can be traced through Bromley (1972) who notes the strong statistical relationship between the column sum of intermediate purchasing coefficients and the value of the corresponding output multiplier. Drake (1976) provides further development, but it is the series of papers by Burford and Katz (1977; 1978; 1981) which are principally associated with the short-cut formula. Burford and Katz demonstrated that the column output multiplier could be closely approximated by the formula:

$$\mu_j = 1 + \frac{1}{1-\overline{\omega}}\omega_j \qquad [3.28]$$

where

ω is the column sum of direct purchase coefficients and the bar denotes the mean of these column sums

Hence, the output multiplier for a sector can be derived with just the knowledge of its local purchase propensity and the corresponding average for the region. Phibbs and Holsman (1980) suggest the knowledge of principal diagonal coefficients improves the estimation process. The basic relationship can be extended to total output, income and employment multipliers with average coefficient information relevant to those factors (Burford and Katz, 1985).

Unsurprisingly, the Burford and Katz methodology has received some heavy criticism. In particular, because the approach essentially by-passes the *raison d'être* of the input-output model - namely inter-linkages, many feel that it's claim of association with the field of input-output analysis is misplaced (see Hewings and Jensen, 1985 for example).

3.54 Iterative Techniques

The problem of updating national input-output tables using minimum information-cost methods was addressed by Stone (1961). The algorithm devised by Stone requires knowledge of the gross sums of output, as well as intermediate sums of sales and purchases for both the 'base' and 'target' years. Defining the following matrices:

${}^{t}\mathbf{U}^{*0}[u_{ij}^{*0}]={}^{c}\mathbf{A}^{0}\hat{\mathbf{X}}$	: the base year transactions table
${}^{t}\mathbf{U}\left[u_{ij}\right]$	: the target year transactions table with u_{ij} unknown;
$g/2$	: the iteration count. Initially g=0
u_i, u_j	: the *known* intermediate row, column sums of ${}^{t}\mathbf{U}^{*}$
$u^{*g}{}_{i}$, $u^{*g}{}_{j}$	: the intermediate row, column sums of ${}^{t}\mathbf{U}^{g}$
${}^{t}\mathbf{U}^{*g}$	: the estimated transactions table at iteration g.

At any stage of iteration, it is possible to define an i dimensional vector of row quotients $\mathbf{r}^g$, with elements:

$$r_i^g = u_i / u_i^{*g} \qquad [3.29]$$

The intermediate row sums of the matrix $'\mathbf{U}^{*g}$ can therefore be made to match those of the target transactions matrix by:

$$'\mathbf{U}^{*g+1} = \hat{\mathbf{r}}^{g}\,'\mathbf{U}^{*g} \qquad [3.30]$$

$$g = g + 1 \qquad [3.31]$$

Intermediate column sums in the target year are known, hence it is possible to define a j dimensional vector of column quotients $\mathbf{s}^g$, with elements:

$$s_j^g = u_j / u_j^{*g} \qquad [3.32]$$

Consistency is then achieved between the column sums of the target matrix and ${}^{t}\mathbf{U}^{*g}$ by:

$$ {}^{t}\mathbf{U}^{*g+1} = {}^{t}\mathbf{U}^{*g}\hat{\mathbf{s}}^{g} \qquad [3.33] $$

And

$$ g = g + 1 \qquad [3.34] $$

completes one full iteration. The bi-proportional adjustment process of steps [3.29] through [3.34] is then repeated until convergence is achieved *i.e.* the row and column quotients defined in [3.29] and [3.30] are simultaneously equal to unity, or within some tolerable limit of unity. Hence the final matrix ${}^{t}\mathbf{U}^{*g}$ is an estimate of the true target matrix ${}^{t}\mathbf{U}$, with gross and intermediate row and column sums consistent with those of the true matrix. The algorithm is known as RAS, which derives from Stone's original notation. Bacharach (1970) and Miller and Blair (1985) show that RAS minimises the function:

$$ D\left[\mathbf{A}^{0}:\mathbf{A}\right] = \sum_{i}\sum_{j}\left\{a_{ij}\ln\left[a_{ij}/a_{ij}^{0}\right]\right\} \qquad [3.35] $$

subject to the known row and column constraints. The following points should be obvious. Firstly, as long as the elements of the base matrix and the target intermediate row and column sums are all nonnegative, RAS will yield a nonnegative solution. Furthermore, as long as the intermediate column sums are less than their respective gross sums, no individual coefficient, or intermediate column sum of coefficients will be greater than unity. Thirdly, because the adjustment process is purely multiplicative, any zero element in the base matrix will remain so throughout. Geary (1973) presents an alternative updating procedure which utilises Lagrangian multipliers, although this is not detailed here.

Stone's original work classifies the observed differences in row sums as substitution effects - those differences arising through import substitution or actual changes in the combinations of inputs in the aggregate production function that occur for what ever reason. Differences in column sums are classified as fabrication effects - for example a move to a more labour-intensive production process, hence the proportion of primary (labour) input purchased will rise in relation to the proportion of inputs purchased from industrial production. Thus, RAS has some economic foundation and is not simply a 'black box' technique. Although many commentators disagree (particularly Miernyk, 1976; to some extent Hewings and Jensen, 1986), it is possible to see how the RAS algorithm can be applied in the context of regionalising a national input-output table. With known regional gross and intermediate row and column sums, the regional purchase matrix **R** can be estimated by adjusting the national **A** matrix to meet with these constraints. Firstly, the base estimate of the regional **A** matrix in transactions can be made from the vector of regional gross outputs $\mathbf{X}^r$ and the national technical coefficients:

$$^{t}\mathbf{A}^{r} = {}^{c}\mathbf{A}^{n}\hat{\mathbf{X}}^{r} \qquad [3.36]$$

Then, in general:

$$ {}^{r}\mathbf{R} = \hat{\mathbf{r}}^{r} \; {}^{r}\mathbf{A}^{r} \hat{\mathbf{s}}^{r} \qquad [3.37] $$

It should be clear that, in the context of regionalisation, the row and column constraints embody differences in regional and national technology through substitution and fabrication effects which are in the spirit of Stone's original application of the algorithm; however, there is an additional *import substitution* effect in the adjustment process which captures the extent to which regional demand (columns) substitutes potential local supply (rows) in favour of imported supply.

One particular problem associated with the use of RAS as a regionalisation technique is that it does not allow the derivation of an imports use matrix because technology and trade are accounted for simultaneously. Specifically, it is not possible to guarantee nonnegativity in the imports use matrix given by:

$$ \mathbf{M}^{r} = {}^{r}\mathbf{A}^{r} - {}^{r}\mathbf{R} \qquad [3.38] $$

where

A, R are the matrices derived from [3.36] and [3.37]

With prior information on other primary inputs, it would of course be possible to calculate the aggregate value of import purchases for each industry as a residual (gross output, less intermediate sum, less other primary inputs).

As Hewings and Jensen (1986 p.311) point out, there is considerable scope for variation in the way in which RAS is employed. For example, supposing only percentage row and column regional import propensities are known. It is then possible to construct a RAS procedure which operates from an assumption of spatially invariant technical coefficients. This would bring RAS closer to a location quotient or RPP methodology, the difference of course being the addition of a column constraint. Given sufficient information, it would be feasible to split the RAS regionalisation process into two, or even three distinct phases of estimation. So, for example, with information on the intermediate row and column sums of the regional technology matrix, the multiplication vectors $\mathbf{r}^a$ and $\mathbf{s}^a$ could be formed and used to estimate all regional technical transactions:

$$ {}^{t}\mathbf{A}^{ar} = \hat{\mathbf{r}}^{a} \; {}^{t}\mathbf{A}^{r}\hat{\mathbf{s}}^{a} \qquad [3.39] $$

Then, with information on intermediate local sales and purchases, one could form the multiplication vectors $\mathbf{r}^m$ and $\mathbf{s}^m$ using this and the intermediate sums from the regional technology matrix. The regional purchase coefficients could then be estimated by:

$$ {}^{t}\mathbf{R} = \hat{\mathbf{r}}^{m} \; {}^{t}\mathbf{A}^{ar}\hat{\mathbf{s}}^{m} \qquad \text{s.t. } {}^{t}\mathbf{R}\left[r_{ij}\right] \leq {}^{t}\mathbf{A}^{ar}\left[a_{ij}^{ar}\right] \qquad [3.40] $$

One could go further and define a situation where the regional technology matrix is known and is adjusted to conform with aggregate information on regional import propensities. As Hewings and Jensen point out, the coefficients of the regional purchase matrix estimated in each case are conceptually and 'significantly' different (see Hewings, 1977).

It should be clear that RAS need not operate over the full (ixj-i-j+1) degrees of freedom. Coefficients which are based upon 'superior' information can be accommodated into the matrix and remain fixed throughout the process of iteration (see Szyrmer, 1989; Dewhurst, 1992). Ironically this may not guarantee a generally closer solution. An example of this phenomenon, known as the Miernyk Paradox is given in Miller and Blair (1985, p.293) One could also form column and row constraints from any given sub-matrix and perform RAS (Israilevich, 1986). Generally, as the number of degrees of freedom available to RAS are reduced, the more accurate the final solution.

Other variations on RAS include the suggestions of McMenamin and Haring (1974) who update a regional model of Southern California on the basis of gross (not intermediate) sums. Phibbs and Holsman (1982) present what is essentially an extension of Round's technology adjustment [3.13, 3.14], although they trace their work through the Burford and Katz literature. The authors generate nonsurvey regional input-output coefficients by means of normal location quotient procedures. The column sums of intermediate coefficients generated by this procedure are then adjusted mono-proportionally to conform with survey-based intermediate column sums, which is where the process ends.

Despite a continuous stream of relatively favourable reports on its ability to simulate input-output data (see the following chapter), within the context of UK, RAS has been used less as a regionalisation technique and more as a method of updating national tables prior to the application of some less sophisticated regionalisation technique (Batey *et al.* 1993; Hill and Roberts, 1996), or as a reconciliation procedure in survey-based tables (Scottish Office, 1994). De Kanter and Morrison (1978) provide some attempt to utilise the algorithm in a hybrid model of Merseyside. The reason why RAS has not been widely utilised as a regionalisation technique is, presumably, that its data requirements cannot readily be met. Hence the two important summary points with respect to RAS are that: it is capable of adjusting for trade and technology, but operates on a data set which is currently unobtainable from published UK sources.

3.55 Summary of Hybrid Procedures

Hybrid procedures seek to combine survey and nonsurvey procedures. Their aim is to achieve optimality between cost and accuracy in the construction of the regional input-output model. A number of techniques have been devised to help identify the areas where survey attention would be most appropriately applied. The sophistication of these techniques ranges from those which identify broad sectors, to those which provide a coefficient ranking. However, the hybrid process may be flawed to the extent that, without an accurate model, the important model components cannot be identified, but without these important components, an accurate model cannot be constructed. These 'partitive importance' techniques are therefore used in the final and post-construction phases of development. However, studies have consistently highlighted the significance of the household sector, and therefore more 'holistic' guidelines on importance are available. Evidence of this nature has suggested that the estimation of trade propensities is more important than technical coefficients. This seems an issue for investigation. Two hybrid procedures have been presented: 'short-cut' multipliers and RAS. The latter of these provokes most interest as a technique for 'regionalisation' in that it is the only one of the broad range of mechanical procedures to account for trade and technology simultaneously. However, because the data set required to implement RAS is not generally available within the context of UK regions, its use as a regionalisation technique has been limited.

3.6 Summary on Methods of Generating Regional Input-Output Data

This chapter has presented the broad approaches to generating regional input-output data. The purely survey approach delivers accuracy but at prohibitive cost. The purely nonsurvey approach offers a low cost alternative but necessarily sacrifices the accuracy of the model. Researchers have responded in two ways to this problem: (*i*) they have sought to improve purely nonsurvey techniques; (*ii*) they have sought to devise procedures and methods which aim to optimise the use of nonsurvey and survey methods. The former response has centred upon improving regional trade estimation, an action which is supported by evidence arising from response (*ii*) that differences in regional and national technology are insignificant. The latter response has also stressed the importance of the household sector.

3.7 Conclusion on Methods of Generating Regional Input-Output Data

The objective of the research is to improve the procedures for generating regional input-output tables. From the discussion of this chapter, it is clear that this improvement should concern hybrid estimation procedures.

There would seem to be three main ways in which one could, simultaneously, seek to improve the hybrid process. (*i*) Can the popular hybrid approach set out in Table 3.1 be improved as a concept?; (*ii*) Is the trade-only hypothesis an appropriate strategy for a hybrid approach?; (*iii*) Can the methods of nonsurvey estimation be improved?

With respect to (*i*), note that the survey-based model *embraces* the idea of an economic *database*, whilst the nonsurvey approach is purely focused upon the model's analytical function. In terms of the ultimate objective of the research -

improving the local economic data set for Devon and Cornwall - the set of hybrid procedures should make absolutely certain that the input-output table is perceived as a database framework.

Points (*ii*) and (*iii*) both centre on the 'trade or technology' debate. Point (*iii*) however raises the additional question of 'how to regionalise', *i.e.* formula parameters, and the mode of nonsurvey application.

The next chapter investigates the scope for improvement.

CHAPTER 4

A Review and Critical Assessment of the Evidence Concerning Nonsurvey and Hybrid Methods of Generating Regional Input-Output Tables

4.0 A Review and Critical Assessment of the Evidence Concerning Nonsurvey and Hybrid Methods of Generating Regional Input-Output Tables

4.1 Introduction

The motivation for developing hybrid procedures has been that the 'classic' set of nonsurvey techniques has been theoretically and empirically evaluated, and subsequently rejected.

Giarratani and Garhart (1991) identify two distinct approaches to the evaluation of regional input-output models. The first is the method of scenario simulation where, in general, the elements of a given table are manipulated and the effects observed. The second approach is the method of direct comparison between an observed and estimated input-output table. Whilst the former approach has been used to shed light on a number of evaluation issues, for example parameter sensitivity and temporal stability, the function of the latter has been limited to the appraisal of the 'classic' mechanical simulation techniques. The first part of this chapter reviews and critically assesses applications of both forms of approach which have attempted to gain some insight into the operation of nonsurvey and partial survey estimation methods.

The evidence of direct comparison studies is initially reviewed. The critique that follows the review considers (*i*) the validity of the direct comparison approach *per se*; (*ii*) how successful these studies have been in achieving their research objectives. Following this, the evidence from the alternative simulation approach which has supported the nonsurvey assumption of spatially invariant technologies is challenged. In the light of the evidence from these two methods of analysis, the second part of the chapter considers the extent to which the attempts to develop classic nonsurvey approaches have formed a useful contribution to hybrid estimation procedure. In doing so the deficiencies of the classic approaches are revealed. One is led to question whether current hybrid philosophy is the most appropriate paradigm under which to construct regional input-output tables. The purpose throughout the chapter is to further develop the research objectives and hypotheses which will form the subject of the next chapter.

4.2 The Direct Comparison Approach to the Evaluation of Nonsurvey and Partial Survey Methods

Direct comparison studies attempt to assess the ability of nonsurvey and partial survey estimation techniques to reproduce the features of a model compiled by survey. The motivation for such studies is fairly clear. The ultimate objective must be to improve the cost-efficiency and/or cost-effectiveness of input-output studies by identifying the techniques, or the opportunities for developing techniques, which best mimic the broad features of the survey based model yet are less demanding in terms of resources. Hence, either a given level of accuracy could be afforded at lower cost, or greater accuracy could be 'purchased' with a given level of resources through increased efficiency of deployment.

Studies gauge error by some measure of overall (dis)similarity between simulated and survey-based matrices. Measures focus upon simulation error in coefficients, inverse elements (and multipliers), trade flows, and solution values of gross output. The general merits of the matrix distance approach are considered in the critical analysis of this chapter. A presentation and critique of individual measures of matrix distance is however reserved for a later chapter.

World-wide, there have been numerous applications of the direct comparison approach to the evaluation of mechanical estimation techniques. In the UK however, studies of this nature are relatively few in number. This is largely a consequence of there having been relatively few survey-based regional input-output tables produced in the UK. One can identify two classic studies: that of Peterborough by Smith and Morrison (1974); Morrison and Smith (1974); and that of Scotland by Harrigan *et al.* (1980a; 1980b). These studies, in particular that by Smith and Morrison, are amongst the most well known, widely quoted, and thorough of their kind. Thus the few contributions UK analysts have made have been important. Willis (1987) provides a lesser-known study of Wales and North Staffordshire. Dewhurst (1992) also provides some evidence on the simulation of the Scottish economy. Studies of this nature have tended to be quite similar in their approach and therefore, what follows is not an exhaustive retrospection. Instead, the review traces the main developments in a broad chronological order, with particular attention to the UK evidence. However, since studies of this nature originated in the United States, it is to these pioneering works that attention is initially focused.

4.21 The Beginnings of Direct Comparison Studies

In 1969, two pairs of authors, Schaffer and Chu, and Czamanski and Malizia provided empirical evidence on the ability of mechanical estimation techniques to simulate a 1963 survey-based table of the Washington State economy (Borque *et al.*, 1967). These studies mark the first attempt to gauge the degree to which the relative accuracy of the survey study is sacrificed by the use of mechanical estimation methods. Prior to these studies, the evidence had remained largely hypothetical (*i.e.* Shen, 1960). Furthermore, the 1969 studies represent probably the first formal statement of the 'classic' set of mechanical regionalisation methods. Moreover, the studies that followed those of 1969 show relatively few developments in terms of style and application. Therefore, the 1969 studies can be regarded as having pioneered nonsurvey methods and the means of evaluation by direct comparison.

4.211 Czamanski and Malizia

Czamanski and Malizia focus very much upon the application of RAS as a regionalisation technique. Working from a 1958 matrix of direct national coefficients, they apply a three-stage process, the first of which is to 'update' the national model for differences in commodity prices between 1958 and 1963. The second stage performs sectoral aggregation which transforms the national and regional model onto a consistent 43 and 36 sector basis. Brand (1997) notes this as a necessary feature of this genre of study:

'The matrices are aggregated prior to simulation in order to create a 'level playing field'*; this enables the relative performance of a range of techniques to be assessed.'*

The final stage is the familiar one-step RAS regionalisation process which works from a national framework and adjusts for input and import substitution simultaneously. The authors apply a permutation methodology in order to isolate individual effects. So, for example, the change in the reported level of error between simulations identical but for the presence of the adjustment for relative price is attributed to that effect. Errors are observed only between coefficient matrices. Moreover, the absence of any purely nonsurvey test means that the gains from the 'superior' information set that fuels RAS are not quantified.

Unfortunately, the methods and results of the Czamanski and Malizia study are not entirely clear, and indeed, the paper tails away in its conclusion. However, it would appear (Table 1 p.71) that the price 'updating' has little effect on observed error. Aggregation would appear to contribute positively to mean percentage simulation error. The area of contention is *Case V* where it appears the authors attempt to simulate regional technical coefficients. Here the largest errors are observed, and hence the authors conclude that

'...adjustments for domestic imports do not seem to add anything to the quality of results.'

By this, the authors would seem to imply the opposite of the Stevens-type approach which focuses upon trade estimation. The authors also conclude that the largest errors in the simulation of technology are felt in Washington's most specialised sectors. However, their conclusions are generally sketchy.

4.212 Schaffer and Chu

Schaffer and Chu consider the use of three quotient techniques: simple, purchases-only, cross-industry; and two commodity-balance methods, including the supply-demand pool outlined in the last chapter. These are applied to a 23 sector aggregation of the 1958 US model, and the results are compared to the 1963 Washington survey table. No attempt is made to update the 1958 national model. The authors consider the errors in coefficients, Type I and II income multipliers, and estimated trade flows.

On the basis of chi-square, the authors conclude that only a third of sectors in the estimated matrices possess coefficients which are not 'significantly' different from their survey-based counterparts. The Chapter 6 questions the suitability of the chi-square measure within this context. In terms of estimated trade flows, notably Appendix B (p.98) reveals the inability to account for cross-hauling by SLQ and SDP as an inherent weakness of these techniques. For example, whilst all 23 commodities were exported to some extent, SLQ and SDP identified only 9. The CILQ, by this criterion, fares better. However, estimates of Type I income multipliers are found, on average, to be just over 20% higher than the survey multipliers for the SLQ and the SDP, whilst they are some 40% for the CILQ. Errors in Type II multipliers are, on average 50% higher for all techniques. This leads to the general conclusion that the SLQ and SDP are the relatively more 'accurate' simulators, although Schaffer and Chu are careful to comment that:

'...it seems that, at the moment, there is still no acceptable substitute for a good survey-based table.'

By excluding RAS from the simulation process, Schaffer and Chu fail to gauge the effect of 'superior' information on simulation performance.

4.22 Development in the UK

4.221 The Peterborough Study

Perhaps the most well known study of this genre is that by Smith and Morrison (1974), which was the first of its kind in the UK. They note the limitations of early indirect attempts to assess the performance of British regional nonsurvey models (Nevin *et al.*, 1966; Hewings, 1971; see Round, 1972 for a criticism of the latter), and set out to conduct the experiments of the two earlier American studies within the context of the UK. The source of their analysis is a survey-derived table for Peterborough, and the UK model, both of which relate to 1968. The authors achieve consistency of definition between these tables at 19 sectors.

Initially, six, mainly quotient-based, estimators are applied to the UK table. The techniques are: SLQ, POLQ, CILQ, Round's LQ, SDP, and RAS. The authors recognise that there is no single 'ideal' test of matrix distance, and therefore utilise six different measures. Perhaps the most interesting innovation within this context is the application of a simple regression model to test the validity of the nonsurvey assumption of identical regional and national technologies. Briefly, in the model represented by [4.1] below,

$$a_{ij}^{r} = \alpha + \beta a_{ij}^{n} + e_{ij} \qquad [4.1]$$

where

a is a technical coefficient;

r, n are the region and nation respectively;

α, β are parameters to be estimated;

e is a stochastic error term.

under the null hypothesis of identical regional and national technologies, the value of α and β should not be significantly different from zero and unity respectively. Smith and Morrison report an intercept equal to 7.71×10^{-4} and a slope parameter of 0.871 and conclude (p.30)

'*These results would suggest that, on the whole, the nonsurvey* [i.e. national technical] *coefficients were in fact reasonably close to the survey estimates...*'

The authors perform other tests of distance upon the observed and hypothesised technical relationships. This is an important innovation because, by considering only differences between technical coefficients, Smith and Morrison have essentially *isolated* the error that is associated with the assumption of identical technologies. However, they conclude on these measures that (p.31)

'...it is not possible to make any rigorous judgement as to whether these values are tolerable.'

Hence they deem to have found general evidence to support the null hypothesis of identical technology, and on this basis, proceed with the application of their selected 'reduction' techniques. However, as a consequence, evidence on a potentially important source of estimation error is kept to within the preparatory phase of the study and is largely forgotten in the analysis that follows.

The initial test of the relative simulation performance of techniques follows the form of the early US studies: the measure of difference is between the survey-based purchasing coefficients and those simulated by applying mechanical techniques to the national technical matrix. Intuitively, one can see that this gives a more general measure of simulation error than the comparison of technology matrices.

In comparing the relative performance of purely nonsurvey methods with a hybrid approach (RAS), Smith and Morrison bridge an important gap that is left by the early US studies. In all five distance tests RAS substantially outperforms the set of purely nonsurvey techniques. To gain some measure of the extent of its superiority, RAS generates a matrix which shares a relatively high positive correlation coefficient, ρ equal to 0.501, with its survey-based counterpart, whilst its nearest rival records a corresponding value of 0.190. Care in the interpretation of these statistics should however be made (see the Chapter 6 for details on the correlation coefficient and its validity). Over the five tests, the following order of performance by mean rank order was found

Table 4.1 Relative Performance of Estimation Methodologies, Smith and Morrison (1974)

Method	Mean Rank
RAS	1
SLQ	2.2
POLQ	3.2
SDP	4.4
Round's LQ	4.6
CILQ	5.6

However, it is important to note that the distance statistics were virtually the same between SLQ, POLQ and SDP (*i.e.* ρ equals 0.160, 0.158, 0.190 respectively). Between these and the cross-industry formulae the differences were slightly more apparent (ρ equals 0.096 for Round's quotient, 0.075 for CILQ). Principally, the analysis illustrates a fundamental difference between the simulation performance of purely nonsurvey and hybrid methodologies.

Smith and Morrison's second test of relative performance represents another innovation on the early US studies. The authors compare the survey-based trade coefficients to those simulated by the application of mechanical techniques to the survey-based technical matrix. Intuitively, it should be clear that, by excluding the use of the national technical matrix, this experiment isolates the error that is associated with regional trade misspecification. If this is so, one could gain some insight as to the relative importance of correctly specifying trade and technology by comparing the two sets of 'isolated' errors. Evidence of this nature would provide facilitating focus and direction for the development of simulation techniques. Indeed, the decomposition of simulation error has been sought by other authors interested in technique development (see for example Park *et al.* later in this chapter). However, it would seem that Smith and Morrison either view the separation of error components as unimportant, or they are oblivious to the possibility of interpreting their analysis in this way since nowhere do they perform the necessary comparative analysis.

All techniques show marked improvements by simulating from regional technology, although the ranking in Table 4.1 does not alter, and differential performance between nonsurvey techniques is minimal.

In addition to the consideration of matrix distance, the authors focus upon errors in aggregate Type I and Type II column output and income multipliers. Type I output multipliers are overestimated on average by just under 20% for the SLQ and POLQ and 25% for the SDP and cross-industry formulae. The errors in the RAS multipliers are negligible. For Type II multipliers, the errors rise to 30% for SLQ and POLQ and around 40% for SDP and cross-industry formulae. RAS errors rise to 8.8%. In terms of income multipliers, the error margins are virtually the same. There are strong similarities between Smith and Morrison's results and the errors reported by Schaffer and Chu.

In general, Smith and Morrison conclude that (p.61)

'...only the RAS approach produced consistently acceptable results..., if approximate answers are all that is needed the SLQ method... might be acceptable on some occasions.'

On the basis of their evidence, Smith and Morrison then go on to suggest improvements to conventional formulae, which were detailed in the previous chapter. Briefly, these improvements relate to the estimation of the principal diagonal of the purchasing coefficients matrix and involve (*i*) replacing cross-industry formula with SLQs; (*ii*) replacing the principal diagonal with zeros. Whilst these steps provide better simulations, they are clearly rather *ad hoc* suggestions of questionable generality.

4.222 The Scottish Study

Smith and Morrison of course study an extremely small geographical area. Peterborough at the time represented just 0.15% of the UK population (Smith and Morrison *op. cit.* p.13). If regional size is an important variable in the simulation exercise (*i.e.* Round, 1978) then evidence relating to a variety of relative spatial dimensions is desirable. Harrigan *et al.* (1980a; 1980b) therefore provide evidence for the UK's largest standard region, Scotland.

The authors utilise UK and Scottish tables for 1973 (ONS, 1978; Fraser of Allender, 1978) and transform them onto a consistent 46-sector basis - over twice the number defined in the earlier Peterborough study.

Harrigan *et al.*'s first study focuses upon differences in regional and national technology. Their second study returns to a more conventional performance assessment of various simulation techniques.

In their study of differences in technology, Harrigan *et al.* present aggregations of the Scottish and UK input-output tables ordered by the *nature* of the commodity in question (*i.e.* Simpson and Tsukui, 1965). Briefly, the commodities are ordered according to their position within the hierarchy of production: broadly speaking, this begins with manufacturing - who are mainly input demanders - and flows through to services - who are mainly input suppliers. This creates a triangular matrix which reveals broad structural similarities between economies and generally facilitates comparison. Through this approach, Harrigan *et al.* are able to show that UK and Scottish economies exhibit similar characteristics.

However, the authors attempt a quantitative analysis of the differences in technology by repeating the regression analysis of Smith and Morrison for Scotland. They extend the early analysis to test whether, under the laws of statistical inference, the estimated model parameters support a null hypothesis of identical technologies. Indeed, Harrigan *et al.* record results which are very similar to those of Smith and Morrison, with α equal to 4×10^{-4} and β estimated at 0.882. Applying a joint test of the null hypothesis $\alpha = 0$, $\beta = 1$ they are able to reject the assumption of identity between technical coefficients at the 1% level of significance. Thus, whilst broad similarities exist between regional and national technology, the authors conclude that (p.806):

'If nonsurvey or mixed survey and nonsurvey methods are to be used to generate regional coefficient matrices, they will need to be sufficiently sensitive to capture certain essential features of regional structure.'

In their second study Harrigan *et al.* test the performance of several coefficient estimators: SLQ, CILQ, the SLQ-adjusted CILQ, Round's CILQ, the commodity balance approach, and RAS. They employ seven measures of matrix distance and also consider the simulation of column sums of coefficients, output multipliers, and intermediate outputs. Once again, the results reflect strongly those of previous studies, Type I output multipliers are overestimated by the purely nonsurvey techniques by around 20%. Although the adjusted CILQ is the 'best' of the purely nonsurvey estimators, which contradicts with earlier evidence, Harrigan *et al.* point out that differences between the purely nonsurvey techniques are marginal; the principal distinction lies between the relative performance of the nonsurvey set and RAS. However, they do note (p.932) that their distance measures indicate substantial estimation errors in individual coefficients under the RAS simulation.

Harrigan *et al.* consider the possibilities for improving nonsurvey techniques. In particular, they consider Round's fabrication adjustment to technology (see the previous chapter) and the simulation of purchasing coefficients from fully specified Scottish technology *i.e.* the isolation of the error due to trade misspecification. Whilst these results show an improvement in simulation performance - and in the latter case, this improvement is marked, the authors conclude that, since the error associated with trade simulation is approximately the same as that associated with the RAS procedure, methodological developments should (p.936)

'remain concentrated on RAS or RAS amended techniques.'

4.23 Miscellaneous Studies

A number of later studies follow the general lines of these classic approaches. General 'copy-cat' studies, for example by Eskelinen and Suorsa (1980), do little more than confirm earlier observations on the poor performance of classic nonsurvey techniques. Sawyer and Miller's (1983) study is about the best of these later studies. The authors base their analysis upon the 1972 table for Washington and the 1967 US table. No attempt is made to make these tables consistent by time period prior to simulation. The authors compare the performance of SLQ and SDP against RAS with an attempt to gain some insight as to (*i*) the effect of Round's fabrication adjustment for technology upon the estimation performance of the SLQ; (*ii*) the effect of aggregation on simulation performance. Distances between purchasing coefficients and Type I and II multipliers are computed.

Sawyer and Miller's analysis confirms the general superiority of the RAS procedure in the estimation of multipliers. However, the purchasing coefficients, as simulated by RAS show, on average, a 50% absolute deviation from the survey-based coefficients.

The authors show that the dimensions of the technical matrix prior to simulation has some influence on SLQ error. Simulating from a 255 sector technical matrix as opposed to a 28 sector model generally reduces simulation error. Indeed the 28 sector simulation produces errors in Type I multipliers which are of the same magnitude as the earlier studies. However, the principal finding of the Sawyer and Miller study is that by using Round's adjustment for technology, the simulation performance of the SLQ can be brought into line with that of RAS. Although this would appear perhaps a slightly extravagant claim, given the simplicity of Round's adjustment (see later), it is further evidence to suggest that the modelling of regional technology should be an important objective.

Returning to the UK context, Willis (1987) attempts to simulate the 1968 survey table for Wales (Ireson and Tomkins, 1978) from the 1968 UK table and the 1977 table for North Staffordshire (Proops *et al.*, 1981) from the 1974 UK table. Willis applies the SLQ and RAS methodologies. The study is poorly executed - indeed, the results of Tables 2 and 3 (p.111-112) suggest that there are some serious methodological errors in the analysis. For example, RAS based Type I output multipliers for Staffordshire share a correlation coefficient of *negative* 0.02 with their survey-based counterparts whilst for the simulation of Wales, the correlation is recorded at positive 0.66. Not only is the difference in correlation coefficients between studies mysterious, but the figure for Wales is low enough to arouse suspicion (the Burford and Katz analysis shows that known column sums of coefficients generate near 'perfect' column sum output multipliers; all previous comparison studies show the measured error in RAS output multipliers to be close to zero). Willis' evidence can therefore be discounted.

A number of other studies focus upon the effect of feeding additional information to the RAS algorithm, for example Szyrmer (1989) and in the UK context Dewhurst (1992), but these are not considered here as the principal interest is with nonsurvey methods.

4.3 A Critical Consideration of the Direct Comparison Approach and its Evidence

The previous review has concentrated upon the main contributions to the direct comparison approach to the evaluation of nonsurvey methods, and has focused in particular on UK evidence. In turning to a more critical consideration of these studies, one can identify two areas where criticism could be directed. The first is towards the method of direct comparison *per se*. The second is to how far the evidence of these studies has met with the objectives one might reasonably expect of such an exercise.

4.31 The Intrinsic Validity of the Direct Comparison Approach

A number of authors have questioned whether the survey model represents an appropriate testing ground for alternative estimation methodologies (*i.e.* Jensen, 1980; Jensen and MacDonanld, 1982; Round, 1983). The basic argument is that direct comparison studies treat the survey based model as the 'truth' where, as Jensen and MacDonald (*op. cit.* p.35) comment

'...since any analyst with experience in the preparation of survey-based tables would make rather modest claims for the detailed accuracy of these tables, such an accuracy test must be largely inconclusive.'

The authors subsequently make a stronger attack on the approach

'...these experiments were doomed at conception; if they had produced regional tables which were "close" to the "genuine" regional tables, we

would probably expect coincidence rather than input-output logic as the cause.'

Does this then mean that the direct comparison approach is invalidated *per se*? Jensen and MacDonald's criticism would seem harsh for a number of reasons. Firstly, whilst the assumption of 'truth' in the survey based model may be implicit within the design of the direct comparison experiment, one cannot seriously believe that the analysts were not aware of this fact. Indeed, that such studies have largely been concerned with reporting results at a matrix wide level as opposed to, for example, a cell-by-cell analysis would suggest that this was the case. A regional table which has been based upon survey work *must* reflect real features of that region's economy at a broad level, otherwise the input-output model is lost as a practical exercise. If it does reflect these broad features they must be observable, and if they can be observed the method of direct comparison has validity.

As for Jensen and MacDonald's second point, the suggestion is that those responsible for direct comparison analysis were motivated by the belief that the need for survey work could be eliminated by their experiments. Any analyst who held such expectations would indeed have to be charged with *'a real boost of professional optimism'* (Jensen and MacDonald, *op. cit.* p.35) but if one's objectives were more modest, such as gauging broadly the benefit of an improved data set, then there is no reason to believe the choice of the direct comparison method should necessarily result in failure. Therefore, as long as the analyst is aware of the inherent limitations of such an exercise, there can be no serious objections.

However, the direct comparison studies can be criticised if they have been a source of misdirection through poor execution and practice. More generally, have the direct comparison studies achieved the set of objectives that might reasonably be expected from them as an approach to analysis?

4.32 The Extent to Which Direct Comparison Studies Have Achieved Their Reasonable Objectives

So what are the realistic objectives of such a study? All authors in the area recognise the potential benefits of employing less resource intensive estimation methods. Hence, the primary objective of the direct comparison study must be to facilitate the procedure of constructing an input-output model by easing the resource burden of the exercise. This it does through identifying the most appropriate low-cost estimation techniques in current use, promoting their application, and noting the inevitable deficiencies of such techniques in order that they may be improved in the future.

Thus it is possible to identify the stages and features of a direct comparison study which would be necessary to meet these objectives effectively. Four stages can be identified:

(*i*) Technique Selection

(*ii*) Testing and Observation

(*iii*) Interpretation and Explanation

(*iv*) Recommendation

It is fundamentally important to realise that the implied empirical exercise *i.e.* the actual act of measuring nonsurvey simulation error from comparison with the survey matrix, is only one way in which the above objectives could be fulfilled. Indeed, it would seem a matter of professional expediency to precede such an exercise with a thorough consideration of the estimation issue at hand, the theoretical validity and suitability of the proposed simulation and measurement techniques, and an *a priori* assessment of each techniques' estimation behaviour. These functions form the first, and most important stage of the process.

In stage (*ii*) the direct comparison study begins, and through the application of distance measures, the analysis would, for example, be able to make some observation on the relative simulation performance of the selected methods.

But the study must also offer an explanation of why, for example, technique *A* outperforms *B*. Where do *A* and *B* differ? Is *A* more resource intensive than

B? Is *A* based upon more realistic assumptions? This is the third stage of the process.

Through this the direct comparison study not only aids technique selection from the currently available set, but it also prepares the foundations for technique development - for example, understanding precisely why technique *B* fails to perform is likely to be an important factor in the development process. Hence, in the fourth stage, the study should be able to make some recommendations on which of the available techniques should be commonly applied, weighing the relative resource cost of each against its simulation performance, and where possibilities for future developments lie.

So how have direct comparison studies fared against this model? The simple fact of the matter is that the first and most crucial stage of the process is largely missing from every example. There is an almost wholesale failure to question whether the selected estimation techniques have any theoretical grounding: studies tend to concentrate on describing the method by which each technique is applied. Fairly deep consideration however is given to the properties of various matrix distance statistics! Furthermore there is a failure to provide any *a priori* consideration of each technique's estimation behaviour. This generally means that the nature of the estimation problem is never clearly defined.

Failure to complete this fundamental level of analysis has, it will be argued, had the following effects which, collectively, have promoted poor practice in the process of estimating regional input-output models. Direct comparison studies have:

1. Identified an inappropriate and arbitrary initial set of nonsurvey techniques;
2. Promoted these methods into conventional use;
3. Made unrealistic recommendations on technique selection;
4. Damaged the process of technique development by:
   (*i*) failing to clearly identify and define the estimation problem;
   (*ii*) promoting inappropriate methods of technique development.

The initial offenders are Schaffer and Chu who present the first nonsurvey direct comparison study. No consideration is given to the estimation issues at hand; hence it would seem difficult to imagine how one could devise a methodology to tackle an undefined problem. Nevertheless three 'families' of techniques appear, which are presumably those which were in practical use at the time. However, there is no consideration as to whether these techniques make economic sense, nor as to how they might perform within the proposed context. Effectively, Schaffer and Chu take a set of *ad hoc* techniques and throw them against a survey model in the hope that one will stick. Whilst their concluding comments convey the right message (p.96)

'...there is still no acceptable substitute for a good survey-based study.'

the damage had already been done. An identifiable set of nonsurvey techniques had been created, almost without thought. It is not until well over a decade later, well after the studies of Smith and Morrison and Harrigan *et al.*, that Round (1983), in a theoretical review of nonsurvey methods, concludes (p.209):

'...many nonsurvey methods are sadly lacking in theoretical ... underpinnings, which tends to bring the whole approach into disrepute.'

Later still, at the beginning of 1990s, Jensen (1990 p.17) observes:

'We have, to date, no extensive debate or argument on the logical/theoretical properties of the nonsurvey methods. We have relied on a general level of acceptability of quotients in their different forms without extended discussion of the circumstances in which we can expect, in logical terms, variations of quotient types to produce acceptable results, based on logical analysis.'

Thus, if these facts had been established by Schaffer and Chu in 1969, it would hardly have been necessary for them to prove empirically the failure of their techniques. Perhaps if they had considered the nature of the problem more carefully, they would have been able to make some more appropriate suggestions, and current input-output practice would stand much improved. As it was, events that followed meant that Schaffer and Chu's arbitrary set of techniques escaped a deep theoretical examination and subsequently became recognised as established nonsurvey practice.

Hewings' (1971) rather cavalier proclamation that nonsurvey techniques showed promise can have only contributed to this event, and whilst Round (1972) was quick to discredit Hewings' analysis, it was largely on the grounds that he had nothing to compare his nonsurvey results against, and not of the merits of the estimators *per se*.

Since both Schaffer and Chu and Hewings had demonstrated, or one might even say advertised, the feasibility of constructing a regional input-output model using these methods, it is undoubtedly the case that, away from the academic field, location quotients and the like were finding application with increasing frequency. Thus, by the time Smith and Morrison's study was published in 1974, Schaffer and Chu's ideas were probably widely known as conventional nonsurvey practice.

Smith and Morrison justify their research by arguing that the studies to date had been concerned with US regions. However, it is a peculiar logic that motivates this exercise *i.e.* it is known that these techniques do not work in America, we would like to see if they do not work here. Whilst elements of their study are inventive, Smith and Morrison pass up on a golden opportunity to strike at the foundations of Schaffer and Chu's nonsurvey methodologies. Instead, they work almost entirely within the original framework, right through to their suggested 'improvements' which, in the spirit of their predecessors are fairly arbitrary. Furthermore, the nature of Smith and Morrison's suggestions encourages the idea that the way to improve nonsurvey methods is to 'fiddle' with them - an approach which can be traced to Teibout's POLQ and later to

Round's semi-log CILQ. This is perhaps the most regrettable feature of nonsurvey history since by making minor 'corrective' adjustments to conventional formulae, the implication is that they are generally well founded. It suggests that there *is* a general nonsurvey answer which can be encapsulated in a mathematical formula or algorithm, whereas in truth, nonsurvey's most potent weapon is probably expert judgement (*i.e.* Miernyk, 1976). The Schaffer and Chu and Morrison and Smith formalisations push nonsurvey thinking away from the conjectural approach. It is a sad fact that, in 1995, UK researchers *still* are blindly 'fiddling' with location quotient formulae (see the criticism of Flegg *et al.*, 1995 that appears later in this chapter).

By the time of Harrigan *et al.*'s study, even the most damning deconstruction of nonsurvey methodology would have probably gone unheeded, because with the publication of Jensen *et al.*'s GRIT procedures in 1979, nonsurvey methods had found a safe-haven which would allow them to be used with respectability.

However, Harrigan *et al.*'s study is just one of a number to offer what is hollow advice to the practitioner faced with a limited budget (p.936):

'...it would seem that survey work would be best directed towards providing the data required by the RAS technique...'

If one recalls, RAS operates upon known row and column intermediate and gross sums of transactions. Nothing approaching this data set is provided by central sources in the UK, even for its principalities. In light of this, three points make RAS an unsuitable candidate for selection by the regional model builder. Firstly, even for a small region, it is probably a very large and expensive survey that seeks to generate gross and intermediate sums for every sector of the economy. Secondly, it is known from the reconciliation debate (*i.e.* Jensen and McGuarr, 1977) that firms have a relatively poor knowledge of the destination of their sales, particularly on a regional basis (although Boomsma and Oosterhaven, 1992 contend this). Therefore, a survey whose only focus was upon generating RAS data may be subject to significant error in its row constraint estimates; and of course there would be no means of cross-checking this fact from purchase data. Finally, if one is purchasing/ constructing the sample population, printing the survey forms, paying the postage, paying people to input data *etceteras*, is RAS cost effective? With such an outlay, mightn't one just as well make a full purchase enquiry?

However, the impracticalities of the advice on RAS is clearly reflected in the subsequent actions of practitioners: a RAS regional data set has never been compiled for a UK region. Faced with little alternative, practitioners have fallen back on the original nonsurvey procedures, and with a sparse - even dubious - supplement of 'superior' data, they have been able to pass them off as hybrid tables under the GRIT label (see for example Johns and Leat, 1986).

Moreover, the fact that RAS consistently outperformed nonsurvey techniques should surely have acted as catalyst and guide for the development of nonsurvey methods. Explaining the superior performance of RAS purely as a function of its superior information set is, again, hollow information. To what extent is RAS's success due to its superior assumptions (*i.e.* allowing trade *and* technology substitution)? What is the possibility of modelling these features? If these issues had been explored within the framework of the direct comparison study one cannot help feeling that the understanding of the estimation problem and the proceeding developments to nonsurvey methods would stand much improved.

4.33 Conclusion

Whilst direct comparison methods have intrinsic limitations these would not seem to preclude their application. However, the studies which have been conducted have largely failed to achieve their implicit objective of improving estimation methodology. Indeed, there is a strong case for arguing that they have served to impair the process of technique development.

In general, they would seem to have generated plenty of empirical evidence on the nature of the estimation problem, but the overall picture is not clear. What is the most important feature to estimate correctly: trade or technology specification? Directly comparable evidence is not produced. Indeed, little recognition is given to the fact that the relative importance of trade and technology misspecification may have relevance, which is a good indication that the estimation problem is never defined with any clarity or concision. The relative effect of trade and technology misspecification *is* however an important point of guidance to technique development, and this is reflected by the fact that approaches other than the direct comparison method have sought to gain an insight into its effect. It is to the evidence of these studies that attention now focuses.

4.4 Trade or Technology? : The Scenario Simulation Approach to the Evaluation of Nonsurvey Methods

In the previous section it was argued that direct comparison studies had failed to give due consideration to the relevant nonsurvey estimation issues. In contrast, the approach typified by Stevens and Trainer (1976) seeks initially to establish some order of importance to the components of the estimation process by experimentation. Generally, the approach introduces hypothetical estimation errors into the relevant components of a survey based input-output model in a manner which allows their relative influence to be assessed. Importance is judged in terms of the degree to which a given level of error corrupts the model's analytical function. Once the most important features have been identified the aim is to use this information to develop an appropriate estimation technique. The Stevens approach would therefore seem to be more soundly base in logic than the examples of direct comparison.

There have been two principal conclusions from these studies. Firstly, that errors within the vectors of household incomes and expenditures, and in particular the latter, have the most distorting influence on the model's function. The experiments of Stevens and Trainer (1976; 1980), Jensen and West (1980), Hewings and Romanos (1981), Park *et al.* (1981), Garhart (1985), Hewings (1986), to name but a few, report this fact. Whilst these studies have established this through observation following simulation, there are some good *a priori* reasons for anticipating the result. Firstly, in terms of the incomes row, for the vast majority of sectors this will be their largest input purchase: income coefficients typically make up one third of gross input (ONS, 1994); the forward linkages of the labour row are therefore almost certainly stronger and more evenly spread than any tradable commodity. The consumers' expenditure vector will similarly exhibit strong backward linkages. Firstly, in contrast to any industry's purchase pattern, consumers of course do not contribute to value added (*i.e.* profits and labour income); hence a relatively large proportion of their expenditure goes directly upon commodities. Consumers also purchase a wide range of products, thus compared to any industry, their expenditure is relatively evenly spread. Thus, with strong, evenly dispersed forward and backward linkages, it is hardly surprising that households are identified as a key sector. As to why the consumers' expenditure coefficients are more important than those of the incomes row, the reason is fairly obvious: the expenditure coefficients are employed in the distribution of *every* income coefficient; thus the transmission of errors appearing in expenditure coefficients will, in general, be more widespread than in the incomes row. Therefore there is no particular argument with the findings of the simulation studies here since they stand up to logical examination. The estimation of the household vector should automatically find a place in any study which sought to improve estimation performance.

However, the work of Stevens and Trainer and Park *et al.* has attempted to go further in the process of identification. Their claim is that errors in estimating technical coefficients are much less significant than equivalent errors in estimating local purchasing propensities. Essentially, their hypothesis is that there is a difference in the transmission of trade and technology errors: the latter are far less serious than the former. This has implications for the hybrid process. From the survey perspective, it implies directing resources towards regional trade estimation. From the nonsurvey perspective, development should concern improving regional import estimators; and indeed they develop a technique which focuses purely upon the estimation of the local purchasing propensities (the RPP of the last chapter). One would expect to find little objection to such an approach since it would appear to follow a sensible order. Indeed, the simulation evidence would appear quite conclusive. Consider the study of Park *et al.* for example.

4.41 Evidence Supporting the Trade-Only Hypothesis : Park *et al* (1981).

Park *et al.* begin by defining the regional input-output specification as follows:

$$\mathbf{x} = \hat{\mathbf{p}}\mathbf{A}\hat{\mathbf{x}} + \hat{\mathbf{p}}\mathbf{f} \qquad [4.2]$$

where

x is the vector of gross outputs;

A is the matrix of regional technical coefficients;

f is a vector of final demands;

p is the vector of local purchasing propensities, $0 \leq p_{ij} \leq 1$

^ denotes a diagonalised vector

The objective is to determine the relative importance of errors when introduced to the components **p** and **A**. The simulations are carried out on a 39 sector model of Utah. Matrices of percentage errors are generated from a normal distribution with mean zero and standard deviation equal to half the percentage error, which is defined at 10, 20, 30 and 40 percent. These errors are applied multiplicatively to the coefficients in question in a manner which allows the relative effects to be isolated. Thus, for example, errors are introduced exclusively to the **A** matrix of interindustry technical coefficients; then exclusively to the **p** vector of regional purchasing propensities. Using a combination of distance measures and regression analysis the authors find that errors in the vector **p** account for above 40% of the simulation error, whilst errors in the matrix **A** contribute little at all, and they are therefore able to conclude (p.335) *'that errors in the multipliers and output are caused much more by errors in the regional purchase* [propensities] ... *than by those in the technical coefficients.'*

It is from these conclusions that Stevens *et al.* (1989) for example justify developing an equation which generates regional purchase propensities for use with national technical relationships.

4.42 Challenging the Trade-Only Hypothesis: An *a priori* Consideration

But there is something inherently odd about the evidence of the Stevens and Park *et al.* simulations. What are the *a priori* expectations about the relative importance of errors in trade propensities and technical coefficients? It is an exercise which Park *et al.* neatly avoid by presenting a highly complex algebraic exposition of the error function:

'The complicated form of the error function involving both true values and errors ... does not permit easy generalisation of the relative importance of different types of error.'

It would appear that the reams of matrix algebra obfuscate one simple fact, namely that the regional purchasing coefficients, r_{ij}, share a multiplicative relationship with the technical coefficients a_{ij} and the local purchasing propensities p_{ij}, *i.e.*

$$r_{ij} = p_{ij} a_{ij} \qquad [4.3]$$

Therefore, with a given percentage error e_{ij} it should not *matter* whether this is applied to the technical coefficients or to the purchasing propensities: the estimation error in r_{ij} will be precisely the same. How then do Stevens and Trainer and Park *et al.* generate results which display a difference between the two effects? The root cause is that the local purchasing propensities which they apply in their simulations are highly stylised and have the effect of generating a *systematic* error in the estimated **R** matrix. Instead of defining an *i* x *j* matrix of local purchasing propensities, the authors assume a constant purchase propensity applies to each *i*th row by specifying a *vector*, **p**. Hence, when the errors are applied to the technical coefficients matrix, the **R** matrix is:

$$\mathbf{R}_{e(\mathrm{A})} = \begin{bmatrix} e_{11}a_{11}p_1 & \cdots & e_{1j}a_{1j}p_1 \\ \vdots & \ddots & \vdots \\ e_{i1}a_{i1}p_i & \cdots & e_{ij}a_{ij}p_i \end{bmatrix} \qquad [4.4]$$

When the errors are applied to the vector **p** the **R** matrix becomes:

$$\mathbf{R}_{e(\mathrm{p})} = \begin{bmatrix} a_{11}p_1e_1 & \cdots & a_{1j}p_1e_1 \\ \vdots & \ddots & \vdots \\ a_{i1}p_ie_i & \cdots & a_{ij}p_ie_i \end{bmatrix} \qquad [4.5]$$

In [4.4] each of the *ij* elements is subject to an independent stochastic shock; in [4.5] there are only *i* stochastic perturbations applied to *ij* elements. This is without doubt the source of difference. Garhart (1985) recognises the potential bias and repeats the experiments - although his analysis is also questionable (see below). His Table 1 (p.360) confirms the statistical equality of error implied by [4.3] above by Student's *t*-tests. The explanation as to why the constant propensity biases the error upwards is probably because there is greater self-cancellation of error terms in the Leontief inverse derived from [4.4] than in [4.5].

Whatever the reason, the two situations are clearly not comparable. The assumption of a constant row consumption propensity is clearly a simplification (see Ralston *et al.*, 1986 for empirical evidence) of the true trading function - and the type of suggested simplification one might expect to *result* from a study which found the estimation of trade propensities to be more important than technical coefficients! To go on and use this evidence as supporting the application of row-constant RPPs is nothing short of audacious. Thus whilst the Stevens type approach seems grounded in logic, its execution is fundamentally flawed. Moreover it would seem they fail to appreciate that the error in their analysis is catastrophic, as they comment in their 1989 paper:

'There seems to be general agreement (Stevens and Trainer, 1976, 1980; Park et al., *1981) although there is some dissent (Garhart, 1985) that the accuracy of the regional purchase* [propensities] *is the most crucial factor in determining any regional input-output table.'*

Garhart's analysis offers further evidence by considering an additive as opposed to a multiplicative error structure. *A priori* one can determine from equation [4.3] that:

$$\begin{aligned} r_{ij,e(A)} &= \left(a_{ij} + e_{ij}\right)p_{ij} \\ &= a_{ij}p_{ij} + e_{ij}p_{ij} \end{aligned} \qquad [4.6]$$

$$\begin{aligned} r_{ij,e(P)} &= a_{ij}\left(p_{ij} + e_{ij}\right) \\ &= a_{ij}p_{ij} + e_{ij}a_{ij} \end{aligned} \qquad [4.7]$$

And since, in general, $p_{ij}>a_{ij}$ the errors applied to technical coefficients will result in greater simulation error. This is precisely what Garhart finds. However, it is hard to see why this result has much relevance to the issue of relative importance. Clearly, every additive error can be expressed as a proportional value, except in the case where either or both a_{ij} and p_{ij} are zero. Placed within the context of nonsurvey estimation, the case where the 'true' regional coefficient is positive and the corresponding UK coefficient zero - the only instance where one cannot define a multiplicative error - is unlikely to be the general one. Consequently, Garhart's analysis makes a rather obvious comparison between two matrices where the induced error in one is bigger than the other.

4.43 Conclusion

There would appear to be no *mathematical* reason for assuming that a given level of error when applied equally to trade propensities and technical coefficients should result in any difference in simulation error. Consequently it would appear that the Stevens-type simulation approach is not the appropriate framework for establishing the relative importance of trade and technology misspecification. It would seem that the issue must be settled by other means. It should be possible to gain some *a priori* insight into the nature of trade and technology misspecification error and its determinants. Ultimately however, the potential levels of misspecification would have to be assessed by observation from empirical data - in other words a direct comparison approach is required. The very nature of regional input-output data means that suitable sets are unlikely to be in abundance - thus the problem that one immediately anticipates is proving the generality of any empirical analysis. The direct comparison study of Smith and Morrison (*op. cit.*) would however appear to indicate that it is at least feasible to separate out the two sources of misspecification. It should therefore be possible to perform some direct comparative analysis in order to gain a narrow assessment each component's relative importance, and this judgement could then be linked back to the *a priori* analysis in an attempt to form some more general conclusions. Of course, any such analysis would have to take into account the frailties inherent within input-output data.

4.5 Recent Developments on Classic Nonsurvey Methods, and Hybrid Procedures: An Assessment

The preceding sections of this chapter have painted a rather dim view of nonsurvey methods. This section considers whether the developments of the classic nonsurvey techniques are likely to have overcome their basic deficiencies. The debate is particularly focused on recent UK innovation.

4.51 Initial Developments of Classic Nonsurvey Techniques

All of the classic approaches fail to consider what, on the evidence presented so far in this chapter, is probably a significant source of difference between region and nation - namely differences in technology. Various suggestions have been made as to why spatially invariant technology is unlikely to be observed in practice. Smith and Morrison (1974, p.22) cite anything from differences in productive efficiency to climatic variation as contributing to the phenomenon. A more formal consideration of the nature of differences in technology is given in the next chapter.

In the light of the evidence on the significance of technology differences, it would seem hard to justify any development which avoids the issue. However, of the suggested developments to classic methodologies, it is only Round's (1972) column adjustment for differences in value-added which attempts to make any physical adjustment to the technical coefficients. Round's adjustment is of course extremely simplistic in that it accounts only for regional substitution between value added and the purchase of inputs (hence the distribution of input purchases is adjusted by a single scalar). Substitution, if it occurs at all, is likely to arise between input purchases - driven by forces such as relative prices. Whilst Round's formula is valuable in that it is the only contribution to have tackled the issue of technology, it is probably too simplistic to be of any great use.

Of the developments to trade estimators, if one begins with Teibout's (1967) Purchases-Only Location Quotient, Table A.6 in Smith and Morrison (1974) reveals that the POLQ makes almost no difference to the value of the location quotient. The POLQ will therefore suffer the same trade misspecification as the SLQ and SDP techniques. It will require an assumption of spatially invariant technology - regional specialisation must result in trade and not in input substitution for it to be a valid indicator of trade orientation. Moreover SLQ and SDP fail to account for cross-hauling, and this results in the classification of a significant proportion of a region's commodities as self-sufficient in supply.

Round's semi-logarithmic CILQ claims to offer an improvement on the basic cross-industry formula by correcting its failure to account for relative regional size. Regional size is identified as an important variable affecting trading propensities because, generally, as regional size diminishes, the economy becomes less diverse and hence less able to meet given input demands. If one recalls from the last chapter, the value of the CILQ, [4.8] below, is not a function of the region's share of total output and is therefore unrelated to its relative size.

$$q_{ij} = \frac{X_i^r / X_i^n}{X_j^r / X_j^n} \qquad [4.8]$$

Round's formula introduces a logarithmic expression to the denominator of the CILQ. Whether or not this has the intended effect has been the subject of the most recent location quotient debate, and this will be considered in the next section. However, since Round's formula is still cross-industry based, if the hypothesised relationship between selling and purchasing sectors is misspecified, then Round's formula must suffer too.

A brief glance at the basic cross-industry formula reveals that its properties are perhaps contrary to the more intuitive ideas of how a region's industrial structure is formed. Consider the local import propensities hypothesised by the CILQ for the given SLQs in Table 4.2 below.

Table 4.2 : Estimated Cross-Industry Import Propensities

$SLQ_{i,j}$	0.25	0.5	1.0	2.0	4.0
0.25	0	0.5	0.75	0.875	0.938
0.5	0	0	0.5	0.75	0.875
1.0	0	0	0	0.5	0.75
2.0	0	0	0	0	0.5
4.0	0	0	0	0	0
Average	**0**	**0.1**	**0.25**	**0.425**	**0.613**

As one can see, the average propensity to import rises with the specialisation of the purchasing sector. In other words, a region's most specialised sectors are assumed to have the weakest indigenous purchasing links. But would this sort of relationship really be expected? Are regional specialisations accurately depicted as the overgrown, almost misplaced sectors of the economy? Or is it more likely the case that, in the spirit of agglomeration and locational economies (*e.g.* Moses, 1958), the needs of strong regional specialisations have, over the period of time in which they have developed, formed some mutual role of support between the local service sector, and shaped the local economy by attracting relevant input suppliers? In other words, mightn't one expect some *negative* relationship between regional specialisation and relative import propensity? The degree to which the cross-industry formula exhibits the relationship between specialisation and import propensity displayed in the hypothetical illustration of Table 4.2, and the extent to which this behaviour is realistic is a matter for empirical investigation. It would seem fundamentally important for anyone in the process of developing some form of cross-industry specification to have at least identified this feature of its behaviour and carried out the necessary empirical validation before recommending some improvement. Round is just one to fail in this respect.

As it was argued earlier in this chapter, nonsurvey methodology probably began to go wrong with the publication of Schaffer and Chu's, 1969 paper which marked a move away from the informal, judgement-orientated approaches of Isard and Keunne (1953) to the strictly formula-based world of the location quotient. It seems that this move extinguished what were, and *still are* nonsurvey's greatest assets: common sense and expert opinion. This former ingredient is certainly lacking in the initial developments of nonsurvey methods. Authors have made marginal adjustments to formulae which were *'sadly deficient'* where real improvement would have required a lateral reassessment of the nature of the nonsurvey approach. Regrettably, as the next section will illustrate, research effort is still being directed to the business of 'tinkering' with original formulae.

4.52 The Most Recent Contribution: the FLQ

Flegg *et al.*, 1995 reject Round's semi-logarithmic location quotient on the grounds that its properties are counterintuitive. Their suggestion is that Round's quotient will generate higher import propensities for larger regions and vice-versa. It can be shown that their thinking is flawed. Whilst, as it has already been expressed, there are more fundamental objections to the general approach of 'fiddling' with formulae, it would seem necessary to make the following points in defence of Round's work.

The authors consider two regions A and B which account for 10% and 20% of national employment respectively. They provide two illustrative industries i and j and assume that:

$$\frac{RE_i^A}{NE_i} \equiv \frac{RE_i^B}{NE_i} = 0.08, \frac{RE_j^A}{NE_j} \equiv \frac{RE_j^B}{NE_j} = 0.12$$

where RE denotes regional employment NE national employment.

They note that Round's formula produces an import propensity (j purchasing from i) of 0.3 (=1-0.7) for region A and 0.41 (=1-0.59) for region B. Hence, they conclude that, since a higher import propensity has been generated for the larger region B, Round's formula is counterintuitive. Clearly however, the *same absolute number* of employees are employed in i and j in *both* regions *i.e.*

$$RE_i^A \equiv RE_i^B, RE_j^A \equiv RE_j^B$$

Round's formula therefore reflects the fact that the employees in B are servicing a region that is twice the size of A and so, generally, should face a greater domestic demand for their output. With the usual assumption of constant labour productivities, the relatively higher demand for *i* and *j*'s output in B will have to be met by a relatively higher proportion of imports than in region A. Of course, the authors focus upon the specific relationship between *i* and *j* in the two regions. With the forces of supply and demand between *i* and *j* unaffected by the increase in total regional size, one might ask why the propensity to import commodity *i* by sector *j* should change *at all*? The answer is that the spatial dimensions of region B are presumably greater than in A. Consequently, in region B the *i* suppliers may be located further away from the *j* demanders and so, for example, may be more difficult to find than they are in the smaller region A. The higher import propensity for the larger region is therefore perfectly intuitive. Since Round's formula performs consistently with expectations, it cannot be rejected on these grounds. By reverse logic however, formula (5, p.551), the ELQ, which transfers the logarithmic expression to the numerator of Round's equation, must be counterintuitive, and so too must the subsequent formula for the FLQ. The fact that the authors' reasons for the revision of Round's cross-industry formula are unfounded suggests that the FLQ is surplus to requirements. Nevertheless, a closer examination of the FLQ is worthwhile since it reveals some of the pitfalls of the development of location quotient approaches.

The authors suggests that the CILQ should be scaled down by a variable which reflects regional size. Clearly however, the suggested scalar, lambda, is anything *but* reflective of differences in regional size! For example, the authors suggest a scalar of 73% for a region which represents 10% of national employment (about the size of a UK Standard Region) and a scalar of 70% for a region which accounts for just 2% (about the size of a UK county). Is this 3% margin likely to be an adequate or significant reflection of the difference in regional size? Realising perhaps that their regional scalar is redundant, the authors then suggest that lambda is raised to a power (1 to 5). Of course, whilst this has the potential to reduce the size of the scalar in general, it introduces little more variation between the scalars for regions of different size. So, it would appear that the dimension the authors have introduced to nonsurvey estimation is a choice between the numbers one to five.

Consider the illustration of Table 6, p.552. Two regions with three common sectors are specified. Region A is precisely ten times the size of Region B; the sectors 1,2 and 3 in A are also ten times the size of those in B. The authors claim success in finding that FLQ is the only formula to produce a higher import propensity for the smaller region. But surely, there is no *real* spatial effect to model here: the sectors in B are one tenth the size of A, but the *dimensions* of region B are (presumably) also one tenth those of A. Furthermore, as in the previous example, with constant labour productivities, there is no reason to believe that the supply and demand relationships between the respective sectors are different between A and B. The FLQ can therefore be rejected outright on the grounds that, in this example, *it* is counterintuitive. Incidentally, the justification (p.552)

'that ... firms in Region B would experience greater difficulty than their counterparts in Region A in satisfying any increases in regional demand'

seems misguided in the context of the conventional linear production function. What if we wished to model contractions in regional demand?

Whilst this discussion has been necessary to prevent the unfair rejection of Round's formula, making any form of 'corrective' detail-adjustments to location quotient formulae is a dangerous form of alchemy for it imparts value into formulae which have little theoretical or empirical foundation. If the trade hypothesis underlying the CILQ is misspecified both Round and Flegg's formulae can be rejected outright. But where in Flegg *et al.*'s paper is the theoretical and empirical analysis which establishes this fact prior to the development of the FLQ? Furthermore, where does the FLQ stand on the estimation of local consumption propensities? Can their concern with the improved estimation of regional trade be justified given the likely significance of the error associated with the misspecification of regional technology? Such issues can really only be explored and resolved within a tight theoretical-empirical framework, which Flegg *et al.* certainly do not provide.

Flegg *et al.* should really question what it is they have achieved by the FLQ. Have they made a useful contribution to nonsurvey methodology - a technique that enhances the holistic features of the regional model, and thus facilitates the overall hybrid process? Or does the estimated regional table lack feature and diversity - the *raison d' être* of the regional model? Examining the authors' Avon FLQ multipliers (Table 7 p.554), which hardly vary from the mean of 1.07, would suggest that the latter is true. What Flegg *et al.* have produced is not an economic model of the county of Avon, but a pancake.

The authors may well be happy with this result. But it seems they could have saved themselves much time and effort by deriving a simple Keynesian multiplier. Alternatively, in their Appendix B1, they go perilously close to repeating the work of Bromley (1972), which is a mere stone's throw away from the Burford and Katz genre of short-cut 'input-output' multipliers (*i.e.* Burford and Katz, 1977). Focusing in either direction may at least have the advantage that, in future, they take more time and consideration, and compile more regional-specific data for the estimation of their multiplier parameters (although of course see Hewings and Jensen, 1985 on the Burford and Katz approach).

One final point, which is slightly out of context in a location quotient debate but which underlines the naiveté of Flegg *et al.*'s treatment of input-output issues is their claim to have discovered a new procedure for aggregating regional input-output matrices. Whilst their advice is correct - that trade adjustment should take place prior to aggregation rather than post aggregation in order to account for regional product mix - this has long been established as common procedure (see for example Shen, 1960; Jensen *et al.*, 1979). Indeed.

the prior-to step is clearly set out on page 67 of the Johns and Leat, 1986 study they refer to. The authors are mistaken in believing that (p.553)

'The studies of Smith and Morrison ... and Harrigan et al. ... can be taken as indicative of the normal procedure used in producing a non-survey regional input-output table.'

as clearly, they are not. As previously mentioned, this genre of studies aggregate prior to the application of a nonsurvey technique in order facilitate direct comparison of simulation performance. Furthermore, the authors should be aware that the errors associated with post-aggregation trade estimation have already been established (see for example Sawyer and Miller, *op. cit.*).

Flegg *et al.* provide some valuable lessons. Clearly, as Hewings and Jensen warned in 1986, *'tinkering'* with nonsurvey trade formulae is a desperate trap to fall into. It seems that all it achieves is to move one further away from the reality of the task in hand - that is building a model which reflects the features of the regional economy. If there is anything to gain in terms of data improvement from purely nonsurvey method it probably lies in regional technology specification. However getting caught up in formula-tinkering must be avoided at all costs, because the rewards are unlikely to justify the effort. It is possible that current hybrid practice requires some rethinking in order to ensure that Flegg *et al.*'s mistakes are not repeated.

4.53 Stevens *et al.* Regional Purchase Propensities

Whilst the Stevens-type approach may be flawed in its justification for the focus upon trade estimating techniques, mercifully they avoid the temptation to play with quotient-based formulae. Instead they settle for the row-proportional import propensity (RPP) which is normally associated with Isard and Keunne (1953) and Miller (1957). It would seem that their concern with the generation of an average row propensity is appropriately defined within the modest objectives of the nonsurvey approach.

What is refreshing about the Stevens *et al.* (1983; 1989) approach is that it is based upon a theoretical and empirical analysis of the determinants of the RPP. The 1983 investigation, which contains a number of regional specific regressors is superseded by the 1989 study. Two things are interesting about the 1989 estimating equation: firstly that it focuses almost exclusively upon supply determinants; secondly, of the two continuous explanatory variables, one is regional specific (the regional industry's share of the national industry's output - essentially the SLQ without the normalisation), the other is commodity specific (the nationally recorded weight-value ratio). The inclusion of this latter variable is an important innovation because it recognises that commodities may possess some *inherent* characteristics which are relatively invariant across regions, yet influence input-output relationships. Thus, for example, the propensity to import manufactured products, which tend to be highly specialised, and narrowly defined will, in general, be higher than for service products. The notion of commodity specific determinants and moreover, the attempt to separate them from those that are regional specific, brings the problem of identifying import structure reasonably close to the problem of identifying Fundamental and Non-Fundamental Economic Structure (*i.e.* Jensen *et al.*, 1988; Jensen *et al.*, 1991).

Thus although Stevens *et al.*'s justification for focusing exclusively on trade is at fault, one would have to concede that the traditional methods of trade determination in use prior to the Stevens analysis required attention. Their work takes an appropriate theoretical and empirical line and is valuable in that it provides food for thought on the subject. Of course, their estimating equation is limited to the context of the USA. Can one therefore separately identify fundamental and localised characteristics within survey based UK regional import propensities which could then find application as a nonsurvey regional tool? Moreover, their approach seems limited to the inter-industry matrix. Whilst estimates of gross household expenditure can be derived for standard regions, there is little to determine the local consumption propensity other than the location quotient. Can a RPP equation be estimated for households? What is the relationship between household and industry estimating equations?

4.54 Popular Hybrid Philosophy: An Appropriate Paradigm?

If the examples of direct comparison analysis have achieved anything at all it has been to demonstrate that one-step, quick-fix methods of estimation do not generate regional input-output information with any useful degree of accuracy. However, deficiencies in the execution of these studies, in particular their failure to offer appropriate guidance on best practice, have at times risked the credibility of the regional input-output model (see Jensen and MacDonald, 1982). In this respect, the hybrid approach to construction has been the saviour of regional input-output because it has served to 'bridge the gap' between prohibitive cost and prohibitive inaccuracy and moreover, re-emphasised the importance of a survey model-content. Hybrid methods are now the accepted approach to regional model construction (Lahr, 1993), with the GRIT procedures finding greatest popularity.

However, it seems ironic that the GRIT procedures are founded on what are largely the nonsurvey methods first formalised by Schaffer and Chu in 1969. There are at least three serious problems with this fact.

Firstly, by recruiting standard nonsurvey techniques GRIT implies that they have something to offer in terms of improving model accuracy. However, there is no particular evidence to suggest that they perform significantly better than a stochastic 'regionalisation' process, and indeed one would suspect that this is a fairly close run thing. So why promote their continued use? Does this not offer encouragement to alchemists such as Flegg *et al.*, 1995?

More seriously, whilst survey and superior data collection should be the quintessence of the input-output approach, the GRIT procedures would seem to portray this process more as an error correction mechanism. The idea that one can 'pick' at a few choice coefficients promotes this idea. There is of course the logical flaw associated with coefficient selection from nonsurvey foundations.

Thirdly, by placing the nonsurvey step ahead of the survey phase GRIT attracts what one might term 'model-sharks'. GRIT permits the generation of a fully operational 'regional' model, but then of course cannot possibly prescribe or dictate the level of 'superior' information that invokes the transformation to hybrid status. Anybody can say that their nonsurvey model 'follows the GRIT procedures' and gain the respectability of having followed 'good practice'. GRIT *can't* ensure good practice. But then who can?

'Good practice' can of course only be encouraged, but *demoting first-best techniques to the back seat does not do this.* And some alarmingly bad impact assessments result from this 'loop-hole' in procedure (see Business Strategies Limited, 1997 - a study which, even more alarmingly, uses the FLQ). Studies of this nature undermine the credibility of regional analysis in general.

In short survey-based data should be the foundation of the regional input-output model, yet in the GRIT model it is the largely unqualified nonsurvey methods that hold this place.

There is however no particular reason why the two approaches could not be reversed. Indeed, by generating some survey data first, it would seem possible to base the proceeding nonsurvey phase upon this information in some way. This would restore the logical order of preference to model construction that is reversed in GRIT and associated hybrid techniques, *i.e.*

Table 4.3 Preference Order for Hybrid Approaches

	Order of	Order of Approach	
Method	**Preference**	**GRIT**	**Proposed**
Survey	First-best	2	1
Survey-based Nonsurvey	Second-best	No general principles	2
Pure Nonsurvey	Third-best	1	3

Thus nonsurvey methods, and in particular classic location quotient approaches, could feasibly be banished from the hybrid process altogether and nonsurvey adjustments would become survey-based. Most importantly of all, survey or superior data collection would become *necessary* in order to gain possession of an operational, respectable regional input-output model. Indeed, this sort of approach is applied in Holland, where regional input-output modelling practice is significantly more advanced (see for example Boomsma and Oosterhaven, 1992, although this is based upon the relatively detailed Dutch regional data set).

Possibly, there has been some over-dramatisation of the costs of survey generation, and this may well extend from the fact that the debate originated in the USA, where 'regions' are equivalent in size to European nations. But, in the UK, it is probably not the case that a postal purchasing enquiry covering a selection of 'local interest' sectors should fail to find financial backing because of its prohibitive expense. This would at least mark the beginning of a coherent regional database.

Therefore the role of the input-output table as a *regional economic database*, and not simply a 'black box' for generating local multipliers, should be promoted by pushing survey-work to the forefront of the hybrid exercise. The objective should be to encourage an approach which is focused not upon the methods of data estimation, but on the process of data collation, ultimately to the level afforded by the Social Accounts Matrix.

However, clearly the basic nature of the input-output estimation problem needs to be properly understood before a set of principles governing the alternative paradigm can be established.

Thus the first part of the research will set out to establish *a priori* the nature of the general estimation problem, and attempt to derive some broad principles. The second part will attempt to provide an empirical explanation of estimation error and determine a more specific set of guidelines for a survey-nonsurvey hybrid approach.

4.6 Summary and Conclusion

This chapter has reviewed the evidence concerning nonsurvey and hybrid methods of generating regional input-output data. Evidence from direct comparison studies has shown that the partial-survey algorithm, RAS, is superior to the classic set of purely nonsurvey techniques. There is also some empirical evidence to suggest that regional technology is a significant source of misspecification. Whilst there would seem little to discredit the direct comparison as an approach to analysis, studies have failed in their application. In particular, the failure to undertake an *a priori* analysis of the range of relevant estimation issues has adversely affected the development of nonsurvey methods.

The view that the specification of the household vector is relatively important was accepted. However, the Stevens-hypothesis that there is a difference in the transmission of technical and trade coefficient error was rejected. Consequently the relative importance of trade and technology misspecification must be established by observation, not simulation. Resolving this issue will have implications for the deployment of survey resources and the future development of nonsurvey methods.

It would seem highly likely that technology misspecification is a significant factor in explaining the failure of classic nonsurvey methods. Therefore, since developments to these techniques have largely failed to embrace the issue of regional technology, seems unlikely that they have significantly improved their performance. Moreover, since most of these techniques make relatively minor adjustments to classic import-specification formulae, their success on *this* aspect of estimation seems doubtful. The techniques that have been developed are also vague on the issue of specifying local consumption propensities for the household vector.

It would seem that the move from judgement-orientated approaches to more formal procedures has resulted in a net loss to nonsurvey methodology. The blame for these failures can, at least in part, be placed with the direct comparison studies.

The only trade development to have made a definite positive contribution to the field of nonsurvey trade estimation would appear to be that of Stevens *et al.*, and this derives from the fact that it has a reasonable theoretical and empirical basis. In particular, accounting for commodity-specific characteristics seems a worthy innovation.

Finally the nonsurvey-survey ordering of approach that characterises popular hybrid methodology such as GRIT does not encourage best practice. A survey-nonsurvey scheme of order would seem to offer significant benefits. The development of principles to this effect will attempt to meet the research objective of improving hybrid procedures.

The method of approach will be, firstly, to give consideration to the nature of the estimation problems in hand and thus form research hypotheses. These hypotheses will then be tested by empirical means, and this evidence will provide the basis for conclusions and recommendations. Hence the next chapter formulates the research hypotheses.

CHAPTER 5

RESEARCH OBJECTIVES AND THE FORMATION OF RESEARCH HYPOTHESES

5.0 Research Objectives and the Formation of Research Hypotheses

5.1 Introduction

This chapter draws together and develops upon the evidence of the previous chapters. A formal statement of research objectives and strategies is made. This is followed by an *a priori* consideration of the nature of nonsurvey error which leads to the formation of research hypotheses.

5.2 Research Objectives and General Strategies

i. The general aim of the project is to promote the input-output model as a regional economic database. Part of this encouragement will be to develop an improved method by which the nonsurvey data for use in regional input-output tables are generated. This is the principal objective of the research.

ii. The principal objective will be achieved by developing a methodology which reverses the paradox in popular hybrid philosophy, namely the nonsurvey-to-survey ordering of approach. The new procedure will attempt to offer some 'holistic guidance' on the deployment of survey resources by resolving the trade versus technology debate.

iii. The approach requires a set of governing principles which will be developed throughout the process of analysis.

iv. Firstly the estimation problem will be considered from the third-best, or nonsurvey perspective. The analysis will seek to establish the general nature, significance, and relative importance of the principal components of nonsurvey estimation error: trade and technology misspecification.

v. On the basis of this analysis, the research will derive an appropriate estimating technique from empirical sources and test its relative performance.

vi. The research will discuss how the parameters for this methodology could be established through 'second-best' methods (*i.e.* survey-based-nonsurvey).

vii. These experiments will mark the completion of the research exercise and thus conclusions on the relative success of the research programme will be drawn.

viii. On the basis of these conclusions, recommendations for the creation of an input-output database in Devon and Cornwall will be given.

This research strategy follows the lines of the general direct comparison model set out in the last chapter. Stage *iv* will initially form an *a priori* analysis, from which hypotheses will be generated. The hypotheses relating to *iv* will be tested empirically using direct comparison methods. Stage *v* will attempt to explain the observations of *iv*. Stages *vi*, *vii* form the basis for recommendation in *viii*.

5.3 *A Priori* Consideration Leading to the Formation of Research Hypotheses

5.31 The Nature of the Nonsurvey Estimation Problem

Nonsurvey estimation error is conventionally expressed in terms of regional technology and trade misspecification (see for example Smith and Morrison, 1974). The former concerns the extent to which the assumed national technical coefficient departs from the observed regional technical coefficient; the latter concerns the extent to which the application of a proxy measure of trade orientation such as the location quotient causes a departure from the observed regional purchasing coefficient. For simplicity, this convention has been applied in previous chapters.

However, following the logical flaws identified in nonsurvey methods by Hewings and Jensen (1986), a more formal definition of the sources of estimation error would differentiate between total use (*i.e.* purchases irrespective of source) and overseas trade. One should also recognise that it is virtually impossible to observe a 'true' input-output table - sampling error is inevitable. Thus, one can define four categories within nonsurvey estimation error:

(*i*) The error associated with the misspecification of regional total use;
(*ii*) The error associated with the misspecification of overseas trade;
(*iii*) The error associated with the misspecification of regional trade;
(*iv*) Stochastic error.

Error (*i*) arises through differences in the 'true' regional and national total use coefficients. Error (*ii*) arises through differences in regional-national overseas import propensities. Error (*iii*) arises through mechanical trade estimation. Whilst it should be feasible to separate out errors (*i*), (*ii*) and (*iii*) (following

Smith and Morrison), stochastic errors will be present in each component and, since they are largely unobservable, their effects cannot be separately measured.

Thus, whilst some authors (*i.e.* Flegg *et al.*, 1995) have been concerned purely with the estimation of the regional purchasing coefficients from national technical coefficients, here the nonsurvey estimation problem is more explicitly defined. There are three regional components to estimate: total use, domestic use (total less overseas import), and regional use (domestic use less regional import).

5.32 The Nature and Relative Importance of Nonsurvey Error Components

The first task, as set out by (*iv*) above, is to establish the nature and relative importance of these error components. Only by identifying the 'significant' contributors to estimation error can one begin to tackle the estimation problem. So what are the expectations on the nature and relative importance of the error components? In what follows, for simplicity, the discussion is centred upon the production functions of industries, although the analysis could be equally applied within the context of consumer purchases.

5.321 The Assumption of Spatially Invariant Production Functions: The Problem of Heterogeneity

All classic nonsurvey methods assume that national total use coefficients - the production functions in full - hold at every conceivable regional dimension. Various suggestions have been made as to why this is unlikely to be observed in practice. Smith and Morrison (1974 p.22) identify anything from differences in productive efficiency to climatic variation as contributing to the phenomenon. However, it is argued here that, apart from variation due to stochastic observation error, differences in total use can be attributed to the violation of one fundamental assumption of the Leontief model, and that is that each defined sector is homogeneous.

Homogeneity implies that the set of coefficients which describes each defined commodity's production is *wholly unique*. Hence there can be *no* variation in the means of production for any defined commodity, and if there is, for whatever reason, the commodity would require separate definition in order for the homogeneity assumption to hold. The reason for the assumption is quite straightforward. If each production function describes a diverse set of commodities, the pattern of linkage is hidden within average relationships (see Table 5.1 below), and hence the precision with which one can calculate, for example, multipliers is eroded.

It is argued here that differences in regional and national total use derive from the fact that the production functions of the national model are heterogeneously defined. The argument is as follows. At the broadest levels of commodity definition *i.e.* agriculture, manufacture, regional production functions are merged together, and as such cannot be identified in the national model. As the definitions of the national model are increased, the production functions of regionally specialised commodities emerge *i.e.* dairy farming, sea fishing, fish farming. At some much higher level of disaggregation - where homogeneity is approached - the definitions of the national model are so fine that it is possible to identify the individual factories and firms operating within the nation. At this point, the national model becomes one of infinite-regions because it is possible to extract the input-output table for any conceivable spatial subset of the nation. Not only the problem of estimating regional production functions evaporated, the trade estimation problem has ceased to exist.

A highly simplified illustration of this effect is presented in Table 5.1 below. Suppose the national input-output table defines a sector *N* as making purchases of commodities *b* and *c*. The respective national coefficients are 0.25 and 0.75, as shown. Suppose though that the national sector is heterogeneously defined, and there are three regional variations in the production process, *A*, *B*, and *C*.

Table 5.1 Heterogeneity in Production Functions

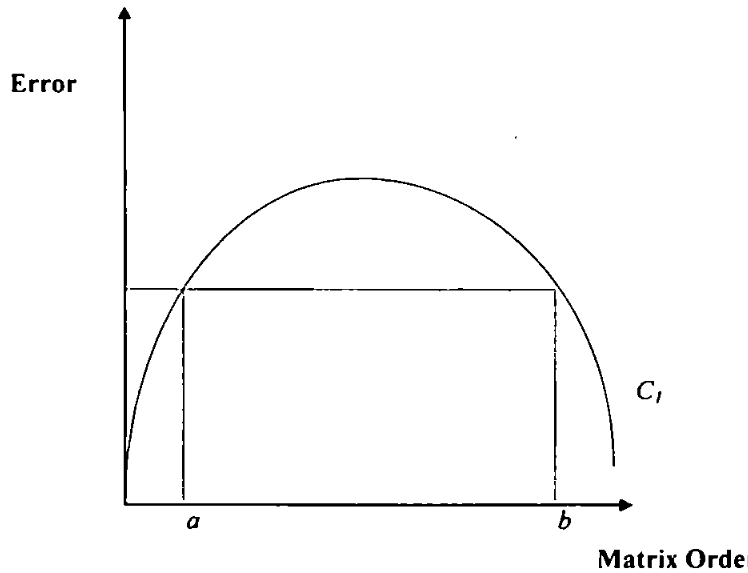
	Transactions				Coefficients			
	N	*A*	*B*	*C*	*N*	*A*	*B*	*C*
b	250	200	50	0	0.25	0.5	0.1	0
c	750	200	450	100	0.75	0.5	0.9	1.0
Total	**1000**	**400**	**500**	**100**	**1.0**	**1.0**	**1.0**	**1.0**

When defined as sector *N*, the nonsurvey modeller in region *A*, *B*, or *C* faces a problem in estimating the regional coefficients. When the national table is defined with sectors *A*, *B* and *C* of course, the regional coefficients can be determined from the national model. Note that the regional coefficients are *distributed* around their weighted mean (weighted by gross regional output), which is the national coefficient.

For a model of a given region, defined consistently with its national counterpart, one would be able to measure the degree of error due to heterogeneity by comparing the regional-national total use coefficients. One could go further and, by measuring the difference between total use coefficients at different levels of aggregation, one could gather a range of observed error values to form what one might call the 'heterogeneity error function' - assuming of course that a measure of error can be applied which is independent of matrix order. Thus, with evidence of some established general functional relationship between total use error and matrix order, one could make some assessment as to whether the national model adequately describes the region's production function. But what would such a function look like?

Expanding upon the analysis of above, at very low orders of disaggregation, say where production is split simply between manufacturing and non-manufacturing, it is unlikely that *any* region will display wild dissimilarities in such fundamental classifications of production. Thus, at low orders of disaggregation, one would expect little difference in the error associated with total use. As order rises, specialisations emerge in all regional production functions, but these are averaged out in the national model. Thus the observed differences in total use will increase. At some point of disaggregation however, the national model must begin to describe these regional specialisations as separate definitions. Thus the associated error will fall until the national model fully describes regional production. Hence the hypothesised form of the 'heterogeneity error function' with respect to matrix order is parabolic, as in Figure 5.1 below (this is of course highly stylised).

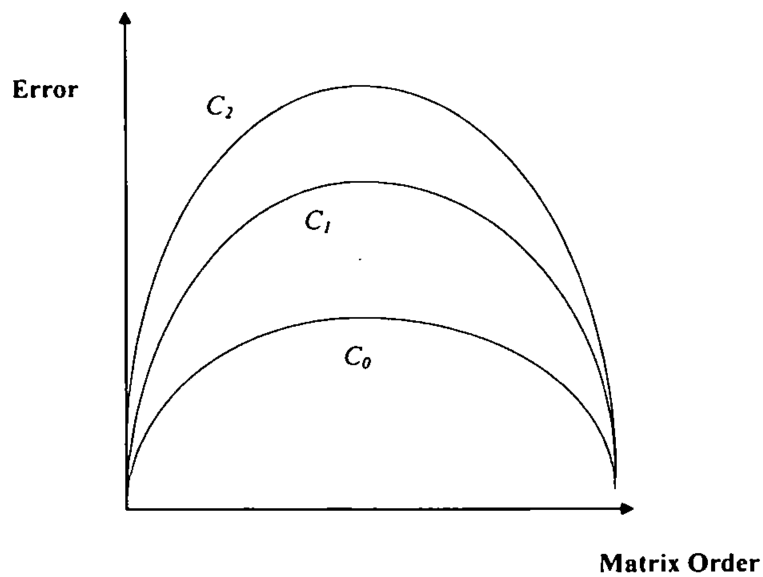
Figure 5.1 The Heterogeneity Error Function



Of course the national model at point *b* this is preferred to that at point *a* because, although the measure of estimation error may be equal, the production functions of model *b* are more homogeneous.

However, the specification error will also be a function of regional size. Consider a single heterogeneous production function. Thus when the region *is* the nation, the specification error must be zero. As one moves down the scale of size however, and one begins to encounter specialisation in production, the specification error will tend to rise. Figure 5.2 below illustrates the hypothesised relationship between total use misspecification with respect to matrix order for regions $C_0 > C_1 > C_2$.

Figure 5.2 Heterogeneity and Regional Size



Hence the two testable hypotheses are as follows:

H_1: the error associated with total use misspecification is parabolic with respect to matrix order, as per Figure 5.1.

H_2: the error associated with total use misspecification is negatively related to regional size.

5.322 The General Nature of Trade Misspecification

The analysis of the behaviour of trade estimators is similar to that of total use. Firstly, with respect to overseas import propensities, at low orders of observation, regions probably have a similar average dependence on international imports, and the error associated with the use of national propensities will be low. As order rises, error should first rise and then fall as the national input-output sectors ultimately become homogeneous. These effects should be accentuated for smaller, more specialised, regions. Thus:

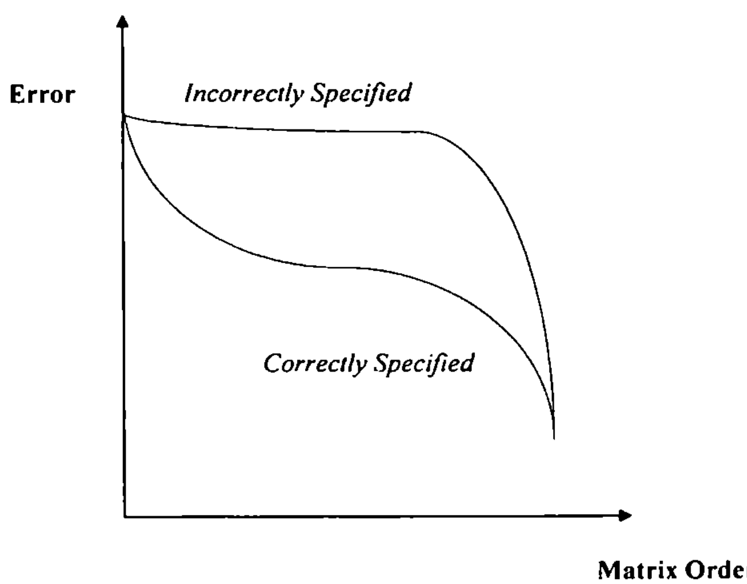
H_3: the error associated with the use of national overseas import propensities is parabolic with respect to matrix order.

Location quotient formulae, such as the SLQ and CILQ are the most common method of generating regional import propensities by nonsurvey means. Hence the general nature of regional trade misspecification is considered from a location quotient perspective.

Consider the case where the transactions table is defined by a single heterogeneous commodity. With only one product, there is no commodity specialisation, and thus location quotients are defined at unity: all regions are classed as self-sufficient. The failure of the location quotient in identifying regional trade is therefore complete, and hence the error associated with regional trade misspecification will be relatively high. Increases in the number of defined sectors however represent an informational gain to the location quotient. Therefore, if they are indeed correctly specified, one would expect the error associated with trade misspecification to fall with increases in matrix order. As regional-specific production functions emerge within the national table, and non-competitive imports are increasingly identified, one should observe some acceleration in the rate at which the error falls.

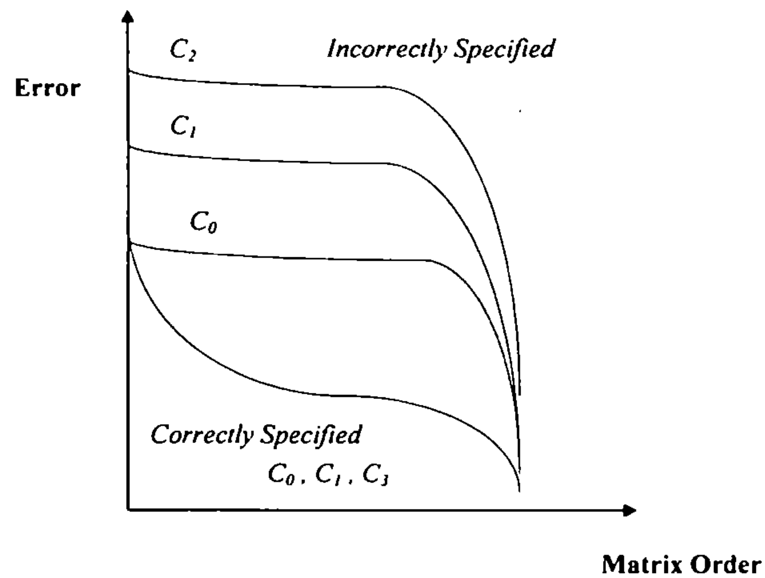
However, if location quotients are largely unrelated to regional trade orientation, the associated regional trade misspecification error should remain fairly constant across matrix order, up until homogeneity is approached (where the need to mechanically estimate regional imports is obviated). Figure 5.3 below illustrates.

Figure 5.3 Regional Trade Error Functions - Single Region



With respect to the behaviour of the error function across regional size, since regions generally become more open to trade as they become smaller, correct trade specification will become more important as regional size diminishes. If quotients are operating correctly, they should account for this, and the level of error should not vary across regional size. However, if quotients are misspecified in their function, error will be negatively related to regional size. Figure 5.4 illustrates for three regions $C_0 > C_1 > C_2$.

Figure 5.4 Regional Trade Error Functions - Variable Regional Size



Given the inherent deficiencies of quotient formulae, hypotheses **H_4** and **H_5** are that the regional trade error functions exhibit signs of misspecification.

H_4: the error associated with regional trade misspecification is largely unrelated to matrix order.

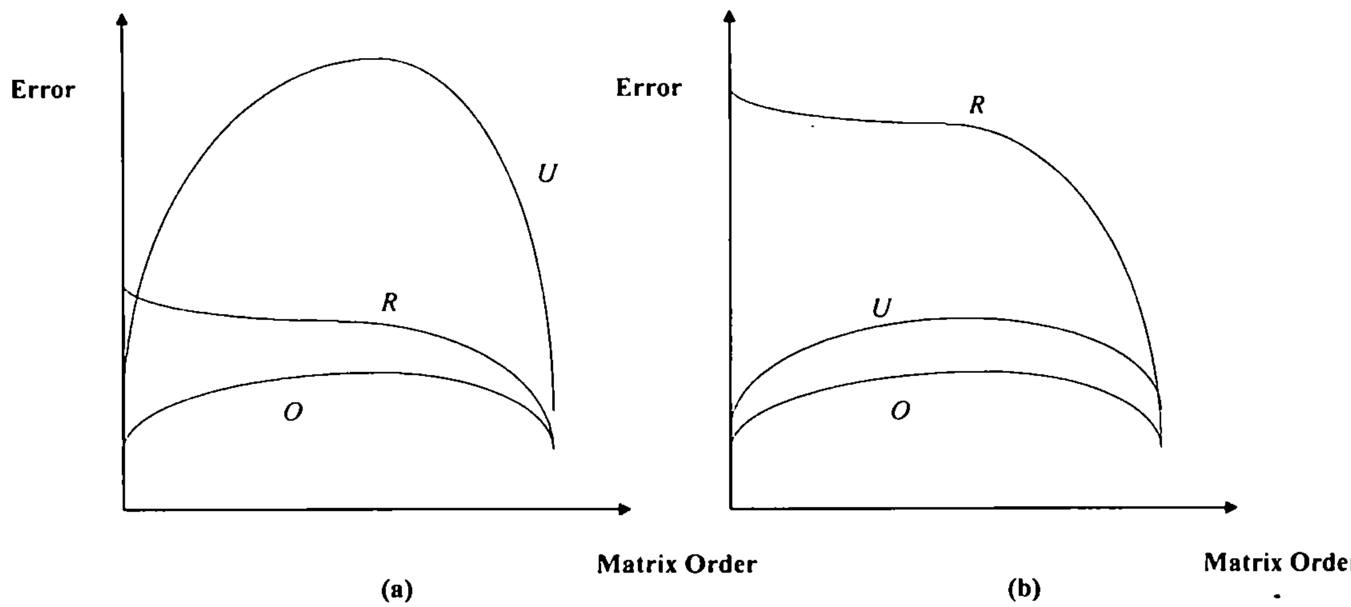
H_5: the error associated with regional trade misspecification is negatively related to regional size.

5.323 The Relative Importance of Trade and Total Use Misspecification

The identification of the most important components of error is intended to provide a focus for the process of technique development. Consider Figure 5.5 below. The functions *U*, *R* and *O* represent the misspecification error associated with total use, regional, and overseas trade respectively; it is assumed that the measures of error are directly comparable. One would suspect that the error associated with the misspecification of overseas imports is the least significant of the three components. The propensities, as given in the national model, will partially be determined through factors which are not specific to location within the nation, for example, relative UK-overseas prices - and as such they may have some general relevance at the regional level. Furthermore, if the national model's overseas import propensity matrix is reasonably based upon survey data, it will represent a pattern of behaviour which is inevitably more realistic and relevant than, for example, that represented by simple quotient regional trade estimators.

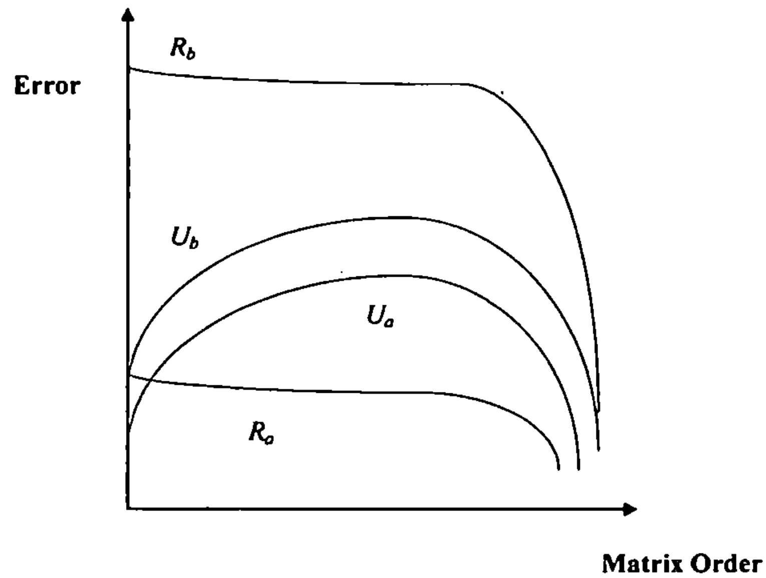
Panel (a) below shows the case where total use is the greatest source of misspecification and thus warrants most attention in any strategy for the development of estimation methods. In panel (b) however, the misspecification of regional trade dominates, and this is the implicit assumption behind Flegg *et al.* who concentrate exclusively upon developing trade estimation methods.

Figure 5.5 The Relative Importance of Error Components



One would suspect that the picture is not one of dominance, but rather that both regional trade and total use functions are significantly misspecified, and thus in need of 'attention'. However, one can perhaps make some prediction with respect to relative importance and regional size. If hypotheses $\mathbf{H_2}$ and $\mathbf{H_5}$ are correct, then as the region becomes smaller, the level of error associated with all components will rise. If small regions are increasingly reliant upon regional imports, the fact that quotients appear ill equipped to capture such an effect would seem to suggest that the specification of the trade function will take on greater importance as one moves down the scale of regional size. Hence the hypothetical relationship between error functions and regional size is shown in Figure 5.6 below, where region $a > b$.

Figure 5.6 Relative Importance and Variable Regional Size



The hypotheses underlying the illustrations of 5.5 and 5.6 are

H_6: nonsurvey assumptions over total use relationships and regional import propensities are misspecified to a significant degree.

H_7: the assumption of national overseas import propensities holds at the regional level and therefore constitutes the least significant source of misspecification.

H_8: regional trade specification is relatively more important than total use specification for smaller regions.

5.33 Towards an Appropriate General Nonsurvey Method of Application: the Intermediate Matrix

At this stage of the research, the general nature of nonsurvey error and the relative importance of its constituents will have been established. The objective is now to develop a method which goes some way to tackling the identified sources of error. The method will be referred to as 'nonsurvey' or 'third-best' throughout. The specification of its parameters from a survey-base will transform it into a second-best approach.

In what follows, it is assumed that $\mathbf{H}_6$ and $\mathbf{H}_7$ are maintained in the direct comparison analysis. In other words, the nonsurvey overseas import propensity matrix has been found to be acceptable for direct use: the discussion therefore centres upon adjusting the total use relationships and estimating regional trade.

Whilst the observation may have been made, for example, that the regional production functions are misspecified by the national model, it may be that this level of error cannot be significantly reduced without the specification of elements which are beyond the necessarily broad limits of the nonsurvey approach. The next stage of the analysis is therefore to make an assessment of the potential for reducing nonsurvey estimation error. This potential can be assessed by specifying the maximum and minimum expected limits of error associated with nonsurvey estimation. These boundaries can then be used as an apparatus for gauging subsequent suggested improvements in estimation methodology.

The maximum limits for error are relatively obvious. For example, in the adjustment of total use coefficients, a suitable benchmark for the maximum misspecification would be the error between the national total use coefficients and the contiguous regional observations.

In specifying a minimum level of error, the choices may be more contentious. One feasible option is to assume that the minimum achievable error is zero. Given the nature of input-output data and nonsurvey objectives, this is inappropriate. The second option is to assess the error associated with the application of a regionalisation technique which is relatively superior in its assumptions and operating data, and yet whose components would be a realistic target for replication by survey-based-nonsurvey means. The staged RAS algorithm seems a sensible choice for this role. RAS is suitably broad - operating at the level of row/column sum, rather than cell-by-cell, and yet manages to accommodate the three components of specification.

Assuming that the RAS data set enables a significant reduction in nonsurvey error (and this is supported by the evidence of previous studies) the generation of the RAS data set becomes the objective of the 'new' estimating technique. So what nonsurvey variables might determine the RAS constraints? Firstly, reconsider the requirements of the staged RAS procedure.

Generally, the RAS algorithm requires the specification of two vectors: 'initial' and 'desired' row and column sums of the intermediate transactions table. Hence, in the first stage of adjustment - the regionalisation of the total use matrix - the 'initial' intermediate row and columns sums can be derived from:

$$ {}^{t}\mathbf{U}^{*} = {}^{c}\mathbf{C}^{*}\hat{\mathbf{X}} \qquad [5.1] $$

where

$^{r}\mathbf{U}^{*}\ [u^{*}_{ij}]$ is the nonsurvey regional combined use matrix in transactions;

$^{c}\mathbf{C}^{*}$ is the national combined use matrix in coefficient form;

$\mathbf{X}$ is the vector of regional gross industry outputs.

Most of the 'actual' totals (*d*) will not be available from published sources. Intermediate column sums for manufacturing sectors may be available from the Census of Production, but others, and in particular the actual intermediate row sums, would have to be estimated. Note that one feasible objective would be to estimate the *ratio*, *t*, of actual to initial sums instead of the transaction value itself, which could of course be inferred once the ratio was known *i.e.*

$$t = \frac{d}{u^{*}} \qquad [5.2]$$

where

d, u^{*} are, respectively, actual and initial (nonsurvey) transactions values of intermediate column/row sums of total use.

Under $\mathbf{H}_7$, domestic consumption propensities are assumed to be 'known' and are denoted by *k*, such that the product *dk* is equal to the 'actual' intermediate sum of the domestic use matrix. Concentrating on the adjustment for regional trade then, the ratio that maps the intermediate sums of the actual domestic use matrix (*dk*) onto those of the actual regional use matrix (*r*) is the local consumption propensity *i.e.*

$$l = \frac{r}{dk} \qquad , 0 \leq l \leq 1 \qquad [5.3]$$

where

r is the actual intermediate column or row regional use in transactions;

Therefore, one is looking to estimate two principal functions: (*i*) the relationship between nonsurvey-survey total use transactions; and (*ii*) local consumption propensities. The sense of these functions is of course dependent upon whether one is generating row or column constraints. Assuming a commodity-by-industry framework, in generating rows constraints, the questions being asked are (*i*) how much more or less of commodity x is used by regional industry than is implied by the national average? and (*ii*) what proportion of commodity x purchased by regional industries from national sources comes from local supply? In generating column constraints, these questions become (*i*) how much more or less of its total input purchase does regional industry x spend on commodities than is implied by the national average? and (*ii*) what proportion of regional industry x's purchase of nationally produced commodities are local? So, given that these functions will have to be estimated, what regional variables might help to form such an explanation?

In many ways this is a moot debate. Regional variables which would have obvious relevance, such as relative delivered costs, receive only partial coverage within the UK data set. In fact, there are only a couple of regional variables which have any potential to be used as estimators of the given functions: industry gross purchases, wages, employment, gross value added, and gross output. These data are derived from the Census of Production and hence the coverage is limited to manufacturers. Furthermore, the data are not cost-free and elements may be suppressed to maintain confidentiality (as one moves down the scale of regional size this of course becomes more of a problem). For service sectors, income estimates of GDP are made for broad divisions (in the ONS's *Regional Accounts*). Thus the only variable where coverage could be considered more or less 'complete' in terms of regions and sectors is employees in employment from the biennial Census of Employment[1].

Hence it is really only the well known location quotient - calculated using output and GDP where possible, employment otherwise (West, 1980) - which is at the regional analyst's general disposal. This constraint is reflected in the historical development of nonsurvey techniques. Even the Stevens *et al.* (1983) analysis, which marks an attempt to extend trade estimation beyond relative specialisation, returns at a later date to what is principally a location quotient methodology (Stevens *et al.*, 1989). So just how far can the location quotient be put to use in the estimation of a RAS-type data set? Consider this firstly from the perspective of generating intermediate row sums.

[1] Self-employment is only available from the Census of Population which is made once every ten years.

5.331 Generating Intermediate Row Sums

5.3311 Adjusting for Total Use

Under $\mathbf{H_5}$, the sectors of the national input-output model are heterogeneously defined. This means that the national total use coefficients may disguise regional specialisations in production processes. It is possible that an apparent 'over specialisation' (*i.e.* $q_i > 1$) in the regional production of a commodity is not so much a mark of net exportation, but of greater-than-average *use* of that input by the region's industries. There are two main ways that this can come about, and these extend from classic location-production analysis (*i.e.* Weber, 1929; Isard, 1956; Moses, 1958). A similar analysis is made by McCann and Dewhurst (1996). Firstly, with a regional specialisation in the production of a particular commodity, it may be that the region's indigenous industries substitute that commodity into their production processes in order to take advantage of, for example, lower transport costs. Secondly, the relative abundance of the locally specialised commodity may, over time, have attracted industries who have a strong dependence on its use to locate close to the source of supply, again, to benefit from advantages such as lower costs of transportation. The result of both these processes is that region's industries are greater-than-average users of the region's specialised commodity. The perspective of these arguments can be altered when considering the region's under-specialised commodities (*i.e.* $q_i < 1$). So, for example, the use of commodities which are not available locally may be substituted in favour of those in greater abundance. In other words, is possible that the simple location quotient picks up the ratio of total use, t.

This implies a fairly radical use of the location quotient. On a practical level it means that the location quotients may be applied 'straight', that is, unconstrained, unlike their conventional use in the role of import/export-base estimation.

5.3312 Adjusting for Trade

However, as one extends the analysis to consider (regional) trade, it becomes clear that the location quotient has some applicability, and therefore its role is not as a separate indicator of either total demand or local consumption, but more as a *composite* of these two functions, *i.e.*

$$q_i = t_i l_i = \frac{r_i}{u_i^{\bullet} k_i} \qquad [5.4]$$

Hypothetically, the quotient describes the local use of a commodity from regional supply in relation to the average region's use of the commodity from national supply. There are four possible cases to consider. Firstly, suppose that the commodity in question is regionally under specialised, $q_i<1$, yet there exists a strong regional demand for the product, $t_i >1$ (typically an essential raw material such as oil or metal which is not available locally). The implication is that the ratio l_i will be low. Hence, equation [5.4] feasibly holds. In the two cases where the location quotient simultaneously identifies the demand and supply orientation of the commodity: $t_i <1$, $q_i <1$ (the commodity is not demand and not supplied locally) and $t_i >1$, $q_i >1$ (the commodity is demanded and supplied locally), as long as $t_i \geq q_i$ the hypothesis

$q_i = t_i l_i$ is not ruled out since $l_i \leq .1$. When $t_i < q_i$ however the hypothesis breaks down, and the most 'severe' case of this is when $1 > t_i \rightarrow 0$, $q_i > 1$. This situation describes a commodity which is supplied by the region, but is not demand, and hence is produced mainly for export. The ability to successfully use the location quotient in the implied manner may well depend on one's ability to identify commodities of this nature: judgement and expert opinion should be allowed a relatively 'free role' in any technique that is developed.

It should be clear that, if the location quotient does indeed proxy a combination of relative total regional demand and local consumption, its use implies a single step process from which it is impossible to determine the specific nature of the adjustment in question. For example, given a row adjustment of q=2, one cannot tell if this implies double the average demand and self-sufficiency in supply (t=2, l=1), or some combination involving an import requirement (t>2 and l fractional).

Under $\mathbf{H_5}$, the location quotient fails to account for regional size in estimating import propensities. Assuming that this is the case, the misspecification could be dealt with by incorporating a scale parameter. The basic equation for estimating intermediate row sums can therefore be expressed as:

$$\frac{r_i}{u_i^{\bullet} k_i} = \alpha q_i^{\beta_0} \qquad [5.5a]$$

where

α is a parameter which reflects regional size;

β_0 is a parameter to be estimated.

These variables could be derived from an empirical data set, and thus the model parameters could be estimated. The value of α should be positive and fractional and reflect an 'average' regional import propensity. The parameter β_0 represents the elasticity with which local demand ($r_i / u^{\bullet}_i k_i$) responds to a relative change in local supply q_i, since

$$\eta = \frac{\mathrm{d}(r_i/u_i k_i)}{\mathrm{d}q_i} \cdot \frac{q_i}{(r_i/u_i k_i)}$$
$$= \alpha\beta_0 q_i^{\beta_0 - 1} \cdot \frac{q_i}{\alpha q_i^{\beta_0}} \qquad [5.6]$$
$$= \beta_0$$

The value of β_0 should be positive since, when q_i rises (*i.e.* commodity *i* becomes more specialised), the purchase of *i* from local production should rise as firms substitute away from using other inputs, and firms substitute away from imported supply in favour of local production. The Stevens *et al.* (1989) analysis found an elasticity of 0.51 for US regions. It should be stressed however that the Stevens study estimates only regional consumption propensities - the dependent variable here captures relative *total* use as well as local consumption propensity. Thus:

H₉ : $$\frac{r_i}{u^{\bullet}_i k_i} = \alpha q_i^{\beta_0} \quad , 0 < \alpha < 1, \ \beta_0 > 0 \qquad [5.5b]$$

The final improvement to [5.5b] is to recognise that its structural parameters may not be constant across the range of *i* commodities. The most obvious case of this is with the value of the local consumption propensity α between service and non-service sectors. Intuitively, the propensity to purchase services locally should be higher than for non-service products. If services are, as one would expect, sourced relatively more from local supply, then the rate at which increases in supply are 'absorbed' by local demand may be greater than for non-services. In other words, the value of β_0 for services may be higher than for non-service sectors. Thus, hypothesis 10 states:

H_{10}: The relationship between dependent and independent variables is significantly different between service and non-service sectors.

and with respect to [5.5b]

$\mathbf{H}_{10}: \alpha_s, \beta_s > \alpha_{ns}, \beta_{ns}$ [5.6]

where

s, ns denote service and non-service sectors respectively.

It is of course feasible that other, less obvious, structurally independent commodity groups may be able to be identified.

5.332 Intermediate Column Constraints

5.3321 Adjusting for Total Use

For manufacturing industries, total purchases should be available from the Census of Production. For service and other industries, whilst an income estimate of GDP may be available from the *Regional Accounts*, without gross industry output, total industry purchases cannot be determined (total purchases are, approximately, gross output less incomes GDP). The remaining source of information, the location quotient, has little relevance here and therefore, it is difficult to see how the traditional nonsurvey estimates of total use (*i.e.* from the UK combined use matrix) can be improved upon for these sectors. In what follows therefore, it is assumed the desired intermediate column totals, d_j are available: these are either known from the Census of Production, or are derived from UK values.

5.3322 Adjusting for Trade

With u_j and k_j specified, the task for the generation of intermediate row sums r_j is to specify values for the local consumption propensities, l_j.

One question raised in the last chapter was: does the relative size of the purchasing industry determine its propensity to purchase locally? The Cross Industry Location Quotient suggests that, as relative specialisation rises, the industry's import propensity will also rise: in a sense, the industry 'outgrows' regional supply. However, if a region's industrial structure has formed through firms having taken advantage of locational economies, the opposite relationship should be observed.

The functional relationship between local consumption and relative specialisation is based upon the Stevens *et al.* (1989) analysis. The Stevens equation of course considers local consumption propensities for commodities (*i.e.* rows). Here, one is looking to establish a relationship between the industry's (column) import propensity and its degree of regional specialisation. The hypothesised model is therefore:

$$l_j = e^{-1/Z_j} \qquad [5.7a]$$

where

$$Z_j = \alpha q_j^{\beta_1} \qquad [5.7b]$$

The estimated values of l_j must fall within the range 0-1 and the [5.7a] ensures that this is the case. The value of α is again believed to reflect regional size, although the value of the local consumption propensity will not be α but $e^{-1/\alpha}$. The parameter β_l reflects the extent to which local supply alters to meet with changes in relative industrial specialisation (termed the supply elasticity). If suppliers seek to meet the needs arising from increases in relative regional specialisation (*i.e.* the locational hypothesis), then β_l will be positive. If however the opposite is true and local suppliers do not respond to changes in demand caused by changes in regional industrial structure, then β_l will be zero or negative. A negative value for β_l would support the cross-industry trade specification. The locational hypothesis seems the most plausible. Therefore hypothesis 11 may be stated:

H$_{11}$: $l_j = e^{-1/Z_j}$ [5.7a]

where

$Z_j = \alpha q_j^{\beta_l}$, $0 < e^{-1/\alpha} < 1$, $\beta_l > 0$ [5.7b]

5.34 Towards an Appropriate General Nonsurvey Method of Application: the Vector of Household Purchases

The bulk of the analysis will concern the estimation of the interindustry transactions. Given the undisputed analytical significance of the household sector, this bias appears inappropriate. However, there are a number of possible justifications. Firstly, it is a reflection of the bias within published input-output tables: generally only a single household vector is specified. Secondly, whilst the household sector may remain analytically more significant, the issues that local Economic Development Agencies are involved in, such as monitoring the role of foreign direct investment within the regional economy and local sourcing initiatives mean that the correct specification of interindustry relationships remains an important objective from a user-perspective.

Whilst there *is* some data available on regional household expenditure patterns (*i.e.* from the Family Expenditure Survey), the atomistic nature of the household sector draws a natural suspicion over the validity of such data. Consequently, the analysis concerning the household vector and any subsequent interpretations should be more restricted than the more 'observable' industrial sector. In short, it would be unwise to push the data too hard.

Thus the analysis is restricted by the following assumptions (*i*) labour incomes could (or rather should) be specified by first-best means; (*ii*) gross household expenditure patterns can be established for the region from published sources.

(*iii*) national import propensities from overseas are assumed to hold at the regional level. The issue that remains relevant therefore is how to assess the local consumption propensity of household purchases.

The conventional estimating approach applies a location quotient methodology (i.e. Johns and Leat, 1986; Garhart *et al.* 1996). As in the case of industry's demand for inputs, the degree of relative specialisation in supply has some justification in positively influencing the extent of local sourcing: standard locational arguments such as lower costs of transportation apply. However, the failure of the simple location quotient to account for factors such as regional size lead one to consider a Stevens-type RPP specification. Defining the local consumption propensities as c_i - consumers' expenditure on locally produced *i* as a fraction of consumers' spending on domestically produced *i*:

$$c_i = e^{-1/\pi_i} \qquad [5.8a]$$

where

$$\pi_i = \delta q_i^{\beta_2} \qquad [5.8b]$$

In the same vein as the previous analysis, $e^{-1/\delta}$ is anticipated to capture regional size, and β_2 reflects the extent to which the degree of regional specialisation influences local sourcing in merchandised goods. The expectations over the orientation of the values for these parameters remains unchanged. Hence, hypothesis 12 states:

$\mathbf{H_{12}}$: $c_i = e^{-1/\pi_i}$ [5.8a]

where

$\pi_i = \delta q_i^{\beta_2}$, $0 < e^{-1/\delta} < 1$, $\beta_2 > 0$ [5.8b]

As with industry's supply of inputs, the export-base argument suggests that the local consumption propensity and the responsiveness of demand to changes in relative supply for service sector products may be higher than for manufacturers. Hence, hypothesis 13 states

$\mathbf{H_{13}}$: $\delta_s, \beta_{2s} > \delta_{ns}, \beta_{2ns}$ [5.9]

One would suspect that the nature of retailing is such that the local consumption propensities out of consumers' expenditure will tend to be lower than for industry input purchases. Thus, in general, hypothesis 14 states:

$\mathbf{H_{14}}$: $\delta < \alpha$ [5.10]

5.35 Implementation and Testing of the New Technique

In terms of the interindustry matrix, with estimates of both intermediate row and column sums of the regional use, an estimate of the full matrix is given by applying the usual RAS algorithm to the nonsurvey domestic use matrix:

$$'\mathbf{R}^{*} = \hat{\mathbf{x}}\ '\mathbf{A}^{*}\hat{\mathbf{y}}$$ [5.11]

where

x $[r_i/a_i]$ is the RAS row multiplier

y $[r_j/a_j]$ is the RAS column multiplier

$a_{i,j}$ are the intermediate row, column sums of $'\mathbf{A}^{*}$ at each iteration.

The estimated regional purchase propensities for the household column are of course applied multiplicatively to the values of household domestic expenditure.

Consideration will be given as to how the parameters from each model could be specified from sample survey, thereby creating a new survey-based-nonsurvey estimation stage within the hybrid process.

The final hypothesis therefore relates to the relative performance of the 'new' technique against standard nonsurvey procedures. One would hope that the new methodology offers a more realistic account of total use and regional trade given its basis within an empirical analysis, and it should therefore be seen to estimate significantly better than its nonsurvey predecessors. Thus:

H_{15}: The 'new' third-best estimation methodology performs significantly better than its nonsurvey predecessors.

5.4 Summary

This chapter began with a statement of research objectives and general strategy. The first stage of this strategy has been implemented within the remainder of the chapter. The nature, significance, and relative importance of the components of nonsurvey error have been considered from an *a priori* perspective, and this has led to the formation of research hypotheses. Regional-national differences in production functions and regional trade are believed to be the significant contributors to nonsurvey error. Consideration has been given as to how these factors might be accounted for within a third-best estimation methodology: the RAS algorithm was selected as the most appropriate framework. Equations have been specified which, mainly through the location quotient, seek to determine values for the intermediate row and column sums of the regional use matrix and household local expenditures. Regional specialisation is believed to have a positive influence on the degree of intermediation.

In addition, it has been hypothesised that a simple consideration of the nature of commodities - for example service and non-service sectors - could provide some useful general information on trading propensities which would supplement the formula-based estimation process.

The parameters of the equations could be estimated from empirical data which would create a new survey-based-nonsurvey step within hybrid procedure.

A summary table of the hypotheses is provided in Table 5.2 below.

Table 5.2 Summary of Research Hypotheses

H_1	The error associated with total use misspecification is parabolic with respect to matrix order.
H_2	The error associated with total use misspecification is negatively related to regional size.
H_3	the error associated with the use of national overseas import propensities is parabolic with respect to matrix order.
H_4	the error associated with regional trade misspecification is largely unrelated to matrix order.
H_5	the error associated with regional trade misspecification is negatively related to regional size.
H_6	nonsurvey assumptions over total use relationships and regional import propensities are misspecified to a significant degree.
H_7	the assumption of national overseas import propensities holds at the regional level and therefore constitutes the least significant source of misspecification.
H_8	regional trade specification is relatively more important than total use specification for smaller regions.
H_9	$\frac{r_i}{u_i^{\bullet} k_i} = \alpha q_i^{\beta_0} \quad , 0 < \alpha < 1, \ \beta_0 > 0$
H_{10}	The relationship between dependent and independent variables is significantly different between service and non-service sectors.
H_{11}	$l_j = e^{-1/Z_j}, \ Z_j = \alpha q_j^{\beta_1}, 0 < e^{-1/\alpha} < 1, \beta_1 > 0$
H_{12}	$c_i = e^{-1/\pi_i}, \ \pi_i = \delta q_i^{\beta_2}, 0 < e^{-1/\delta} < 1, \beta_2 > 0$
H_{13}	$\delta_s, \beta_{2s} > \delta_{ns}, \beta_{2ns}$
H_{14}	$\delta < \alpha$
H_{15}	The 'new' estimation methodology performs significantly better than its nonsurvey predecessors.

5.5 Conclusions

This chapter has played a crucial role in laying down the path of the research exercise. It should be clear that the proposed 'new' nonsurvey approach is an extension and development of the work by Stevens *et al*. There are however four important points of difference:

(*i*) The proposed approach in the row generating equation attempts to account for differences in total use, which are expected to be significant.

(*ii*) The estimation of local consumption propensity is extended to a column constraint.

(*iii*) The estimation of local consumption propensity is extended to consider household expenditures.

(*iv*) Most crucially of all, the approach is not intended for use as a traditional single-step nonsurvey methodology. Its parameters must be specified by survey.

The empirical exercises implied by the analysis of this chapter, however, cannot yet be implemented. In order to test the given set of hypotheses, analytical tools, such as an appropriate measure of error, have to be developed, and of course, a suitable data set has to be found. These issues are addressed within the next two chapters.

CHAPTER 6

TOOLS FOR DIRECT COMPARISON ANALYSIS

6.0 Tools for Direct Comparison Analysis

6.1 Introduction

In a direct comparison analysis, the usual objective is to assess whether the estimated input-output matrix of transactions or coefficients is an 'acceptable proxy' for the observed set. The intrinsic validity of such an analysis has already been questioned. However, it would seem that as long as the analysis remains at a relatively broad level of detail it should maintain some effect. This chapter develops a matrix distance statistic that will be used to test some of the research hypotheses set out in Chapter 5.

There are a variety of measures for assessing the overall (dis)similarity between two matrices, and direct comparison studies have not hesitated in borrowing from this range. The first part of the chapter reviews some of the methods that have been employed. The review illustrative rather than exhaustive, principally because most techniques are very similar, and because Knudsen and Fotheringham (1986) provide good guidance on the general problem of matrix distance testing. The second section therefore takes the Knudsen and Fotheringham approach in deriving a range of 'critical values' for appropriate measures of error.

6.2 The Use of Matrix Distance Tests in Direct Comparison Studies

The reason why standard tests of statistical inference cannot be applied to the comparison of input-output data is that, usually, nothing is known about the distribution from which each observed element derives. Jackson (1989; 1991) provides a novel way around this by 'generating' a distribution from aggregation; West's (1982; 1986) laudable attempts to derive confidence limits for multipliers rely upon fairly restrictive assumptions about the distribution of each observed element.

The early direct comparison studies (i.e. Schaffer and Chu, 1969) attempted to utilise the chi-square distribution to assess the statistical significance in differences between observed and estimated coefficients. The general chi-square approach assesses whether the differences between observed and expected frequencies could have occurred by chance. The early direct comparison studies took the notion of 'observed frequency', equating it with the set of nonsurvey coefficients, whilst those 'expected frequencies' were taken as the survey based observations. Hence chi-square, here calculated on a matrix comparison between coefficients, is given as

$$\chi^2 = \sum_{i,j} \frac{\left(a_{ij}^* - a_{ij}\right)^2}{a_{ij}} \qquad [6.1]$$

where

* denotes the nonsurvey estimate, survey otherwise.

As with all distance measures, the comparison can be applied at a matrix level on any sub-section, such as individual rows and columns, or upon intermediate row and column sums.

A number of inherent deficiencies in this approach (see the discussion below) led to later studies, such as Smith and Morrison (1974) using the chi-square formula as a measure of relative distance rather than a mark of statistical 'goodness of fit'.

Smith and Morrison (1974) and Harrigan *et al.* (1980b) calculate the mean absolute difference between observations:

$$\lambda = \frac{1}{n^2}\sum_{i,j}\left|a_{ij}^{\bullet} - a_{ij}\right| \qquad [6.2]$$

where

n is the number of observations

Following Isard and Romanoff (1968) both studies employ a 'similarity index' which, has, for non-negative matrices, the range 0-1, where unity indicates a 'perfect' simulation:

$$s = 1 - \frac{1}{n^2}\sum_{i,j}\frac{\left|a_{ij}^{\bullet} - a_{ij}\right|}{\left(a_{ij}^{\bullet} + a_{ij}\right)} \qquad [6.3]$$

Both studies utilise an information-based statistic of the Kullback and Leibler (1951) ilk

$$I\left(A^{\bullet}:A\right)=\sum_{i,j}\left|a_{ij}\log_2\frac{a_{ij}^{\bullet}}{a_{ij}}\right| \qquad [6.4]$$

The function of [6.4] is to evaluate (Smith and Morrison, 1974 p.28)

'..the additional bits of information contained in $A^{\bullet}$, given the prior probabilities as expressed in A. The more additional information contained in $A^{\bullet}$, the less accurate it is in terms of the prediction contained in A.'

Correlation analysis is also used by both these UK studies, and by Willis (1987)

where

$$R=\frac{\sum_{i,j}\left(a_{ij}^{\bullet}-\bar{a}_{ij}^{\bullet}\right)\cdot\left(a_{ij}-\bar{a}_{ij}\right)}{\left[\sum_{i,j}\left(a_{ij}^{\bullet}-\bar{a}_{ij}^{\bullet}\right)^2\cdot\left(a_{ij}-\bar{a}_{ij}\right)^2\right]^{0.5}} \qquad [6.5]$$

- denotes a mean value of observations

The Harrigan *et al.* (1980a) study attempts to determine the validity of the assumption of identical national and regional technologies by regression analysis

$$a_{ij} = \alpha + \beta a^{*}_{ij} + \varepsilon_{ij} \qquad [6.6]$$

where

α, β are parameters to be estimated by regression.

ε is a stochastic disturbance term

Harrigan *et al.* attempt to test the avidity of the restriction α=0, β=1 which would hold under a spatially invariant pattern of technology.

Butterford and Mules (1980) provide a slightly over-the-top approach to direct comparison analysis. They recommend a battery of tests which includes a chi-square test for differences in the size distribution frequency of coefficients, regression analysis, and a mean absolute difference measure.

These examples cover the main tools used by the relatively early direct comparison studies.

There a number of problems associated with their use. Firstly, the chi-square test for goodness of fit relates to a frequency distribution, and input-output data are not frequency observations. As such the measure [6.1] is 'value sensitive'. So, for example, given a transactions value measure, where $X_j > 1$, then

$$\sum_{i,j} \frac{\left(a_{ij}^{\bullet} X_j - a_{ij} X_j\right)^2}{a_{ij} X_j} = \sum_{i,j} \frac{X_j \left(a_{ij}^{\bullet} - a_{ij}\right)^2}{a_{ij}} > \sum_{i,j} \frac{\left(a_{ij}^{\bullet} - a_{ij}\right)^2}{a_{ij}} \qquad [6.7]$$

Since the degrees of freedom do not change between comparisons (matrix order remains the same), the critical value of χ^2 does not change. Hence it is possible to induce 'statistically significant' differences simply by shifting the value measure of the data set. Chi-square is therefore inappropriate as a statistical test of 'goodness of fit'.

A more general criticism relates to formulae [6.1] and [6.4] and their variants (see Knudsen and Fotheringham, 1986). Clearly, when $a^{\bullet}_{ij} > 0$, $a_{ij} = 0$ the statistics are undefined - and one can expect this to be a likely occurrence within a nonsurvey-survey comparison. Possible solutions include assigning arbitrarily small values to a_{ij}, or aggregating relevant cells (Schaffer and Chu, 1969; Butterford and Mules, 1980).

The second criticism relates to the use of coefficients in preference to transaction values. Results from studies such as Jackson (1991) highlight the relative importance of large transactions, and thus in later studies, such as Szyrmer (1989), measures are computed across value matrices. In these later studies, one also sees the use of *standardised* measures of distance which allow greater cross-comparison between results for matrices of different orders or which contain different types of data. The ability to cross-compare results is essential to the analysis of Chapter 5. So, for example, the mean absolute difference of equation [6.2] reports a mean value in the units of the original data set (*i.e.* coefficients, £m, 1000 employees *etc.*). Equation [6.7] below however provides a standardised measure of error, and is referred to as the Standardised Total Percentage Error (STPE, *e*).

$$e = \frac{\sum_{i,j} \left| {}^{t}a_{ij}^{*} - {}^{t}a_{ij} \right|}{\sum_{i,j} {}^{t}a_{ij}} \times 100 \qquad [6.8]$$

where

t denotes a transactions flow.

Equation [6.8] may be interpreted as the mean absolute difference as a percentage of the mean observed value. Note the importance of calculating across transactions: if coefficients were the basis for [6.8], then the denominator of the equation would change with aggregation and comparison across a range of matrix orders would be extremely difficult. Miller and Blair (1983), Israilavich (1986), and Szyrmer (1989) each find favour with [6.8].

The use of the general regression equation [6.6] seems questionable, given that the distribution of a_{ij} will probably consist of a few relatively large coefficients and many small observations. Indeed Butterford and Mules (1980) illustrate that the acceptance of the null hypothesis varies with different sets of size interval observations. Moreover, given that a high level of measurement error can be expected within the partitive input-output data set, the ordinary least squares regression parameters will be biased towards zero (Pindyck and Rubenfeld, 1976, p.130).

Therefore, it would seem that measures such as the correlation coefficient, standardised total percentage error and the similarity index emerge as preferred measures on the ground that their application is relatively problem free.

However, the experiments by Knudsen and Fotheringham (1986) help to narrow this choice.

The essential problem with using the given measures of goodness of fit is in interpretation. What does a value of R=0.35 actually mean in terms accepting or rejecting the null hypothesis of equality between observed and estimated flows? Essentially what each statistic lacks is a range of critical values.

The authors consider the introduction of different levels of random error into a an observed matrix of flows. They note the first desirable property of any distance statistic as (p.134)

'An "ideal" goodness-of-fit statistic ... would be one for which the relationship between the value of the statistic and the level of error is linear'

This facilitates general comparison. So, for example, if the estimated transactions matrix m is associated with a value of the distance statistic equal to x, and the estimated matrix n is associated with a distance $2x$, then one should be able to conclude that the accuracy of m is twice that of n.

Knudsen and Fotheringham therefore introduce different levels of error into a known flows matrix and calculate values for a range of distance measures. They are able to conclude that R^2, chi-square, and the information gain statistic are not satisfactory in this respect because they exhibit non-linear responses to induced levels of error. They do not consider the STPE or the similarity index.

Thus, with an association between error level and the value of each statistic, the second section of Knudsen and Fotheringham's experiments consider the circumstances under which statistic values may be interpreted as 'significant'. The authors state two general decision rules concerning the test of a hypothesis (p.139):

'(1) a ... model should not be retained when error in the estimated matrix exceeds 50 percent; (2) a ... model should not be rejected when error is less than 10 percent.'

The range 10-50% would require a more explicit test of significance in order to determine the validity of the null hypothesis. Without this test, the comparison must remain inconclusive. The decision rules are, of course, subjective but seem to provide a sensible framework within which to operate. More is said below.

Critical values were therefore derived for the STPE and similarity index following the Knudsen and Fotheringham methodology.

6.3 The Derivation of Critical Values for Matrix Distance Statistics

The methodology was as follows. A 40 by 40 matrix of 'transactions' between 0-1000 was initially generated from a uniform distribution. This was taken as the observed transactions matrix, **O** $[o_{ij}]$.

The elements of **O** were then subject to the introduction of a random percentage error in order to generate an 'estimated' transactions matrix **P** $[p_{ij}]$. The random percentage disturbance was drawn from a uniform distribution

$$\varepsilon_{ij} = 1 - \alpha + 2r_{ij}\alpha \qquad [6.9]$$

where

α is the predetermined percentage error divided by 100 (*i.e.* α=10%=0.1)

r is a random number, drawn from a uniform distribution with limits 0-1

It should be clear from [6.9] that the disturbance has the range

$$(1-\alpha)\leq\varepsilon_{ij}\leq(1+\alpha). \qquad [6.10]$$

The random disturbances were then applied multiplicatively to the elements of **O** in order to form **P**

$$p_{ij} = o_{ij}\varepsilon_{ij} \qquad [6.11]$$

Distances were then calculated between **O** and **P** by using the STPE and a version of the similarity index which, for comparability with STPE was calculated as

$$s = \frac{1}{n^2}\sum_{i,j}\frac{\left|a_{ij}^{\bullet} - a_{ij}\right|}{\left(a_{ij}^{\bullet} + a_{ij}\right)} \times 100 \qquad [6.12]$$

The distance was recorded and the process repeated a further 19 times. This gave 20 values for each distance statistic at a known level of percentage error. The experiment was repeated at different levels of error. The results are plotted in Figures 6.1 and 6.2 below. Figure 6.3 compares the response of each statistic to changes in the level of error.

Figure 6.1 Observed Values of the Similarity Index

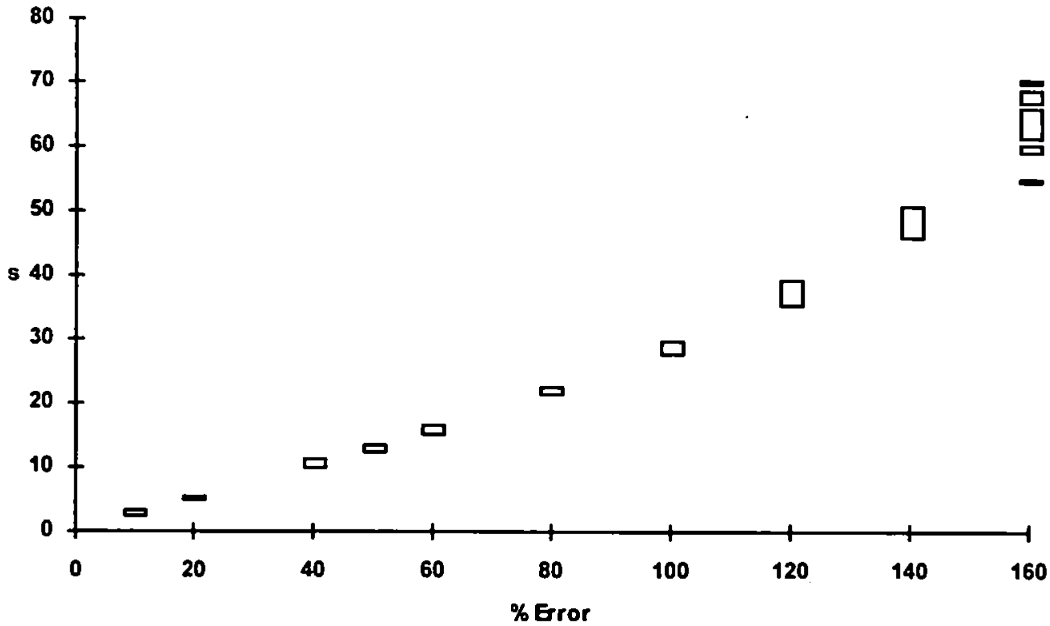


Figure 6.2 Observed Values of the Standardised Total Percentage Error

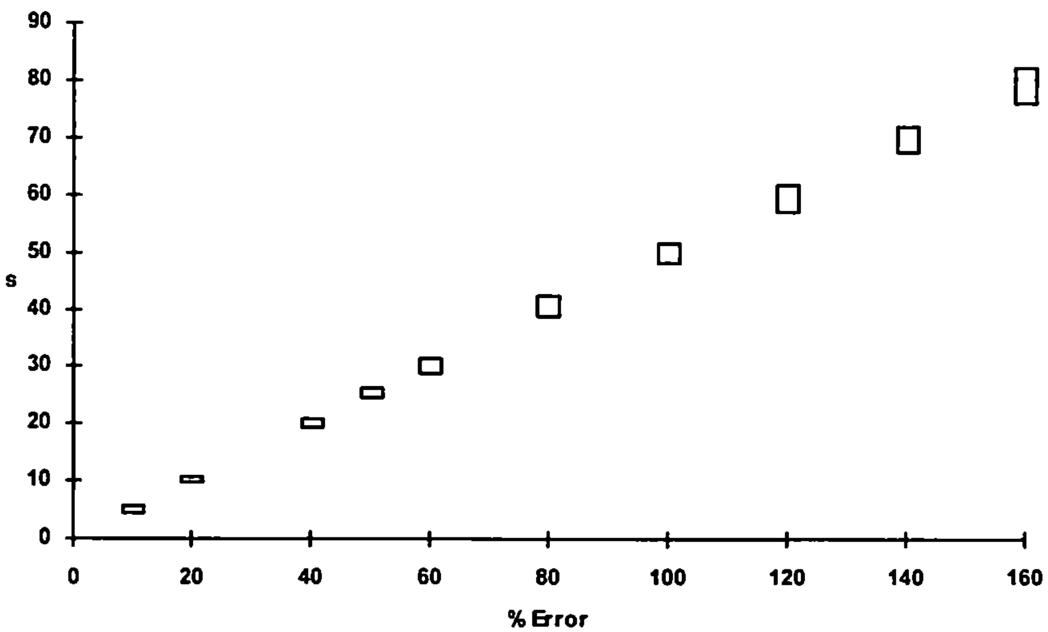
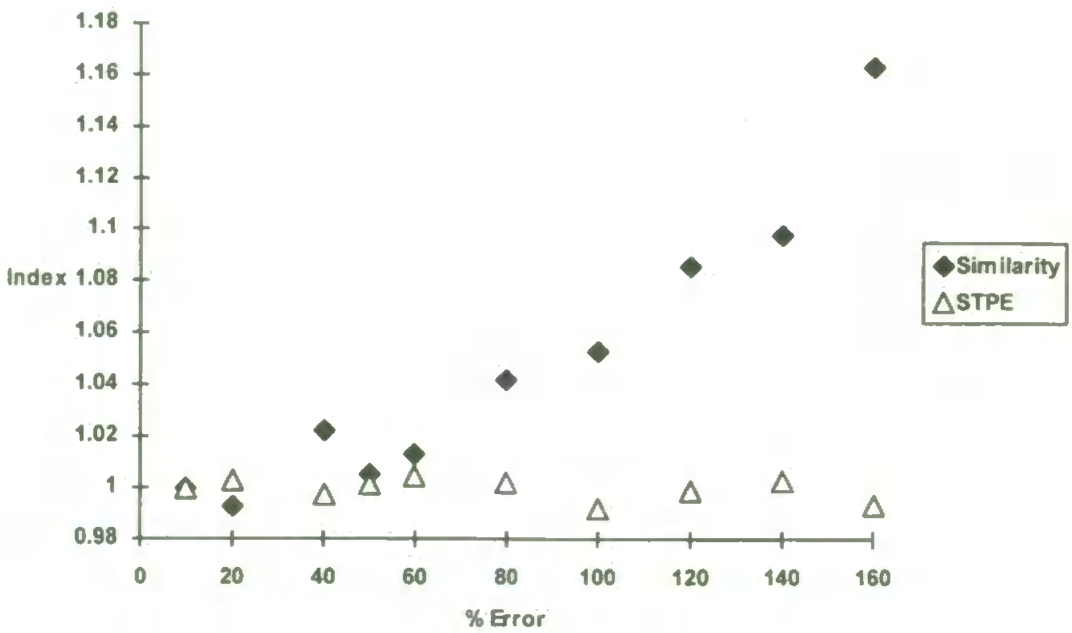


Figure 6.3 Sensitivity of Statistic Mean to Change in Induced Error Level



The relationship between the similarity index and the level of error is non-linear. The STPE exhibits a more or less linear relationship. Figure 6.3 provides the comparative analysis. The relative change in the mean of the observed statistics is expressed as a ratio of the relative change in induced percentage error: a 1:1 relationship signifies linearity. Quite clearly, the similarity index becomes more sensitive at higher levels of error. This is reflected in the increase in the variance of the observed values of s at higher levels of error; the STPE seems to exhibit some increase, but its mean value is clearly much more stable. This probably has something to do with the fact that, at errors in excess of 100%, the elements of the estimated flows matrix can become negative, and this destabilises the denominator of [6.12]. However, even at errors *less* than 100%, the mean value of the similarity index is still relatively unstable. The STPE was therefore the preferred statistic.

Table 6.1 below reports the mean (*i.e.* the 'critical values') and sample standard deviation of the distribution of observed STPE.

Table 6.1 Experimental Critical Values for STPE

α%	10	20	40	50	60	80	100	120	140	160
e	4.99	10.00	19.96	24.99	30.12	40.26	49.93	59.83	70.00	79.46
sd(e)	0.112	0.136	0.317	0.357	0.559	0.843	0.714	0.987	0.951	1.610

The relationship between STPE and the level of error is quite clear: an observed STPE of e can be associated with the introduction of a uniformly distributed random error of $2e$%. Thus, within the Knudsen and Fotheringham broad criteria, as observed values for STPE exceed 25, the null hypothesis of equality between observed and estimated transactions can be rejected with increasing force. Note that the limit for automatic rejection of 50% is fairly generous: so, with an observed transaction of £1000m, the estimate should fall within £500m either side in order to avoid automatic rejection. Observed values of STPE of less than 5 should be result in automatic acceptance of the null hypothesis of equality: the estimate falls within £100m either side of the observed £1000m, and is judged to be within the reasonable bounds of stochastic error.

Two main concerns remain with respect to the use of STPE as a broad significance test. The first is, does the mean and variance of the statistic remain stable as the range of values for the observed flow matrix is made variable? So, for example, the above critical values relate to a flows matrix with the range 0-1000. But if this range becomes 0-1000000, do the critical values change at all?

In order to provide some indication, the initial experiment was extended slightly. Observed flow matrices, of order 40, were generated from uniform distributions with ranges 0-10000 and 0-1000000. Errors were introduced into these flow matrices at the 50% level and the STPE between observed and estimated matrices were calculated. This was repeated 19 times.

Assuming that the distribution of observed statistics is approximately normal[1], standard tests of statistical inference can be used to determine whether the mean and variance between these samples is stable. The test of means is a *t*-test which assumes unequal variances. Following Berenson and Laidler (1986 p.380) this is given as:

[1] The samples passed the Anderson-Darling normality test comfortably. See DAugostino and Stevens (1986) for detail.

H$_0$: $\bar{e}_1 = \bar{e}_2$

$$t = \frac{\bar{e}_1 - \bar{e}_2}{\sqrt{\frac{\delta_1^2}{n_1} + \frac{\delta_2^2}{n_2}}} \qquad [6.13]$$

where

$n_{1,2}$ is the number of observations in sample 1, 2

δ^2 is the sample variance;

- denotes a mean value.

The mean of the sample (0-10000) was calculated at 25.065 and had a variance of 0.248. Comparison with the mean from the sample (0-1000), 24.985, variance 0.128 gave a t-value of 0.584 which, on a 2-tailed test gave a critical value for t of 2.032 and therefore H$_0$ could be accepted, assuming a 5% level of significance. The mean of the sample (0-1000000) was calculated at 25.08 and had a sample variance of 0.195. The calculated value of t between the means of this and the (0-1000) sample gave a value of 0.746; the associated two-tailed critical value was 2.028 and therefore H$_0$ could be accepted at the 5% level of significance. The variability in the range of the observed flows matrix therefore had no effect on the sample mean of e.

The test of variance is an F-test, which is given simply the ratio of sample variances. Following Berenson and Laidler (1986 p.373):

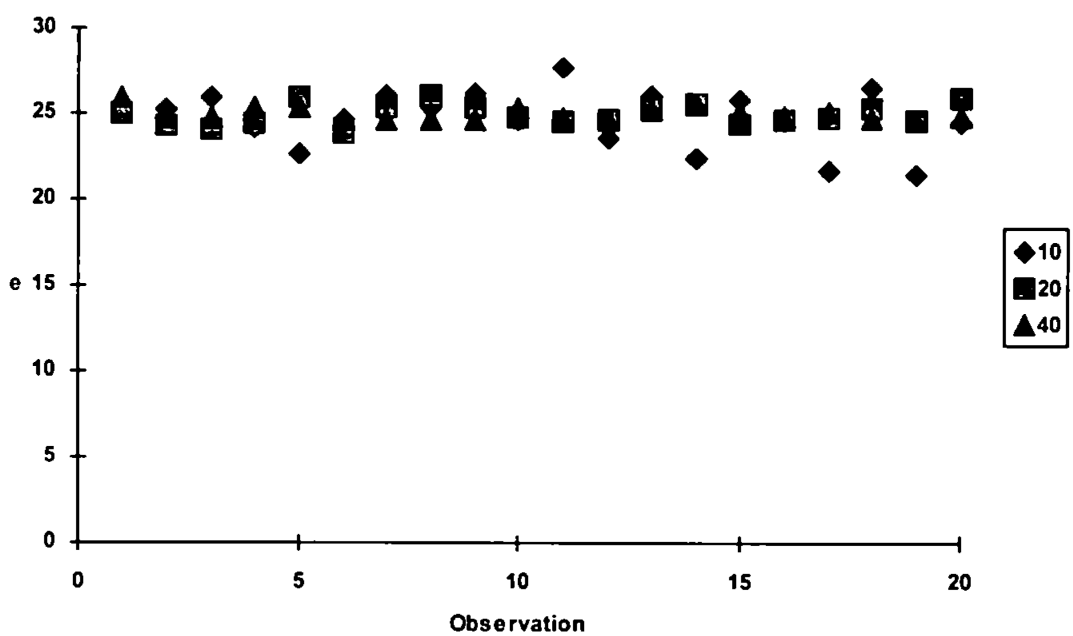
H$_0$: $\delta_1^2 = \delta_2^2$

$$F_{(n_1-1,n_2-1)} = \frac{\delta_1^2}{\delta_2^2} \qquad [6.14]$$

The statistic has degrees of freedom (n_1-1) in the numerator and (n_2-1) in the denominator. The calculated value for F between (0-1000) and (0-10000) was 1.940, with the critical value for a two-tailed F(19,19) at 5% significance equal to 2.526. For the groups (0-1000) and (0-1000000) the F-ratio was 1.530. Therefore, the null hypothesis of equal variances could be accepted. Variability in the range of observed flows did not significantly affect the variance of the STPE.

The second concern is, does the mean and variance of STPE vary with matrix order? Clearly, in the case of the variance, this *is* going to occur because, essentially, the sample size upon which STPE is calculated is much lower. Therefore, as STPE is calculated across a smaller range of observations, it should become more difficult to associate the observed STPE with its associated level of random error. Clearly, one could go into great detail on this issue, and this is not proposed. However, a few exploratory calculations were made. Sets of 20 observations for STPE were calculated for matrices of orders 20 and 10. The means for these samples were 24.935 and 24.730 respectively, neither of which were significantly different from the 40 order mean (t values, 0.305, 0.671 respectively). The sample variance however did increase as expected. The sample variance at order 20 was 0.410, F = 3.21, which was significantly different from 0.128 at 5% significance, but could be considered equal at 1% on a two-tailed test. The 10 order matrix yielded a sample variance of 2.757, $F = 21.6$, which was significant at 5%. However, as Figure 6.4 below illustrates, the increase in variation is unlikely to seriously compromise the intended use of STPE because the 'significance test' is so broad. However, caution on interpreting observed values of STPE across low order matrices should, naturally, be applied.

Figure 6.4 Observed Values of STPE with Variable Matrix Order



6.4 Conclusion

In the absence of information on the underlying distribution of an observation in an input-output table, direct comparison analysts have relied upon measures of matrix distance which, generally, are not associated with a test of statistical inference. This section has reviewed the main measures that have been used in direct comparison analysis. The correlation coefficient, similarity index, and Standardised Total Percentage Error emerged as the preferred measures. However, experiments by Knudsen and Fotheringham suggested that the correlation coefficient was not well suited to matrix distance assessment. Following their methodology, experimental critical values for the similarity index and STPE were derived. The STPE was found to be the preferred technique. The guidelines for using STPE are very broad - but this creates no real problem because the nature of the intended direct comparison analysis will, likewise, be broad. Following Knudsen and Fotheringham, an estimated transactions matrix will be considered 'significantly different' from its observed counterpart at errors exceeding 50% (STPE>25); the estimated matrix will be considered equivalent to the observed set at error levels below 10% (STPE<5). Errors between these values are indeterminate.

The next chapter prepares the data for the analysis of Chapter 8.

CHAPTER 7

THE SELECTION AND PREPARATION OF DATA FOR ANALYSIS

7.0 The Selection and Preparation of Data for Analysis

7.1 Introduction

Before the hypotheses of Chapter 5 could be tested, suitable data sets had to be selected. The basic data requirements for the analysis were a set of national input-output tables and a set of UK regional input-output tables. Clearly, the two tables had to meet a number of consistency criteria inorder to enable comparison. In reality, the selection of the regional model was something of a foregone conclusion since there is only one recent survey-based UK regional model - that for Scotland, 1989 (Scottish Office, 1994). Nevertheless, an assessment of its comparability with national data was clearly necessary because of the strong implications for analytical procedures and the interpretation of results.

The genre of direct comparison studies which precede this exercise seem to overlook the issue of data comparability, or at least they give only a passing consideration to a single issue - that of aggregation. However, it should be clear that there are a number of other issues to be considered. The first section of the chapter therefore considers the criteria which should be met in order to allow direct comparison between two input-output tables. The second section reviews the features of published national and regional input-output data in the UK. The third section assesses the comparability of these data, and therefore provides the basis for selection. The remainder of the chapter gives details of the transformation procedures that were necessary to achieve comparability.

7.2 Criteria for Comparability

There are essentially four main features of the input-output table which should be consistent if one is to draw comparison between them. The first is consistency by time period.

Input-output tables of course relate to a given period, usually a year. Whilst it is perfectly feasible to make comparisons between input-output tables for a given area over time (see for example Sonis and Hewings, 1989), in the current context - that is, drawing comparison between an observed regional model and an estimate from the national table - it is desirable for the two tables to relate to the same year. If this were not the case then one would not be able to distinguish between differences that were due to the misspecification of the given simulation function, and those that were due to structural changes that had occurred over time (*e.g.* technical progress).

The second feature is similarly obvious, and that is that the definitions of the respective tables must be consistent.

The third feature relates to the system of accounts. Hewings and Jensen (1986) illustrate a number of possible way of representing the input-output accounting system. The most popular of these consists of the make and use matrices. Briefly, each is defined with commodities in rows, industries in columns - hence the term 'commodity-by-industry'. The make matrix shows the composition of industrial output in terms of commodities (*i.e.* industries produce a range of commodities which are essentially either principal, secondary or by-products). This distinction of means that the tables are not symmetrical - that the output of each commodity will not be equal to the output of the respective industry. The intermediate section of the use matrix shows the structure of each industry's input purchases in terms of commodities - distinction may be given to purchases from regional, domestic (*i.e.* imported from the UK) and overseas sources in the form of separate use matrices (*i.e.* 'the regional imports use matrix'). The two tables may be manipulated in order to form a symmetrical input-output matrix (see the Scottish Office, 1994 for details). The two points of relevance are that it is not strictly possible to (*i*) directly compare symmetrical and non-symmetrical tables, or tables where one is commodity-symmetrical and the other industry-symmetrical; (*ii*) derive the Leontief inverse matrices from commodity-by-industry tables. This latter point will be returned to. However, tables should share a common system of commodity/industry classification.

Finally, tables should share a common system by which goods and services are valued. In general, the price of a good as it is purchased by the final user will be different from its price as it was sold at the factory-gate. The final price of a good will include transporter and distributors' margins, taxes on expenditure less subsidies - costs which are not included in the factory-gate price. Final-user prices are referred to as purchaser prices; factory-gate prices are referred to as producer prices. When production taxes are removed from producer prices, the goods are said to be valued at basic prices. Clearly, because sales and purchases are represented simultaneously in the input-output matrix, a single valuation assumption must be imposed.

7.3 Published Input-Output Tables in the UK

This considers the data that is available to the analysis. The section is split between a consideration of national and regional tables.

7.31 National Input-Output Tables

The first official UK input-output table was published in 1961 and related to the year 1954. The early publications concentrated upon the manufacturing sectors of the economy. However, the relative growth of the UK's service sector has meant that subsequent publications have described services in increasing detail.

Symmetric tables have been published for the years 1963, 1968, 1974, 1979, 1984 and 1990. In 1989, a symmetric table for 1985 was derived from the 1984 accounts using the RAS procedure. This was mainly to correct for the distortions of the miners' strike of 1984-5.

Tables which relate to years prior to 1989 were constructed using National Accounts data as a constraint. Since then, the ONS has produced an annual 'input-output balance' (see below) which has become an integral part in the construction of consistent National Accounts. The next two sections considers the two most recent UK publications at the time of writing.

7.311 The 1989 UK Input-Output Balance (ONS, 1992)

Input-output balances are distinct from the symmetrical tables. They appear as a purchaser price combined use matrix, domestic output at producers' prices, and total commodity supply tables. The latter of these tables shows overseas imports, distributors' trading margins, and expenditure taxes for each commodity. The domestic output table gives some very broad indication of the form of the make matrix, showing the value of the principal diagonal. The tables are defined across 102 sectors, details of which can be found in Table A7.1 in the appendix to this chapter.

7.312 The 1990 UK Input-Output Tables (ONS, 1995)

This publication consists of five main tables. The make matrix is accompanied by commodity-by-industry domestic and imported use matrices. Commodity-symmetrical tables are also derived.

The tables are defined across 123 productive sectors. This represents a significant expansion of the 102-sector definition used prior to 1990. The main area of expansion is in the definition of service sectors, specifically the SIC divisions 8 and 9 relating to business services and general government services. However, forestry and fishing have been allocated separate categories; the retail and distributive trades (SIC division 6) are expanded from two to four categories; and there has also been some minor reclassification of the energy industries. Table A7.1 at the end of this chapter illustrates.

The other main difference between the 1990 table and its predecessors is the price-valuation assumption. Previous symmetric tables had been valued at producer prices. Transactions in the 1990 table however are valued at basic prices.

7.313 Sources and Accuracy

A discussion of the sources and methods used to construct the UK input-output tables can be found in any one of a number of articles published in *Economic Trends*, for example ONS (1992). Briefly, the main sources are: the Annual Census of Production; the Purchase and Sales Inquiries; MAFF and the Forestry Commission; the Annual Census of Construction; the Annual Wholesale, Retail and Motor Trades Inquiry; the Annual Catering Enquiry; the Annual Service Trades Inquiry; the Family Expenditure Survey; data from the Bank of England, tax and employment data, company reports, and from Government expenditure records.

The ONS quite openly acknowledges that the data in the national table is of variable quality, and indeed approximations are often used. For example, the ONS (1995) concede that the allocation of imported inputs to specific industry groups in the imports use matrix is a somewhat indefinite procedure; distributors' trading margins are also subject to approximation. It would of course be a mistake to view the national, or indeed any input-output table as representing 'truth'. However, the analytical exercise has been designed with this in mind.

7.32 Regional Input-Output Data: 1989 Scottish Input-Output Table (Scottish Office, 1994)

As it has already been mention the 'selection' of a regional table was a foregone conclusion - the only recently produced survey-based UK regional table at the time of writing was for Scotland, 1989. Indeed, Scotland is the only region in the UK to have any input-output history. Tables were produced for a number of years during the 1970s, with the last publication relating to 1979. The publication of the 1989 tables marks the beginning of a major commitment to the input-output model by the Scottish Office. Tables are currently being prepared for 1992 and they will subsequently appear on an annual basis. Social and Environmental Accounting Models have also been developed from the basic input-output structure (Scottish Office, 1995).

The Scottish tables consist of six main tables. The commodity-by-industry tables are the make, regional use, imported use from the UK, and imported use from overseas. Both commodity and industry symmetrical tables are derived. All tables are defined across 114 productive sectors. Details can be found in Table A7.1 at the end of the chapter. All matrices are valued at producer prices.

7.321 Sources and Accuracy

In terms of the main objectives of the research exercise, that is, to identify and explain the relationship between regional and national input-output tables, clearly, if the regional data has been generated through proxy measures and national data sources, then the analysis is invalidated.

Of course, the data restrictions that create problems at the national level are exacerbated at the regional level. Consequently, a regional model with even a relatively large budget behind it will inevitably be constructed using a proportion of data that has been derived from national input-output tables. And so the question here should not be: 'is the regional model based entirely upon survey information?' but rather: 'does the regional model contain sufficient regional-specific data to allow a reasonable assumption of independence from national data sources?'

Volume Two of the Scottish Office (1994) publication provides a full discussion of the sources and methods used to construct the Scottish tables. A number of points make it clear that the model is sufficiently independent of any national table. Firstly, it was possible to extract the Scottish returns from a number of the UK inquiries listed above. Secondly, Scotland has a number of regional-specific government departments and companies which were able to provide relevant information, in particular, the Scottish Office Agriculture and Fisheries Department, Scottish electric and nuclear power companies, Scotrail, and the Scottish banks. Thirdly, four surveys of industry, covering manufacturing, construction, wholesaling and retailing and financial services were undertaken, with the principal objective of determining regional trade flows. These surveys were quite extensive, for example, the Manufacturing Trade Flows Survey covered 42% of manufacturing gross output, and the reported response rates are generally good. Indeed, there appears to have been minimal reliance on the 1989 UK table, with only the purchases of the heterogeneous 'other services' group having been wholly borrowed. In a number of cases, proportions from the 1979 Scottish tables were applied. It would therefore seem reasonable to assume that the Scottish and UK data sets are independent of each other. However, to attach any idea of partitive accuracy to the Scottish model is as inappropriate as it is for the national model.

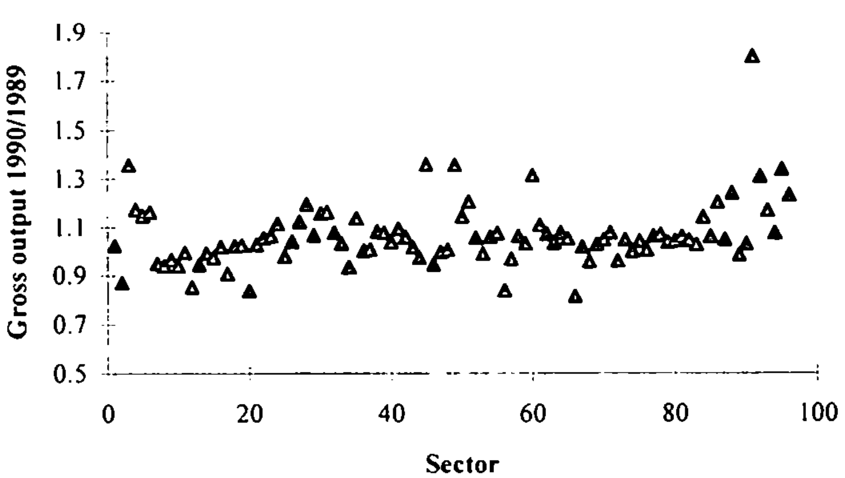
7.4 Data Comparability

It should be clear from the preceding section that neither the 1989 and 1990 UK tables were fully comparable with those for Scotland 1989. In the case of the 1990 UK tables, the time period represents the principal difference; for the 1989 tables, the difference in pricing convention and the absence of an imports use matrix are the main obstacles to comparability.

The first point to consider was whether the difference between 1989 and 1990 was likely to be significant. If it was not significant, then the 1990 UK tables would be the favoured source of comparison.

In order gain some broad inference on the changes that occurred between 1989 and 1990, one can compare the producer price estimates of gross industrial output that are published in the respective balance articles. The comparison is made across 96 sectors - the schematic details are not particularly important - suffice it to say that the aggregation mainly occurs within the service sectors. Figure 7.1 below, which shows the change in gross industrial output, reveals that there was a varied pattern of growth and decline across industrial sectors between 1989 and 1990. Sectors such as electricity, which was privatised during the period, showed a marked growth in output (36% growth for the combined sector coal, electricity and mineral oil) whilst others, such as metal industries (sectors 7 to 10) showed decline (around 5.5%). Generally speaking, the service sectors, which are located at the latter end of the distribution, showed significant growth.

Figure 7.1 Change in Gross Industry Output, UK 1990/1989



The degrees of change illustrated in Figure 7.1 were felt to be too large to ignore. It was felt that 1989-based UK data should be used wherever possible. However, on the face of it, the UK 1989 data set seemed to fall a long way short of the comparability criteria. There were three main concerns: symmetry, price valuation, and the absence of a 1989 UK imports use matrix.

These problems are considered in that order. Two short sections deal with the adjustment for financial services and aggregation.

7.5 Preparation of Data

7.51 Non-symmetrical versus Symmetrical Tables

The 1989 UK tables were of the commodity-by-industry form. As it has been mentioned, it is not strictly possible to derive the Leontief inverse - and hence multipliers - from such tables, *i.e.* the system of equations represented by [7.1] cannot be solved for either vector of gross output.

$$ {}^{c}\mathbf{B}\hat{\mathbf{X}}^{i} + \mathbf{f} = \mathbf{X}^{p} \qquad [7.1]$$

where

${}^{c}\mathbf{B}$ is the commodity-by-industry matrix of use coefficients;

$\mathbf{X}^{i}$ is a vector of gross industry inputs;

$\mathbf{X}^{p}$ is a vector of gross commodity outputs;

$\mathbf{f}$ is a vector of final demands.

There are essentially three solutions. The first is to derive a make matrix which describes the composition of commodity and industry outputs. The 1989 UK balance article allows the construction of the principal diagonal of the make matrix - all off-diagonal elements would have to be estimated. One feasible approach would be to constrain the 1990 make matrix to 1989 dimensions using RAS. The second is to make some simplifying assumption about the relationship between $\mathbf{X}^{i}$ and $\mathbf{X}^{p}$ in 1989 such as $\mathbf{X}^{i} \equiv \mathbf{X}^{p}$: the matrix ${}^{c}\mathbf{B}$ could now be said to represent the coefficients of a symmetrical table. The third option is to simply exclude multipliers from the analysis.

Whilst it would be true to say that the analysis of survey and nonsurvey multipliers has formed a substantial part of the direct comparison genre of studies, such analyses offer very few insights into the nature of estimation error above and beyond those afforded by the comparison of direct transactions. Given the power series expansion of equation [2.7], apart from the multiplicand, the direct transactions will form the most significant round of multiplication. Given [2.10], errors in income and employment multipliers are largely re-weighted errors in output multipliers, which offer no particular further insight. Moreover, [2.7] would represent a subjective transformation to the 1989 data set and a potential source of error. Thus, whilst the most feasible approach would be to RAS-constrain the 1990 matrix, it would seem that the programme of analysis could be satisfactorily completed without calculating multiplier values. Where a multiplier analysis would be useful is in an assessment of the influence of errors in household consumption coefficients. However, this data - in particular that derived for the UK - was not felt to be sufficiently robust to make such an analysis worthwhile (see Figure 7.4 below). The commodity-by-industry format was therefore accepted as the framework for analysis.

7.52 Derivation of a Combined Producer Price Use Matrix from the 1989 UK Input-Output Balance

The goods and services in the 1989 UK tables were valued at purchaser prices, those in the Scottish input-output tables were valued at producer prices. It was however possible to transform the UK data into producer prices. This involved two main stages: the removal of taxes upon expenditure; the allocation of distributors' margins.

Aggregate information on the value of these items for each commodity is contained within the 1989 balance articles. The ONS originally indicated that these items could be dealt with on a *pro rata* basis. However, it soon became clear that this was not an appropriate transformation - certain advisory wires had become crossed. What follows therefore is an account of the initial derivation of the UK producer price matrix and the revised approach. Unless otherwise indicated, the transformations were conducted at the 102 sector level, which can be inferred from Table A7.1.

7.521 Taxes on Expenditure

The first step in this transformation was to determine a figure for gross taxable expenditure on each commodity. A comparison of the tax rows of the 1990 purchaser price and basic price use matrices revealed that the stock adjustment and the sales of exports should not form part of gross taxable expenditure (a zero entry in the tax row of the basic price matrix for these two categories suggested that this was the case). For all other elements, the value of taxation in the basic price matrix was greater than at purchaser prices, which implied that expenditure by these groups had been subject to taxation. Total taxable expenditure, U^{*}, was therefore calculated for product i as:

$U_i^* = X_i - S_i - E_i$ [7.2]

where

X is the total supply of the product;

S is the physical increase in the value of stocks;

E are export sales.

The next step was to form a vector of coefficients, **t**, which described the ratio of net expenditure to gross taxable expenditure for each product. The vector **t** had a typical element t_i, given by:

$t_i = 1 - T_i/U_i$ [7.3]

where

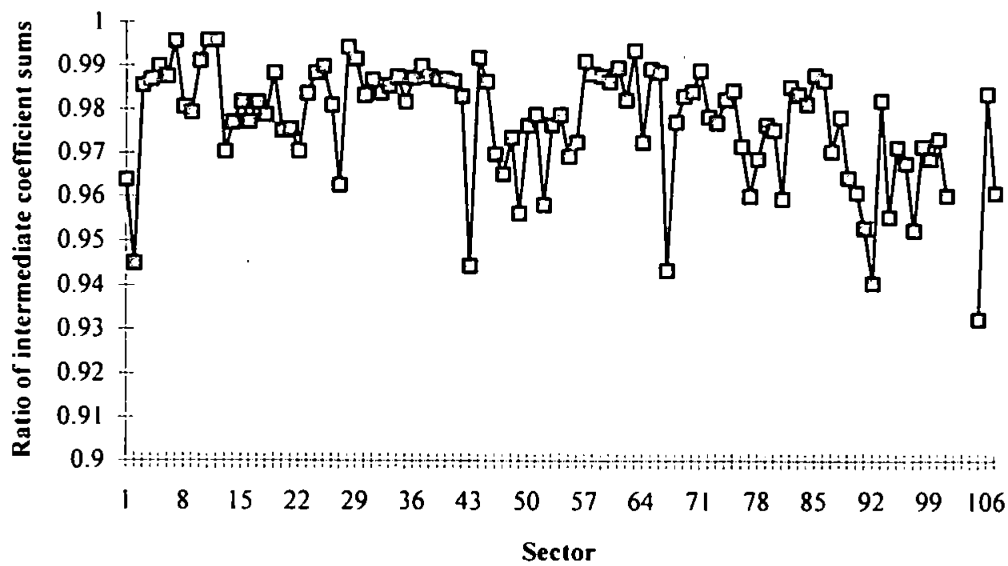
T_i is the value of expenditure tax on product i.

Letting $^{\prime}\mathbf{P}$ denote the n x m combined purchaser price use matrix in transactions, where n is the number of commodities, m equal n plus the number of final demand vectors that are subject to taxation, the net-of-tax combined use matrix under the *pro rata* assumption is then given by:

$^{\prime}\mathbf{D} = \hat{\mathbf{t}}^{\prime}\mathbf{P}$ [7.4]

Some £36,340m, or 4% of gross intermediate output had been redistributed to primary inputs. The effect this had upon the coefficients of the use matrix was relatively minor. Figure 7.2 below shows the proportional reduction in the sum of the intermediate coefficients that resulted from the tax adjustment. It can be seen from this that the reduction was no more than 6% for any intermediate sector (in the chart, sectors beyond 102 relate to final demand categories).

Figure 7.2 Proportional Reduction In Intermediate Sums of Column Coefficients Through Tax Adjustment



7.522 Distributors' Trading Margins

The methodology for the reallocation of trading margins from non-distributive to distributive sectors was similar to the adjustment for expenditure tax. Again, the first step was to calculate a figure for gross expenditures subject to distributors' margins for each product. Comparing the 1990 purchaser price and basic price matrices suggested that only the stock adjustment should be excluded from this calculation - all other sectors showed a positive increase in sales by the distributive sectors at basic prices. The matrix **'D** above was therefore expanded to include the export column. The next step was to remove the two rows that related to distribution from the matrix **D**, as clearly, they would *receive* the margins as a final residual calculation. Matrix **D** now had dimensions k x l where k equals n-2 and l equals m+1. The sum of gross expenditure subject to distributors' margins O was therefore given for product i as:

$$O_k = \sum_l d_{kl} \qquad [7.5]$$

where

d_k $\quad {}'\mathbf{D}[d_{kl}]$

The vector **z** of coefficients showing the value of expenditure net of distributors' margins to gross expenditure therefore had a typical element:

$$z_k = 1 - Z_k / O_k \qquad [7.6]$$

where

Z_k $\quad$ is the value of the distributors' margin for product k

The producer price combined use matrix, under the *pro rata* assumption of apportionment, $'\mathbf{C}$ could now be formed by:

$$'\mathbf{C} = \hat{\mathbf{z}}'\mathbf{D} \qquad [7.7]$$

The distributive rows were then inserted back into the matrix $'\mathbf{C}$, $[c_{ij}]$ which allowed the row of distributors' trading margins to be calculated as a residual. So, for industry j:

$$Z_j = X_j - \sum_i c_{ij} - I_j \qquad [7.8]$$

where

I_j $\quad$ is the value of primary inputs for j.

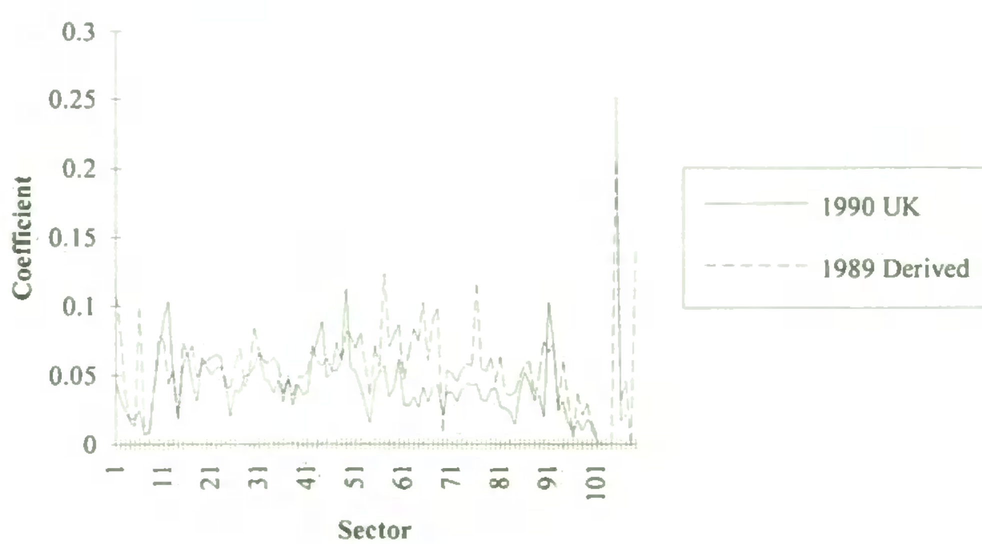
The final task was to allocate a proportion of the value of Z to each of the two distributive sectors. Table 2 of the 1989 balance revealed that the total value of distributors' margins was £89,592m, 94.9% of which accrued to sector 89, distribution and repair; the remaining 5.1% to sector 90, hotels and catering.

For each industry, the value of Z was assigned to the row elements on this basis of apportionment.

Once the column relating to stocks had been reintroduced to matrix **C**, the methodology was checked by comparing the row sums of **C** with the figure for combined gross output given in Table 2 of the 1989 balance. The figures revealed no discrepancies.

However, further checks for consistency with the 1990 table were carried out, and it was at this stage that a serious flaw in the allocation methodology was detected. The first check which gave an indication of error was the comparison of the 1990 distribution row coefficients and the estimated 1989 set. Figure 7.3 below illustrates.

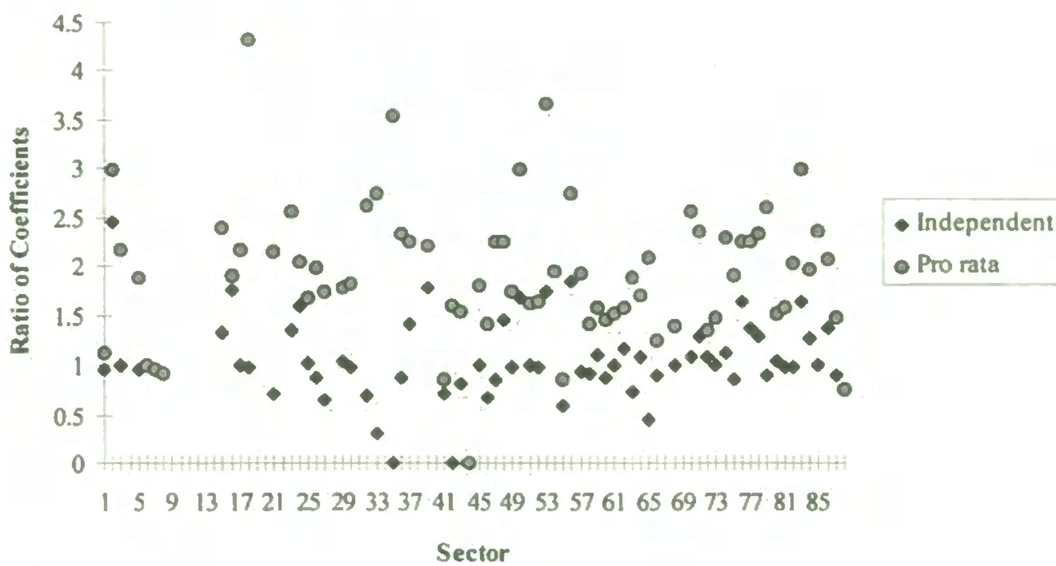
Figure 7.3 Sales Coefficients of the Distributive Sectors: 1989 Estimates and 1990 Actuals



Clearly, over certain portions of the row, particularly sectors circa 50 to 100, there would appear to be some systematic upward bias in the 1989 set. A closer examination of the final demand coefficients revealed that the most probable source of the problem lay with the underestimation of the 1989 coefficient for consumers' expenditure on distribution. The 1990 combined coefficient stood at 0.27 (£94,229m); the estimated 1989 figure at 0.21 (£68,023m). Although in coefficient terms, this difference seems slight, it was sufficient to suggest that some £20,000m had been misallocated. The most likely explanation for the error seemed to be that the proportion of distributors' margins accruing from consumers' expenditure was significantly greater than the simple *pro rata* allocation allowed. The solution was therefore to estimate the producer price vector of consumers' expenditure *independently* and to apportion the residual distributors' margins across the rest of the matrix using the above methodology.

The problem was how to make some reasonable assessment of the portion of distributors' margins that were contained within the purchaser price valuation of consumers' expenditure. Information from the 1990 purchaser price and basic price combined use matrices was used as a proxy. Firstly, tax on expenditure was removed from the value of purchaser price consumers' expenditure, on a proportional basis, for each product. The difference between this value and the value of the basic price consumers' expenditure must then be the distributors' margin. This can then be expressed as a proportion of total distributors' margins for each commodity. These proportions were then applied to the value of distributors' margins for 1989 and the resulting figure was subtracted from the estimated figure for consumers' expenditure (net of expenditure tax). This gave a producer price estimate of combined consumers' expenditure for each product in 1989. However, in ten out of the eighty-eight cases, where 1989 consumers' expenditure was small or indeed zero, this resulted in a negative estimate, in which case, either a zero was imposed, or a value was derived using the 1990 coefficient for combined consumers' expenditure. The figure for consumers' expenditure on distribution in 1989 was then calculated as a residual, and had a value of £88,418m - a coefficient of 0.27. This gave some indication that the problem had been reconciled. However, it seemed wise to make a full check on the consumers' coefficients before apportioning the remainder of the margins. Figure 7.4 below plots (*i*) the ratio of 1989 *pro rata* estimated consumer coefficients to 1990 consumers' coefficients; (*ii*) the ratio of the independently estimated coefficients to 1990 consumers' coefficients. An average one-to-one relationship would suggest successful estimation.

Figure 7.4 Ratio of 1989-1990 Consumer Expenditure Coefficients Under Alternative Estimation Methodologies

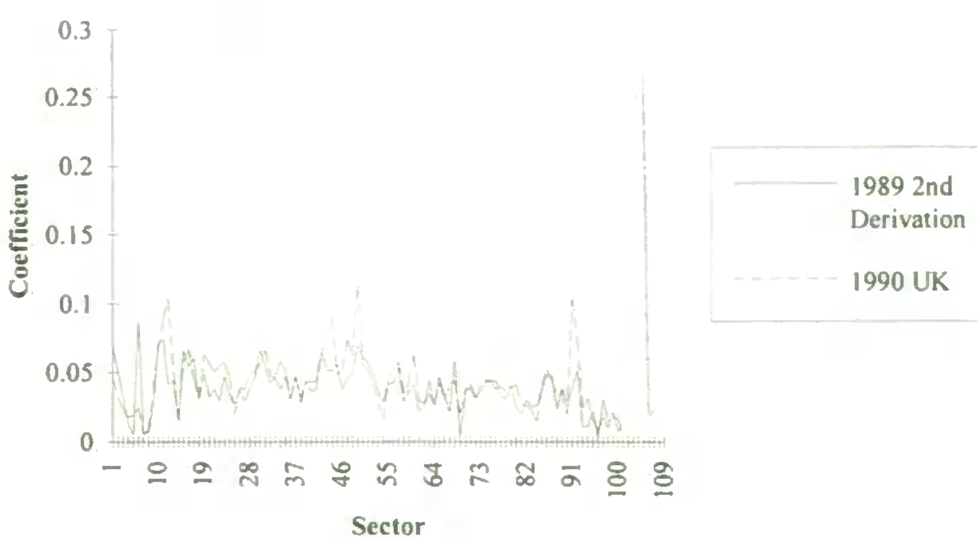


Clearly, this reveals that the initial methodology resulted in a strong upward bias in the estimates which has largely been removed by the alternative estimation procedure. Consequently, one can conclude that the resulting producer price estimates of consumers' expenditure are much more satisfactory.

The remaining margins could now be distributed across the rest of the matrix using the *pro rata* method of apportionment described above. Only 28% of the value of distributors' margins remained to be allocated in this way. The only difference in the method described above of course was that the column of estimated producer price consumers' expenditure was excluded from matrix **D** in the initial step.

As before, in the final matrix **C**, comparison of the estimated combined gross outputs and the figures in Table 2 of the 1989 balance revealed no discrepancies. However, the true test of the methodology seemed to be in the comparison between the estimated distribution sales coefficients for 1989 and their corresponding 1990 values. By comparing Figure 7.5 below and Figure 7.3 above, one can conclude that the coefficients have now been estimated with a relative degree of success. One may of course question why the 1990 coefficients were not simply used in the 1989 model. The reason for this is that one would then have to split each distributor coefficient into its constituent commodity margins (in 1989, this would be around 50 products) in order to subtract the margin from the relevant purchaser price transaction. This would be extremely difficult to implement successfully.

Figure 7.5 Sales Coefficients of the Distributive Sectors: 1989 2nd Estimation, 1990 Actuals



7.523 Comments on the Estimation Process

These transformations were conducted at a relatively early stage of the analysis. Retrospectively, they appear somewhat clumsy, and, in places, slightly naive. Possible points of weakness relate to the *pro rata* allocation of expenditure tax which, on reflection, should have been applied firstly to consumers' expenditure - which is where the bulk of it falls. However, the evidence of Figure 7.5 suggests that such quibbles are probably relatively minor, and the 1989 tables are a fair approximation of producer prices. The transformations serve to highlight the degree of error inherent in the process of assembling input-output data, and thus reaffirm the idea that any subsequent analysis should remain suitably broad.

7.53 Derivation of the Imports Use Matrix for 1989

The only information on imports in 1989 was at the aggregate commodity level. However, a full matrix of imports from overseas was required to complete the analysis. The problem was therefore how to allocate the aggregates across the rows of the 1989 combined use matrix in order to enable the derivation of the 1989 domestic use and imported use matrices. The only feasible approach seemed to be to make use of the 1990 imports use matrix.

The most plausible method of allocation was to assume that the domestic consumption propensities were constant between the two years. The domestic consumption propensity k_{ij} can be derived from the 1990 combined $\mathbf{C}^{\bullet}$ $[c^{\bullet}_{ij}]$ and domestic use matrices $\mathbf{A}^{\bullet}$ $[a^{\bullet}_{ij}]$ as follows:

$$k_{ij} = a^{\bullet}_{ij} / c^{\bullet}_{ij} \qquad [7.9]$$

The k_{ij} can be applied multiplicatively to the corresponding elements of the 1989 combined use matrix to arrive at initial estimates of the 1989 domestic values. The estimated imports use matrix in transactions ${}^{t}\mathbf{S}^{1}$ is then the difference between the combined and domestic transaction use matrices. However, it should be clear that, the row sums of the estimated imports use matrix will almost certainly depart from their published values because each k_{ij} is only an *estimate* of the 1989 propensity. The most obvious method of dealing with such differences - which one would hope would be relatively small - is to distribute them row-proportionally. Let s_i^n be the row sum of the imports use matrix at the nth iteration. Let m_i equal the 'actual' value of imports for i in 1989. Thus it is possible to define a vector of row multipliers, $\mathbf{v}^n$ $[v^n{}_i]$:

$$v_i^n = m_i / s_i^n \qquad [7.10]$$

And this can be applied to the imports use matrix as follows

$${}^{t}\mathbf{S}^{n+1} = \hat{\mathbf{v}}^{n}\ {}^{t}\mathbf{S}^{n} \qquad \text{s.t.} \qquad {}^{t}s^{n+1}{}_{ij} \leq {}^{t}c_{ij} \qquad [7.11]$$

The constraint ensures that the estimated domestic use matrix remains non-negative. The process then iterates until $s_i^n = m_i$.

In implementing this procedure, the first step was to transform the 1990 matrices from basic prices into producer prices. Only three sectors were affected by the basic price valuation: oil processing; alcohol; and tobacco. The ratio of producer price to basic price commodity output in 1990 was calculated. This was then applied across the respective row of the 1990 combined use matrix. The tax row was then recalculated as a residual. The figure for gross industry output at producer prices then replaces the respective basic price valuation. The difference between the two valuations was credited to the relevant element of the tax row.

Secondly it was necessary to aggregate the 123 sector 1990 matrices into a form that was consistent with the 102 1989 definition. Table A7.1 at the end of this chapter shows that it is not possible to translate the 123 sector 1990 UK definition directly onto the 102 sector scheme of 1989. The only real problem that this creates concerns the energy sectors. Table 7.1 below provides a clear illustration of the problem.

Table 7.1 Disaggregation of Energy Sectors, UK 1989-90

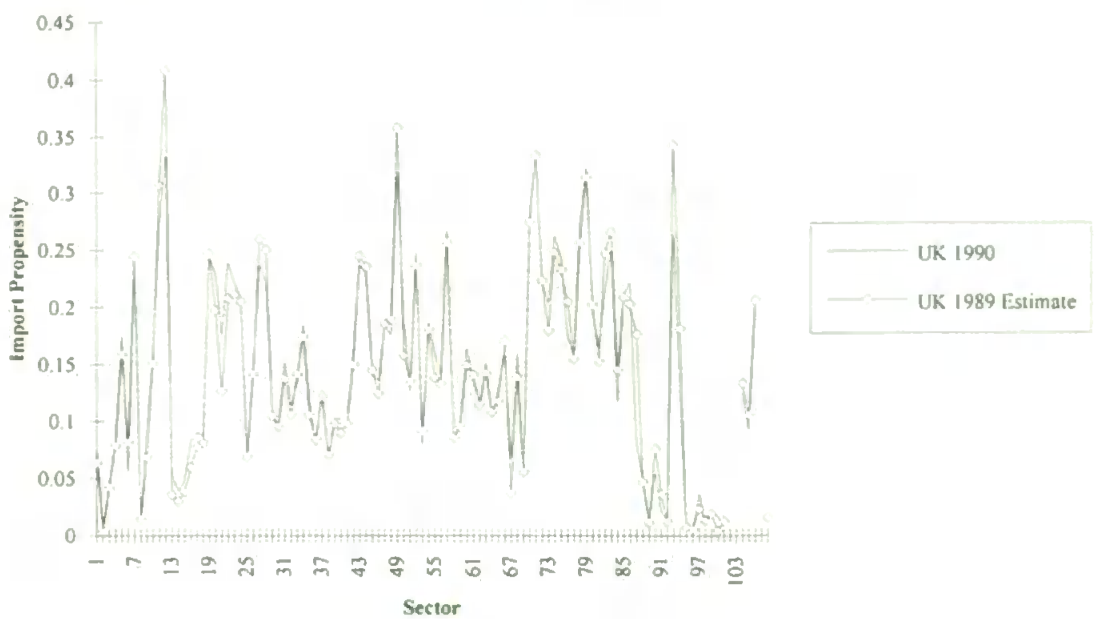
Sector	UK 1989 definition by SIC	UK 1990 definition by SIC
Coal extraction etc.	1113, 1114, 1115, 1200	1113, 1114, 1115
Mineral oil processing etc.	1401, 1402	1200, 1401, 1402, 1520
Electricity production etc.	1520, 1610, 1630	1610, 1630

A consistent SIC definition could be achieved by aggregating coal extraction, mineral oil processing, and electricity production into a single sector - but then the ability to make an individual comparison of these sectors between Scotland and the UK in 1989 would be lost. Clearly, it was impossible to isolate the import propensities of SICs 1200 and 1520 from the 1990 matrix. Nevertheless, since the oil processing sector, SICs 1401 and 1402, accounted for the bulk of import propensity between the three (65% in 1989), and since the general method of import allocation was undoubtedly subject to a significant degree of error, it was felt that the 1990 import propensities for the individual sectors would serve as acceptable approximations for 1989. It was of course impossible to determine the degree of error associated with this approximation. A more satisfactory solution to the other area of inconsistent definition, that of SIC division 9, was found. Since there were no intermediate import purchases made of the commodities relating to the 1989 definition of public administration (sector 102), the import propensities of sectors 116 to 121 in the 1990 matrix could be wholly attributed to the 'other services' sector (101).

The first estimate of the 1989 imports use matrix revealed that the imports of electricity sector had been significantly under-allocated. The initial estimate of £40m was well below the target value of £610m. This could be traced to the large increase in the own-sector domestic purchase that occurred in 1990 through privatisation. The increase in domestic purchases relative to import purchases of course had the effect of reducing the 1990 own-sector import propensity relative to the value in 1989. The £570m discrepancy in 1989 imports was therefore allocated to the electricity sector's own purchase. No other anomalies were detected, and after the third derivation of the imports use matrix, all but two of the values of v were equal to unity. Sectors 27 and 43 showed a 2.6% (£17.7m and £111m respectively) under-allocation of imports. These discrepancies were traced to the row elements of the 1989 combined use matrix which were considerably larger than their 1990 values and seemed most likely to be the source of underestimation (for sector 27 this was the purchase made by sector 33; for sector 43, this was the purchase made by sector 84).

As one would expect given the nature of the allocation methodology, the 1989 estimated import propensities for each industry and final demand category were closely related. Figure 7.6 below illustrates the high degree of correlation between the two (ρ=0.97). Whilst this is evidence to suggest that the allocation methodology has been implemented successfully, it gives no indication as to the accuracy of the estimated 1989 domestic and imported use matrices. The only guarantee the methodology has to offer is that the row sums in each matrix match their published values. Undoubtedly, significant degrees of error will exist in the individual cells of the estimated domestic use matrix for 1989, but since the 'true' value of the transactions are unknown, these errors cannot be assessed. The comments at the end of section 7.523 apply.

Figure 7.6 Industry Import Propensities: 1989 Estimates, 1990 Actuals



7.54 Adjustment for Financial Services

The non-symmetric use matrices include an adjustment column for financial services (net interest income). This was distributed across the elements of the relevant financial services row on a *pro rata* basis. A counterbalancing adjustment was made to the value of gross profits.

7.55 Aggregation

The system of classification used in the 1989 UK balance is closely related to that of the Scottish model. The main difference between the two schemes lies within SIC divisions 8 and 9. Consistency can be achieved by combining sectors 2 to 5 and 103 to 113 in the Scottish model, and sectors 100 and 101 in the UK model into a single definition. This is unfortunate since it surpresses any detail in the definition of SIC division 8 and 9 and creates two heterogeneous blocks of productive activity (some 10% and 12% of UK gross intermediate output in 1989 respectively). Poor definition of financial and public services within the tables undoubtedly imposes some limit upon the significance one is able to attach to the results of any subsequent analysis. In relation to other UK studies of this nature, the level of detail offered to SIC 9 is the same as in the current exercise; however, the current study has the potential to conduct analysis at at a significantly higher level of detail since it includes approximately twice the number of sectors that were available to Harrigan *et al* (1980a) and nearly six times the number of the Smith and Morrison (1974) analysis. The 99 sector definition can be found in Table A7.2 at the end of this chapter.

The research hypotheses require comparison across a range of matrix orders. In addition to the 99 sector definition - the maximum level of detail afforded by the data, three subsequent points of aggregation were chosen: at 40, 21, and 7 sectors. Of course, one can aggregate to a single sector, and this gives five points of observation in total. The sector schemes are not arbitrarily defined:

commodities are grouped according to the nature of their production. Details of the aggregations can be found in Table A7.3 of the appendix.

7.56 Main Variables

Table A7.4 of the appendix lists the main variables used in the analysis.

7.6 Summary and Conclusion

This chapter has dealt with the selection and preparation of data for analysis. The 1989 UK and Scottish data sets have been chosen to form the basis for the investigation. Whilst there was no real choice in the selection of the regional set, the 1989 national tables have been chosen in preference to the 1990 publication. The basis for this decision was that the analysis would be distorted by economic changes that occurred between 1989 and 1990: 1989 UK data was preferred wherever possible. However, the 1989 UK tables required considerable manipulation in order to make them consistent with the Scottish set. In particular, expenditure taxes and distributors' margins and overseas imports had to be accounted for. For this latter adjustment, 1990 UK import propensities were applied. Both transformations produced results that were relatively satisfactory. Whilst the 1989 estimated tables undoubtedly contain significant error, this is an inherent feature of input-output data. The empirical analysis has been designed with these frailties in mind.

APPENDIX

Table A7.1 Input-Output Definitions: UK 1989, 1990, Scotland 1989

UK 1990 Input-Output Sectors		Scotland 1989	UK 1989
1 Agriculture and horticulture	0100	1	1
2 Forestry	0200	2 7 3 (SIC 0200 split)	2
3 Fishing	0300	4 7 5 (SIC 0300 split)	2
4 Coal extraction and manufacture of solid fuels	1113, 1114, 1115	6 (plus SIC 1200)	3 (as Scots)
5 Extraction of mineral oil and natural gas	1300	7	4
6 Coke ovens, mineral oil processing and nuclear fuel production	1200, 1401, 1402, 1520	8 (less SIC 1200 and 1520)	5 (as Scots)
7 Electricity production	1610, 1630	9 (plus 1520)	6 (as Scots)
8 Gas	1620	10	7
9 Water supply	1700	11	8
10 Extraction of metalliferous ores and minerals nes	2100, 2330, 2396	12	9
11 Iron and steel, and steel products	2210, 2220, 2234, 2235	13	10
12 Aluminium and aluminium alloys	2245	14	11
13 Other non-ferrous metals (including precious metals)	2246, 2247	15	12
14 Extraction of stone, clay, sand and gravel	2310	16	13
15 Structural clay products	2410	17	14
16 Cement, lime and plaster	2420	18	15
17 Concrete, stone, asbestos and abrasive products	2436, 2437, 2440, 2450, 2460	19	16
18 Glass	2471, 2478, 2479	20	17
19 Refractory and ceramic goods	2481, 2489	21	18
20 Inorganic chemicals	2511	22	19
21 Organic chemicals	2512	23	20
22 Fertilisers	2513	24	21
23 Synthetic resins and plastic materials, synthetic rubber	2514, 2515	25	22
24 Paints, dyes, pigments, printing ink	2516, 2551, 2552	26	23
25 Specialised chemicals for industry and agriculture	2562, 2563, 2564, 2565, 2567, 2568, 2569	27 (plus SIC 3290)	24
26 Pharmaceutical products	2570	28	25
27 Soap and toilet preparations	2581, 2582	29	26

28 Chemical products nes	2591, 2599	30	27
29 Man-made fibres	2600	30	28
30 Metal castings, forgings, fastenings, springs, etc	3111, 3112, 3120, 3137, 3138	31	29
31 Metal doors, windows, etc	3142	32	30
32 Packaging products of metal	3164	33	31
33 Metal goods nes	3161, 3162, 3163, 3165, 3166, 3167, 3169	34	32
34 Industrial plant and steelwork	3204, 3205	35	33
35 Agricultural machinery and tractors	3211, 3212	36	34
36 Metal-working machine tools	3221	37	35
37 Engineers small tools	3222	38	36
38 Textile machinery, machinery for working other materials	3230, 3275, 3276	39	37
39 Process machinery and contractors	3244, 3245, 3246	40	38
40 Mining, construction and mechanical handling equipment	3251, 3254, 3255	41	39
41 Mechanical power transmission equipment	3261, 3262	42	40
42 Other machinery and mechanical equipment	3281, 3283, 3284, 3285, 3286, 3287, 3288, 3289	43	41
43 Ordnance, small arms and ammunition	3290	27	42
44 Office machinery and computer equipment	3301, 3302	44	43
45 Insulated wires and cables	3410	45	44
46 Basic electrical equipment	3420	46	45
47 Electrical equipment for industry, batteries, etc	3432, 3433, 3434, 3435	47	46
48 Telecommunication etc equipment, electronic capital goods	3441, 3442, 3443	48	47
49 Electronic components and sub-assemblies	3444, 3453	49	48
50 Electronic consumer goods, records and tapes	3452, 3454	50	49
51 Domestic electric appliances	3460	51	50
52 Electric lighting equipment, etc	3470, 3480	52	51
53 Motor vehicles and parts	3510, 3521, 3522, 3523, 3530	53	52
54 Shipbuilding and repairing	3610	54	53
55 Aerospace equipment manufacturing and repairing	3640	55	54
56 Other vehicles	3620, 3633, 3634, 3650	56	55
57 Instrument engineering	3710, 3720, 3731, 3732, 3733, 3740	57	56
58 Oils and fats	4115, 4116	58	57

59 Slaughtering and meat processing	4121, 4122, 4123, 4126	59	58
60 Milk and milk products	4130	60	59
61 Fruit, vegetables and fish processing	4147, 4150	61	60
62 Grain milling and starch	4160, 4180	62	61
63 Bread, biscuits and flour confectionery	4196, 4197	63	62
64 Sugar	4200	64	63
65 Confectionery	4213, 4214	65	64
66 Animal feeding stuffs	4221, 4222	66	65
67 Miscellaneous foods	4239	67	66
68 Alcoholic drink	4240, 4261, 4270	68 & 69 (SIC 4270)	67
69 Soft drinks	4283	70	68
70 Tobacco	4290	71	69
71 Woollen and worsted	4310	72	70
72 Cotton etc spinning and weaving	4321, 4322, 4336, 4340	73	71
73 Hosiery and other knitted goods	43633, 4364	74	72
74 Textile finishing	4370	75	73
75 Carpets and other textile floorcoverings	4384, 4385	76	74
76 Jute etc yarns and fabrics, and miscellaneous textiles	4350, 4395, 4396, 4398, 4399	77	75
77 Leather and leather goods	4410, 4420	78	76
78 Footwear	4510	79	77
79 Clothing and furs	4531, 4532, 4533, 4534, 4535, 4536, 4537, 4538, 4539, 4560	80	78
80 Household and other made-up textiles	4555, 4556, 4557	81	79
81 Timber processing and wood products (not furniture)	4610, 4620, 4630, 4640, 4650, 4663, 4664	82	80
82 Wooden furniture, shop and office fittings	4671, 4672	83	81
83 Pulp, paper and board	4710	84	82
84 Paper and board products	4721, 4722, 4723, 4724, 4725, 4728	85	83
85 Printing and publishing	4751, 4752, 4753, 4754	86	84
86 Rubber products	4811, 4812, 4820	87	85
87 Processing of plastics	4831, 4832, 4833, 4834, 4835, 4836	88	86

88 Jewellery and coins	4910	**89**	**87**
89 Sports goods and toys	4941, 4942	**89**	**87**
90 Other goods	4920, 4930, 4954, 4959	**89**	**87**
91 Construction	5000, 5010, 5020, 5030, 5040	90	88
92 Wholesale distribution	6110, 6120, 6130, 6149, 6150, 6160, 6170, 6180, 6190, 6210, 6220, 6300	**91**	**89**
93 Retail distribution	6410, 6420, 6430, 6450, 6460, 6470, 6480, 6530, 6540, 6560, 6720, 6730	**91**	**89**
94 Distribution & repair of vehicles,filling stations & other goods	6148, 6510, 6520, 6710	**91**	**89**
95 Hotels, catering, public houses, etc	6611, 6612, 6620, 6630, 6640, 6650, 6670	92	90
96 Railways	7100	93	91
97 Road and other inland transport	7210, 7220, 7230, 7260	94	92
98 Sea transport	7400	95	93
99 Air Transport	7500	96	94
100 Transport services	7610, 7630, 7640, 7700	97	95
101 Postal services	7901	98	96
102 Telecommunications	7902	99	97
103 Banking and finance	8140, 8150	**100 & 101**	98
104 Insurance	8200	102	99
105 Auxiliary financial services	8310, 8320	**103 & 104**	**100**
106 Estate agents	8340	105	**100**
107 Legal services	8350	106	**100**
108 Accountancy services	8360	107	**100**
109 Other professional services	8370	108	**100**
110 Advertising	8380	109	**100**
111 Computing services	8394	110	**100**
112 Other business services	8395, 8396	**111 & 112**	**100**
113 Renting of movables	8410, 8420, 8430, 8460, 8480, 8490	113	**100**
114 Owning and dealing in real estate	8500	113	**100**

115 Public administration	9111, 9112, 9120, 9130, 9140, 9150, 9190	114	102
116 Sanitary services	9211, 9212,	114	102
	9230	113	101
117 Education	9310, 9320, 9330,	114	102
	9360	113	101
118 Research and development	9400	113	101
119 Health services	9510, 9520, 9530,	114	102
	9540, 9550, 9560	113	101
120 Recreational and welfare services	9611, 9631, 9660, 9690, 9770	114	102
	9711, 9741, 9760, 9791	113	101
121 Personal services	9811, 9812, 9820, 9890	113	101
122 Domestic services	9900	114	102
123 Ownership of dwellings	n/a	114	102

Table A7.2 Input-Ouput Definitions: 99 Sectors, 1989 UK and Scotland

99 Sector Definition of UK 1989 and Scottish 1989 Tables	Mapped to UK 1989	Mapped to Scots 1989
1 Agriculture and horticulture	1	1
2 Forestry and fishing	2	2, 3, 4, 5
3 Coal, coke, solid fuels, oil processing	3	6
4 Extraction of mineral oil and natural gas	4	7
5 Mineral oil processing	5	8
6 Electricity and nuclear fuel production	6	9
7 Gas	7	10
8 Water supply	8	11
9 Extraction of metalliferous ores and minerals nes	9	12
10 Iron and steel, and steel products	10	13
11 Aluminium and aluminium alloys	11	14
12 Other non-ferrous metals (including precious metals)	12	15
13 Extraction of stone, clay, sand and gravel	13	16
14 Structural clay products	14	17
15 Cement, lime and plaster	15	18

16 Concrete, stone, asbestos and abrasive products	16	19
17 Glass	17	20
18 Refractory and ceramic goods	18	21
19 Inorganic chemicals	19	22
20 Organic chemicals	20	23
21 Fertilisers	21	24
22 Synthetic resins and plastic materials, synthetic rubber	22	25
23 Paints, dyes, pigments, printing ink	23	26
24 Specialised chemicals for industry and agriculture, ordanance	24, 42	27
25 Pharmaceutical products	25	28
26 Soap and toilet preparations	26	29
27 Chemical products nes, man made fibres	27, 28	30
28 Metal castings, forgings, fastenings, springs, etc	29	31
29 Metal doors, windows, etc	30	32
30 Packaging products of metal	31	33
31 Metal goods nes	32	34
32 Industrial plant and steelwork	33	35
33 Agricultural machinery and tractors	34	36
34 Metal-working machine tools	35	37
35 Engineers small tools	36	38
36 Textile machinery, machinery for working other materials	37	39
37 Process machinery and contractors	38	40
38 Mining, construction and mechanical handling equipment	39	41
39 Mechanical power transmission equipment	40	42
40 Other machinery and mechanical equipment	41	43
41 Office machinery and computer equipment	43	44
42 Insulated wires and cables	44	45
43 Basic electrical equipment	45	46
44 Electrical equipment for industry, batteries, etc	46	47
45 Telecommunication etc equipment, electronic capital goods	47	48
46 Electronic components and sub-assemblies	48	49
47 Electronic consumer goods, records and tapes	49	50
48 Domestic electric appliances	50	51

49 Electric lighting equipment, etc	51	52
50 Motor vehicles and parts	52	53
51 Shipbuilding and repairing	53	54
52 Aerospace equipment manufacturing and repairing	54	55
53 Other vehicles	55	56
54 Instrument engineering	56	57
55 Oils and fats	57	58
56 Slaughtering and meat processing	58	59
57 Milk and milk products	59	60
58 Fruit, vegetables and fish processing	60	61
59 Grain milling and starch	61	62
60 Bread, biscuits and flour confectionery	62	63
61 Sugar	63	64
62 Confectionery	64	65
63 Animal feeding stuffs	65	66
64 Miscellaneous foods	66	67
65 Alcoholic drink	67	68, 69
66 Soft drinks	68	70
67 Tobacco	69	71
68 Woollen and worsted	70	72
69 Cotton etc spinning and weaving	71	73
70 Hosiery and other knitted goods	72	74
71 Textile finishing	73	75
72 Carpets and other textile floorcoverings	74	76
73 Jute etc yarns and fabrics, and miscellaneous textiles	75	77
74 Leather and leather goods	76	78
75 Footwear	77	79
76 Clothing and furs	78	80
77 Household and other made-up textiles	79	81
78 Timber processing and wood products (not furniture)	80	82
79 Wooden furniture, shop and office fittings	81	83
80 Pulp, paper and board	82	84
81 Paper and board products	83	85

82 Printing and publishing	84	86
83 Rubber products	85	87
84 Processing of plastics	86	88
85 Other manufacturing	87	89
86 Construction	88	90
87 Wholesale & retail distribution, including repair of vehicles	89	91
88 Hotels, catering, public houses, etc	90	92
89 Railways	91	93
90 Road and other inland transport	92	94
91 Sea transport	93	95
92 Air Transport	94	96
93 Transport services	95	97
94 Postal services	96	98
95 Telecommunications	97	99
96 Banking and finance	98	100, 101
97 Insurance	99	102
98 Other services	100, 101	103 to 113
99 Public administration	102	114

Table A7.3 Input-Output Definitions: 40, 21 and 7 Sectors

40 Sector	Description	99 Sector	21 Sector	7 Sector
1	Agr, For, Fish	1 & 2	1	1
2	Coal Extrct	3	2	2
3	Oil, Gas Extrct	4	2	2
4	Mineral Oil	5	2	2
5	Electricity	6	3	2
6	Gas	7	3	2
7	Water	8	21	7
8	Metal Extrct	9	4	3
9	Iron & Steel	10 - 12	5	3
10	Stone Extrct	13	6	4
11	Concrete	14 - 16	7	4
12	Glass	17 & 18	7	4
13	Inorg Chem	19	8	4
14	Organic Chem	20	8	4
15	Other Chemical	21 - 24	8	4
16	Chem Products	25 - 27	8	4
17	Metal Goods	28 - 31	9	3
18	Machinery	32 - 41	10	5
19	Electrical	42 - 49	11	5
20	Motor etc	50 & 53	12	5
21	Ships etc	51 & 52	13	5
22	Instrument Eng.	54	11	5
23	Food Process	55 - 60	14	4
24	Misc Food	61 - 64	14	4
25	Drink & Tobacco	65 - 67	15	4
27	Clothes	74 - 76	16	4
26	Textiles	68 - 73 & 77	17	4
28	Wood	78 & 79	7	4
29	Printing	80 - 82	18	4
31	Construction	86	21	7
30	Other Manuf	83 - 85	7	4
32	Distrbtn & Hotels	87 & 88	20 (less hotels)	6
33	Other Transp	89 & 91 - 93	19	6
34	Road Transp	90	19	6
35	Post	94	21	7
36	Telecom	95	21	7
37	Banks	96	21	7
38	Insurance	97	21	7
39	Other Service	98	21 (plus hotels)	7
40	Public Admin	99	21	7

Table A7.4 List of Main Variables

Key	
r_i	Scottish Regional Intermediate Row Purchase (£m, 1989)
k_i	Scottish Propensity to Purchase Domestically from Total Intermediate Row Purchase
u^*_i	Nonsurvey Total Intermediate Row Purchase (£m, 1989)
$r_i/u^*_i k_i$	Dependent Variable for Intermediate Row Sums
r_j	Scottish Regional Intermediate Column Purchase (£m, 1989)
l_j	Scottish Propensity to Purchase Regionally from Domestic Supply (Columns)
q_i	Location Quotient
$c(r_i)$	Scottish Consumers' Expenditure on Local Goods (£m, 1989)
c_i	Scottish Consumer's Propensity to Buy Local Goods from Domestic Production

Sector	r_i	k_i	u^*_i	$r_i/u^*_i k_i$	r_j	l_j	q_i	$c(r_i)$	c_i
1	1085.11	0.86	1327.36	0.95	565.25	0.80	1.05	58.02	0.36
2	226.77	0.96	93.48	2.52	229.29	0.86	6.27	61.63	0.99
3	178.46	0.66	317.11	0.85	48.47	0.70	0.75	25.57	0.40
4	362.31	0.45	1083.78	0.74	102.50	0.21	1.38	0.00	n/a
5	741.04	0.78	1000.85	0.95	998.56	0.90	1.64	226.28	0.80
6	581.53	0.99	735.35	0.80	747.89	0.74	1.07	587.72	1.00
7	144.92	1.00	175.43	0.83	3.17	0.02	0.92	327.30	1.00
8	77.04	1.00	52.79	1.46	43.87	0.79	1.00	60.23	1.00
9	8.24	0.13	81.91	0.77	6.17	0.78	0.70	0.00	n/a
10	353.23	0.92	725.19	0.53	298.80	0.61	1.21	0.00	n/a
11	60.38	0.87	185.81	0.37	95.02	0.67	1.37	0.00	n/a
12	38.73	0.73	196.74	0.27	3.29	0.15	0.33	1.65	1.00
13	35.21	0.89	28.61	1.38	22.99	0.73	1.41	0.00	n/a
14	37.93	0.97	46.79	0.84	10.82	0.57	0.59	0.32	0.21
15	49.98	0.88	67.45	0.84	36.25	0.71	0.84	1.80	0.63

16	215.63	0.95	292.40	0.78	193.44	0.68	0.96	7.48	0.56
17	123.23	0.91	165.15	0.82	64.11	0.72	1.30	1.41	0.08
18	36.84	0.59	70.28	0.88	33.23	0.69	0.88	20.69	0.30
19	4.86	0.40	137.54	0.09	20.66	0.58	0.49	0.00	n/a
20	250.00	0.74	268.07	1.27	187.18	0.48	1.71	0.00	n/a
21	64.98	0.68	54.79	1.74	32.11	0.73	1.30	1.99	0.53
22	33.60	0.86	257.18	0.15	75.16	0.66	0.88	0.45	0.03
23	14.22	0.78	116.87	0.16	75.33	0.65	0.78	1.21	0.08
24	43.96	0.83	240.23	0.22	94.45	0.76	0.53	0.69	0.05
25	19.25	0.86	80.65	0.28	84.56	0.41	0.99	2.63	0.03
26	3.08	0.96	27.10	0.12	6.73	0.55	0.19	7.92	0.05
27	3.93	0.62	118.80	0.05	27.99	0.45	0.65	0.65	0.10
28	146.31	0.87	334.40	0.50	80.70	0.61	0.68	0.07	0.00
29	22.22	0.88	34.30	0.73	4.31	0.18	0.71	0.28	0.03
30	64.46	0.72	193.50	0.46	22.92	0.44	0.73	0.00	0.00
31	63.01	0.86	159.96	0.46	96.01	0.63	0.62	0.38	0.01
32	41.13	0.75	166.83	0.33	213.32	0.68	2.15	0.00	n/a
33	2.54	0.67	31.06	0.12	4.50	0.43	0.25	0.00	n/a
34	6.47	0.77	16.11	0.52	7.33	0.42	0.40	0.00	n/a
35	14.20	0.91	44.24	0.35	15.71	0.57	0.81	0.00	0.00
36	8.82	0.90	36.08	0.27	9.87	0.48	0.27	0.00	0.00
37	26.98	0.81	79.26	0.42	21.52	0.56	0.49	0.00	0.00
38	44.59	0.89	132.73	0.38	169.38	0.80	1.27	0.00	0.00
39	29.19	0.76	58.67	0.65	4.14	0.16	0.93	0.00	0.00
40	48.57	0.81	437.29	0.14	204.69	0.60	0.91	0.00	0.00
41	147.38	0.49	566.15	0.54	379.73	0.21	6.22	28.96	0.75
42	28.44	0.60	87.96	0.53	8.69	0.43	0.43	0.01	0.00
43	34.17	0.90	190.11	0.20	62.39	0.37	1.06	5.16	1.00
44	4.27	0.82	53.33	0.10	26.78	0.57	0.56	2.07	0.14
45	79.24	0.63	127.70	0.99	61.32	0.29	0.95	7.32	0.12
46	161.19	0.64	834.50	0.30	49.51	0.18	2.31	3.33	0.66
47	18.05	0.89	18.30	1.11	76.32	0.68	1.40	30.55	0.40
48	8.52	0.91	16.95	0.55	17.97	0.39	0.53	30.54	0.24
49	16.03	0.77	34.85	0.60	33.26	0.62	0.79	0.29	0.02
50	71.16	0.87	275.97	0.30	163.86	0.65	0.22	6.59	0.13
51	16.87	0.74	65.90	0.34	72.51	0.42	3.33	6.09	0.79
52	36.18	0.67	196.25	0.28	190.90	0.50	0.86	0.00	0.00
53	7.32	0.83	33.97	0.26	5.94	0.71	0.32	3.35	0.84
54	29.48	0.43	42.73	1.61	75.39	0.87	0.92	20.29	0.59
55	31.01	0.94	95.96	0.35	41.13	0.63	1.02	15.12	0.24
56	148.69	0.85	241.89	0.72	569.14	0.92	1.46	187.53	0.48
57	68.58	0.87	151.34	0.52	282.68	0.91	0.97	138.87	0.38
58	66.31	0.66	154.23	0.65	249.38	0.86	2.29	101.45	0.58

59	79.17	0.83	157.20	0.61	79.52	0.73	1.28	12.99	0.14
60	37.81	0.96	139.17	0.28	186.78	0.63	1.51	112.38	0.46
61	44.07	0.79	112.33	0.50	14.40	0.53	0.84	8.66	0.20
62	9.21	0.74	49.67	0.25	29.02	0.77	0.31	35.11	0.52
63	146.07	0.89	202.95	0.81	148.41	0.89	1.05	18.87	0.12
64	26.91	0.73	162.02	0.23	118.38	0.81	0.57	65.20	0.37
65	258.96	0.82	305.62	1.04	1281.73	0.74	3.67	240.02	0.70
66	67.70	0.97	56.34	1.24	60.01	0.67	0.93	56.22	0.42
67	11.77	0.95	20.26	0.61	203.14	0.67	0.89	90.86	0.28
68	115.54	0.78	157.88	0.94	83.55	0.66	2.14	5.38	0.52
69	7.36	0.68	247.55	0.04	30.15	0.70	1.12	2.33	0.18
70	27.80	0.87	29.38	1.09	78.47	0.59	1.96	28.52	0.24
71	18.05	1.00	47.74	0.38	12.53	0.43	0.79	1.44	1.00
72	5.82	0.57	34.72	0.30	25.79	0.45	1.03	6.79	0.19
73	16.59	0.55	92.68	0.33	55.99	0.64	2.18	0.95	0.08
74	9.75	0.92	51.18	0.21	48.72	0.76	1.58	7.31	0.51
75	2.40	0.38	7.60	0.84	3.78	0.30	0.38	0.91	0.01
76	94.43	0.95	23.23	4.28	200.45	0.73	1.56	33.40	0.10
77	27.64	0.88	24.62	1.27	29.11	0.62	0.91	1.62	0.13
78	288.83	0.75	433.22	0.89	180.16	0.63	1.69	34.69	0.88
79	22.50	0.79	70.68	0.40	59.07	0.68	0.47	21.14	0.11
80	95.22	0.49	498.55	0.39	163.38	0.62	2.83	4.52	0.58
81	282.28	0.93	431.35	0.71	99.45	0.51	0.90	27.45	0.37
82	233.80	0.71	451.79	0.72	209.45	0.66	0.77	94.93	0.26
83	31.84	0.33	150.28	0.64	82.51	0.63	1.23	9.03	0.18
84	198.82	0.95	599.60	0.35	152.84	0.48	0.80	0.76	0.02
85	25.81	0.72	62.81	0.57	63.09	0.65	0.72	31.49	0.22
86	988.03	1.00	994.42	0.99	1407.65	0.67	0.62	86.02	1.00
87	1197.40	1.00	1693.89	0.71	1243.16	0.74	0.86	3420.87	0.93
88	95.44	0.96	158.38	0.63	528.29	0.87	1.34	1359.85	0.85
89	99.39	0.98	122.20	0.83	53.18	0.82	0.94	121.01	0.89
90	677.24	0.97	777.22	0.90	346.45	0.89	1.15	380.21	0.95
91	51.13	0.36	226.70	0.62	199.70	0.82	1.29	29.85	0.98
92	246.94	0.95	141.07	1.85	104.00	0.32	1.53	39.36	0.17
93	355.54	0.91	616.70	0.64	43.03	0.43	0.67	18.04	0.97
94	63.17	0.92	208.25	0.33	8.60	0.20	0.71	56.26	0.94
95	251.11	0.99	385.84	0.66	18.61	0.42	0.69	172.29	0.98
96	798.55	0.99	2301.17	0.35	290.25	0.74	0.46	166.80	0.44
97	41.71	0.93	393.66	0.11	212.55	0.46	0.66	46.84	0.31
98	2689.69	0.99	3517.34	0.77	543.40	0.67	0.70	885.09	0.58
99	n/a	n/a	n/a	n/a	0.00	n/a	1.20	2028.02	1.00

CHAPTER 8

EMPIRICAL ANALYSIS :

DEVELOPING A TECHNIQUE FOR

ESTIMATING REGIONAL INPUT-OUTPUT TABLES

8.0 Developing a Technique for Estimating Regional Input-Output Tables

8.1 Introduction

The hypotheses set out in Chapter 5 are now tested empirically. The first part of the chapter therefore seeks to investigate the nature, significance and relative importance of nonsurvey estimation error components. On the basis of this analysis, the remainder of the chapter develops and tests the new technique for use within the hybrid estimation process. The data set used for the analysis is that identified in the previous chapter. The tools for the analysis are those presented in Chapter 6.

8.2 The Nature, Significance and Relative Importance of Nonsurvey Estimation Error Components

This section presents an empirical investigation of hypotheses one to eight.

8.21 The Nature of Nonsurvey Estimation Error Components

8.211 The Nature of Total Use Misspecification

In Chapter 5 it was stated that:

H_1: the error associated with total use misspecification is parabolic with respect to matrix order.

In order to test this hypothesis, nonsurvey and observed combined use matrices were compared over a range of matrix orders.

The nonsurvey combined use matrix in transactions was derived by 'grossing up' the national combined coefficients use matrix with the vector of Scottish industry outputs:

$$ {}^{r}\mathbf{U}^{*} = {}^{c}\mathbf{C}\hat{\mathbf{X}} \qquad [8.1]$$

where

${}^{r}\mathbf{U}^{*}$ $[u_{ij}^{*}]$

Nonsurvey matrices were generated at orders 1, 7, 21, 40 and 99 - as specified in the previous chapter. The Standardised Total Percentage Error (STPE, e)

between these matrices and their observed counterparts, $'\mathbf{U}$ $[u_{ij}]$ was calculated, *i.e.*

$$e = \frac{\sum_{i,j} \left| u_{ij}^{*} - u_{ij} \right|}{\sum_{i,j} u_{ij}} \times 100 \qquad [8.2]$$

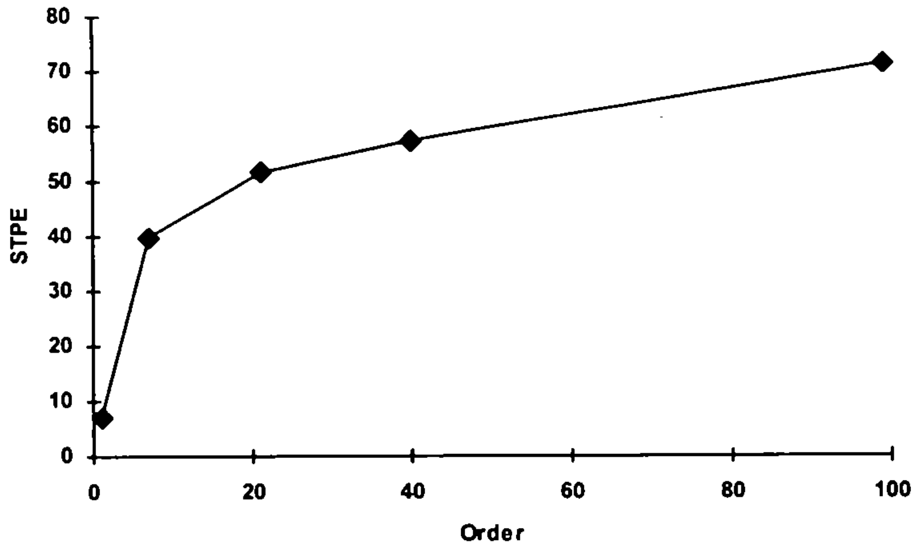
This effect of this action is largely to isolate the error that is due to differences in regional and national production functions. As mentioned in previous chapters, stochastic errors present in either of the subject matrices are not excluded from [8.2] because they cannot be observed. The measure of error provided by [8.2] is therefore not entirely 'pure'. This point is applicable to all subsequent matrix comparisons.

The results of the exercise are displayed in Table 8.1 and Figure 8.1 below.

Table 8.1 STPE Between Combined Use Matrices

Order	STPE
1	7.0
7	40.0
21	51.6
40	57.3
99	71.2

Figure 8.1 STPE Between Combined Use Matrices



The results show that, at low orders, the error associated with the misspecification of regional production functions is relatively low. This error rises with matrix order. This fits in with *a priori* expectation: when the production functions are at their broadest definition, regional-national differences in input use are disguised; as homogeneity improves, the functions that emerge to describe regional production in the regional model are 'averaged out' in the national model, and hence the misspecification rises. Whilst there is some evidence of the error function becoming shallower at higher orders, there is no sign of the hypothesised 'peak', and subsequent fall, in error which would mark the separation of regional functions within the national model. However, given that the error function behaves consistently with expectations over the range of observations, the general hypothesis concerning the effect of heterogeneity seems plausible. In other words, it would seem that, at the given 99 sector definition of the national input-output table, the production functions of Scottish industries remain, to a large extent, merged with those of other regions. However, there seems sufficient empirical evidence and *a priori* reason to support the belief that these Scottish functions would become identifiable at some higher - and perhaps much higher - level of definition. The curve depicted in Figure 8.1, whilst clearly asymptotic, may therefore, at some higher level of disaggregation, become parabolic. Hypothesis one is therefore maintained. The significance of the observed error is discussed below.

8.212 The Nature of Trade Misspecification: Overseas Import Propensities

In Chapter 5 it was stated:

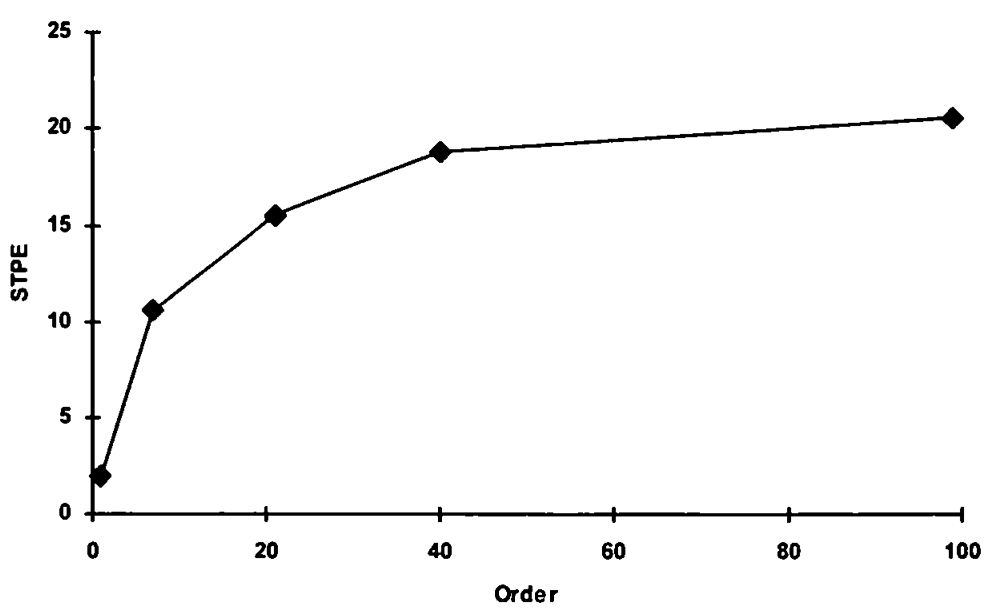
H_3: the error associated with the use of national overseas import propensities is parabolic with respect to matrix order.

In order to test this hypothesis, the error associated with the use of national overseas import propensities had to be isolated from other sources of misspecification. Applying observed Scottish overseas import propensities to the Scottish combined use matrix clearly gave Scotland's domestic use matrix, $'\mathbf{A}$ $[a_{ij}]$. Applying the UK model's overseas import propensities to the Scottish combined use table gave $'\mathbf{A}^{*}$ $[a^{*}_{ij}]$, the Scottish domestic use matrix, had its industries imported from overseas at average UK rates. Differences between the matrices $'\mathbf{A}$ and $'\mathbf{A}^{*}$ could only result from differences in observed and nonsurvey overseas import propensities. Hence the STPE calculated from the elements a_{ij} and a^{*}_{ij}, in the manner of equation [8.2], gave a measure of the error due to the misspecification of overseas import propensities. The process was completed for the five given orders of disaggregation. The results are shown in Table 8.2 and Figure 8.2 below.

Table 8.2 STPE Due to Differences in Observed and Nonsurvey Overseas Import Propensities

Order	STPE
1	2.0
7	10.6
21	15.6
40	18.8
99	20.6

Figure 8.2 STPE Due to Differences in Observed and Nonsurvey Overseas Import Propensities



The results suggest that hypothesis three has some validity. The average dependence on overseas imports is not markedly different across regions: hence at low orders, misspecification error is relatively low. As order rises, the regional model begins to display its characteristics, and these include a particular pattern of dependence on overseas supply. However, the national model is not sufficiently defined to reflect these characteristics and misspecification error rises. The predicted 'downturn' in the error function, which comes as a result of regional-specific representation within the national model, has not been reached at the maximum level of disaggregation. As before however, the evidence and reasoning is strong enough to support the view that, at some further point of disaggregation, the specification error may begin to fall. Hypothesis three is therefore upheld. The significance of the observed error is considered below.

8.213 The Nature of Trade Misspecification: Regional Import Propensities

In Chapter 5 it was stated:

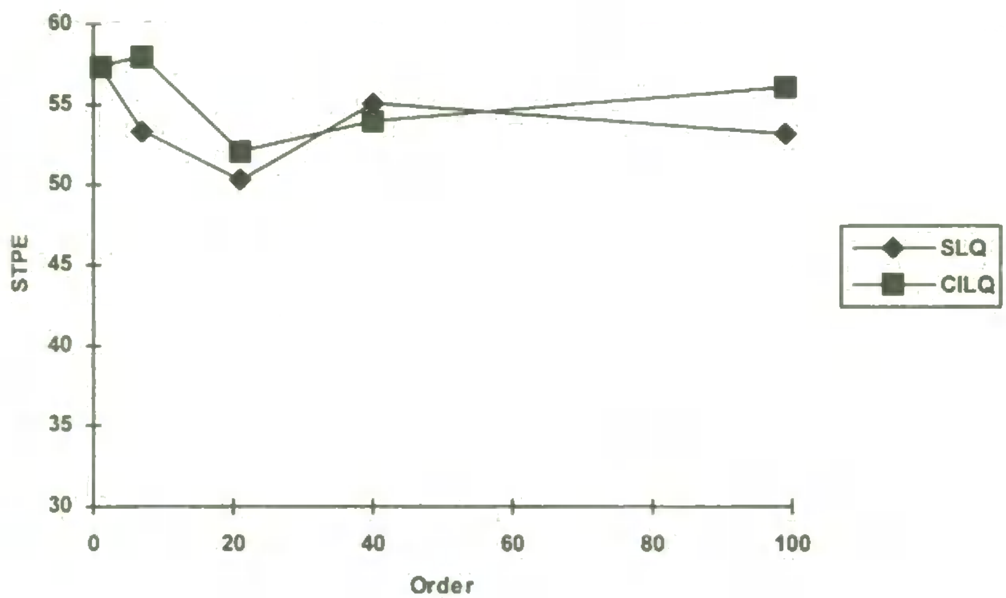
H_4: the error associated with regional trade misspecification is largely unrelated to matrix order.

Once again, the test of this hypothesis required the isolation of the error that was due to the misspecification of the regional trade function. Two specifications of the regional trade function were selected for the test: the SLQ and the CILQ. The justification for this choice was that these methods are the two most commonly applied. The Scottish domestic use matrix, $^{r}\mathbf{A}$, was the starting point for the error isolation exercise. The Scottish regional use matrix, $^{r}\mathbf{R}$ [r_{ij}], was given by the applying the observed regional import propensities to the Scottish domestic use matrix. Applying the location quotient formulae to the observed domestic use matrix $^{r}\mathbf{A}$ gave a nonsurvey estimate of the regional use matrix, $^{r}\mathbf{R}^{*}$ [r^{*}_{ij}]. Any difference between the matrices $^{r}\mathbf{R}$ and $^{r}\mathbf{R}^{*}$ would have to be due to differences between quotient estimated and observed regional import propensities. STPE were therefore calculated between the elements r_{ij} and r^{*}_{ij} at each of the given orders of aggregation. The results are displayed in Table 8.3 and Figure 8.3 below.

Table 8.3 STPE Due to Differences in Observed and Nonsurvey Regional Import Propensities

Order	STPE for SLQ	STPE for CILQ
1	57.4	57.4
7	53.4	58.0
21	50.3	52.1
40	55.1	54.0
99	53.2	56.1

Figure 8.3 STPE Due to Differences in Observed and Nonsurvey Regional Import Propensities



It was argued in Chapter 5 that, if a quotient is a correct specification of the regional import propensity, the improvement in the information used for its construction afforded by an improvement in homogeneity will result in a reduction in trade estimation error. Figure 8.3 illustrates that the measured level of error is more or less constant across the range of aggregations for both SLQ and CILQ, and hence hypothesis four is maintained. The suggestion is that both methodologies suffer some form of misspecification - the relative 'significance' of this will be considered in the next section. The similarity between the degrees of measured error is consistent with the findings of other studies, and may be interpreted as a further indication of some general deficiency. One should also note that the fall in error that is predicted as homogeneity in the national model is approached (regional imports become part of non-competitive supply) has not been reached at the maximum level of disaggregation.

8.22 The Significance and Relative Importance of Nonsurvey Estimation Error Components

In Chapter 5, the following hypotheses were stated:

H_6: nonsurvey assumptions over total use relationships and regional import propensities are misspecified to a significant degree.

H_7: the assumption of national overseas import propensities holds at the regional level and therefore constitutes the least significant source of misspecification.

Firstly, consider the absolute significance of each error component. In Chapter 6 the STPE statistic was calibrated to given levels of random error. The value of STPE, e was found to be approximately equivalent to the introduction of a stochastic, uniformly distributed error of +/-2e percent into the 'true' matrix. Following Knudsen and Fotheringham's (1986) broad decision rules, one should always be looking to reject a matrix of estimated flows as a suitable proxy for the set of observed flows if the level of error exceeds 50%. One should always accept the estimated matrix as a suitable proxy at error levels below 10%. Hence, the broad ranges of automatic acceptance and rejection for STPE are 5 and 25 respectively.

Referring back to Table 8.1 therefore, only from the broadest possible perspective could one consider accepting the hypothesis that there are no differences between Scotland and the average region's total demand for commodity inputs. As soon as the analysis extends into what are, at the 7 sector level, still only very broad sector definitions, one can find evidence of 'significant' differences between regional-national production functions. The STPE of 39.9 suggests that the error level between regional-national production functions around 80%. At the highest order of observation, this error is rises to around 140% - nearly three times above the boundary at which one would automatically reject the hypothesis of no differences. Similarly, the error associated with the use of location quotients as regional trade estimators is significant at all levels of disaggregation. The recorded values of STPE in Table 8.3 are equivalent to error levels in excess of 100%. Thus hypothesis six is maintained.

The error associated with the use of national overseas import propensities however appears to be less significant, as anticipated. There is evidence to suggest that the dependence on overseas supply by Scottish industries as a whole is not different from the average region's - the associated random error at the one sector level is about 4% (Table 8.2). However, this error rises at higher levels of disaggregation into an area where its significance cannot be determined. The implied 40% error, whilst on the 'high side' of the upper boundary, is not an automatic indication of significant differences. However, it cannot be concluded that the use of the national model's overseas does not constitute a significant source of misspecification. Hypothesis seven is therefore only partially maintained.

Tables 8.1, 8.2, and 8.3 describe the error in generating the combined, domestic and regional use tables respectively. Whilst these errors are comparable, it is perhaps more appropriate to compare the effect of misspecification in each component on the regional use matrix alone. In other words, in measuring the error due to the misspecification of the total use relationships, a regional use matrix, $'\mathbf{R}^{\bullet}$ $[r^{\bullet}_{ij}]$, is generated from the nonsurvey combined use matrix [8.1] and the observed overseas and regional import propensities. The error is measured between $'\mathbf{R}^{\bullet}$ $[r^{\bullet}_{ij}]$ and $'\mathbf{R}$ $[r_{ij}]$, rather than between combined use matrices. In isolating the misspecification due to nonsurvey overseas import propensities, the matrix $'\mathbf{R}^{\bullet}$ $[r^{\bullet}_{ij}]$ is generated from the observed combined use matrix, the nonsurvey import propensities, and the observed regional import propensities. The measure of error due to the misspecification of regional trade of course is exactly the same as that reported in Table 8.3. The advantage of this alternative is that the three errors are calculated with respect to the same observed matrix. As Table 8.4 illustrates, the effect is to give greater weight to errors in the combined use matrix, and less to the overseas propensities - although the changes are quite slight. Figure 8.4 brings together the four error functions.

Table 8.4 STPE in the Observed Regional Use Matrix

Order	STPE (Total Use)	STPE (Overseas)
1	7.0	2.0
7	44.5	8.8
21	56.2	11.1
40	63.5	14.4
99	81.5	15.3

Figure 8.4 The Relative Importance of Error Components

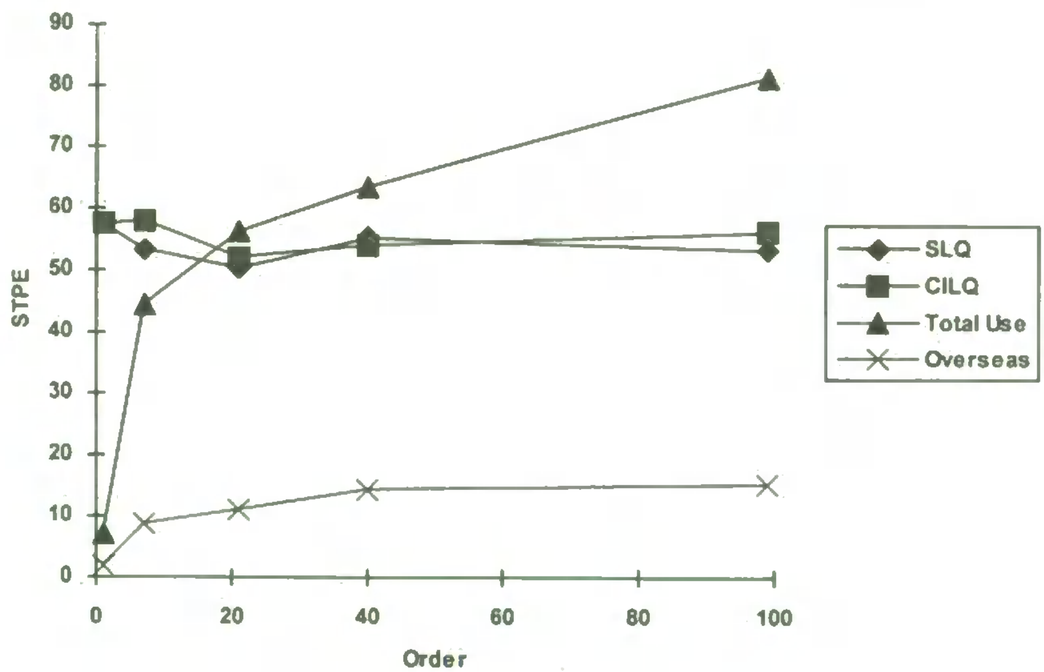


Figure 8.4 clearly illustrates that, for the Scottish model, at anything more than the very broadest levels of detail, the correct specification of the total use matrix is relatively more important than the specification of the regional trade function, in terms of minimising errors in the regional transactions table. This important piece of evidence confirms the view that the analysis of Park *et al.* (1981) is flawed. Therefore, it must be concluded that studies such as Stevens *et al.* (1983) and Flegg *et al.* (1995) are unjustified in their assumption that differences in regional-national production functions are, in general, relatively

unimportant. But is this there any justification across the range of regional sizes?

In Chapter 5, it was stated:

H_2: the error associated with total use misspecification is negatively related to regional size.

H_5: the error associated with regional trade misspecification is negatively related to regional size.

H_8: regional trade specification is relatively more important than total use specification for smaller regions.

In short, it is believed that whilst both total use and regional trade misspecification rise as regional size diminishes, the latter rises quicker than the former.

Unfortunately, the test of these hypotheses is inhibited by lack of suitable UK regional data. In fact, there is only one other regional study which is able to provide any insight, and this is limited. The study in question is that of Peterborough, 1968, by Smith and Morrison (1974). The survey and UK models are reproduced within the appendices of their publication at a consistently defined order of 19 sectors. Thus it is possible to calculate some values for the STPE. A number of points should however be stressed. Firstly, one should of course be wary of attaching any real significance to the relative error values between two studies whose relevant model attributes are not consistent (*i.e.* definition of sectors, time period). The values must be interpreted as no more than crude indicator. Secondly, the presentation of the data is such that only differences between coefficients can be calculated. In the current study weighted coefficients (*i.e.* transactions) are the basis for the computation of error. Thirdly, it is not possible to compare combined use matrices - only the **A** matrices are presented. Fourthly, the comparison can only be made over 19 sectors: the definitions cannot be expanded beyond their published level, but aggregation is also impossible in the absence of transaction value matrices. Despite these severe limitations, the comparison serves some useful purpose.

The only other feasible source of comparison between regions of different 'size' is obtained through subtraction. That is to say that, given the UK model's combined use matrix, netting from it the Scottish counterpart yields a region which is bigger than Scotland. Unfortunately, this comparison cannot be extended to examine the larger region's import propensities because Scottish export destinations are not available.

The levels of error in the coefficients of the Peterborough technology matrix are around 200% (STPE approximately 100). This compares with a value for Scotland at 21 sectors of around 120% (STPE between **A** matrices approximately 60). The error between the combined use matrices of the UK and induced region at 21 sectors is about 30% (STPE equals 13). The levels of error associated with the use of location quotients are something around 600%

(STPE 300) for Peterborough, which compares sharply to the 100% measured in the Scottish model.

This latter result is hardly surprising, given that the Peterborough model relates to an area which is about as small as is practically feasible for an input-output matrix. Hypothesis five therefore seems to find some justification - location quotients fail to account properly for regional size.

Hypothesis two however, which relates to differences in 'technology' lies within a relatively grey area. The apparent rise in error between the three regional models cannot be considered conclusive, simply because the experiment's data and methodology are both very 'dirty'. Nevertheless, the result is in line with *a priori* expectation, and therefore hypothesis two is given weak support.

It would therefore seem reasonable to conclude that hypothesis eight is probably correct in its statement. Of course nothing is known about behaviour of the error functions for regions between these relative extremes of size. It is feasible for example that the sharp rise in import specification error occurs only with economies as small and as open as Peterborough. Whilst the justification for 'trade-only' approaches to the development of nonsurvey techniques is marginally improved by this analysis, it seems fairly clear that the correct specification of the regional production functions should remain a priority for all but the very largest of regional models.

8.23 Summary and Conclusion

The main findings of this stage of the analysis are as follows. For Scotland, both trade and the total use functions are misspecified to a significant degree at the level of detail that available from the national model. Whilst the use of national overseas import propensities may or may not constitute a significant source of misspecification, *relatively* speaking it contributes far less to nonsurvey error, and is therefore of lesser concern in the estimation process. Evidence across the range of regional sizes is patchy. Nevertheless, there seems some evidence to support the view that the specification of the regional import function take on increasing importance as regional size diminishes. However, the specification of regional total use functions remains a necessary objective. Any attempt to develop nonsurvey procedures should therefore attempt to account for *both* total regional use and regional trade. The next stage of the analysis prepares the way for the development of such a procedure.

8.3 Assessing the Potential for Reducing Nonsurvey Estimation Error in the Interindustry Matrix

The previous section measured the level of error associated with the use of standard nonsurvey procedures. However, given that any 'improved' nonsurvey methodology cannot extend beyond a broad process of specification and adjustment, how far can the levels of error be reduced? This section investigates this question. The analysis concerns the interindustry transactions matrix - households are dealt with later.

The potential for reducing nonsurvey specification error is assessed by expressing the maximum and minimum limits of error - the latter being derived from the application of the specific RAS algorithm defined in section 3.54[1]. There are two particular ways in which this can be examined. The first is to consider the potential for reducing the error associated with each specification component. The second is to take a more general view and assess the overall potential for reducing nonsurvey error.

Consider the first of these. In terms of total use, the maximum misspecification of the observed combined use matrix, $'\mathbf{U}$ $[u_{ij}]$, is taken as the standard nonsurvey equivalent, $'\mathbf{U}^{*}$ $[u^{*}_{ij}]$. The maximum STPE is therefore that reported in Table 8.1: 71.2. It is assumed that the most feasible 'target' for nonsurvey estimation is the RAS data set - the intermediate row and column sums of $'\mathbf{U}$. If this can be achieved, the 'best' nonsurvey estimate of $'\mathbf{U}$ is the matrix $'\mathbf{U}^{**}$:

$$'\mathbf{U}^{**} = \hat{\mathbf{r}}\,'\mathbf{U}^{*}\hat{\mathbf{s}} \qquad [8.3]$$

where

r, s are RAS row/ column multipliers, whose targets are derived from $'\mathbf{U}$.

The minimum nonsurvey error in specifying the combined use matrix is therefore the difference between $'\mathbf{U}$ and $'\mathbf{U}^{**}$. At the 99 sector level, the STPE was measured at 59.1.

[1] There may of course an infinite number of RAS algorithm solutions, but for a *given* algorithm with *given* row/column parameters, the solution is unique.

Given that the error associated with the use of national overseas import propensities was relatively low, interest in the extent to which this error could be reduced was, likewise, relatively low. The maximum misspecification error was therefore that associated with the use of unadjusted national propensities - 20.6, as reported in Table 8.2; the minimum error was taken simply as zero.

The maximum misspecification of the regional import propensities was taken as the STPE between the Scottish domestic use matrix, $'\mathbf{A}$ and the Scottish regional use matrix $'\mathbf{R}$. In other words, the maximum 'error' one could make would be to assume no regional importation by specifying the 'technical' matrix. This was measured at 57.4. The closest nonsurvey approximation to the $'\mathbf{R}$ matrix was assumed to be the 'RAS'ed domestic use matrix, $'\mathbf{R}^{**}$, *i.e.*

$$'\mathbf{R}^{**} = \hat{\mathbf{x}}\,'\mathbf{A}\hat{\mathbf{y}} \qquad [8.4]$$

where

$\mathbf{x}, \mathbf{y}$ are RAS row/ column multipliers, whose targets are derived from $'\mathbf{R}$.

The STPE between $'\mathbf{R}$ and $'\mathbf{R}^{**}$ was calculated at 17.3. Table 8.5 below summarises these results.

Table 8.5 The Potential for Reducing Error in the Components of Nonsurvey Estimation

	STPE	
	Maximum	Minimum
Total Use	71.2	59.1
Overseas Imports	20.6	0
Regional Imports	57.4	17.3

These results provide a crucially important insight into the nature of nonsurvey misspecification. The implication is that, whilst the level of error in the specification of regional import propensities could, feasibly, be reduced by around 80% to a level which may well be insignificant, the potential for reducing the misspecification of the total use coefficients is a relatively marginal 20%; and moreover, the implied 120% error that remains is a significant concern. Somewhat ironically this implies that the 'trade-only' approach to the development of nonsurvey methods are justified! The justification however arises *not* because the nonsurvey error associated with total use misspecification is insignificant, but because such errors are, largely, *intractable*. It seems that the decision to make use of the national framework of input-output relationships involves an inevitable and significant departure from the regional relationships that would be observed by survey means. More is said on the implications of this below.

However, the view is confirmed by looking at the more general potential for the reduction in nonsurvey error. From this perspective, the maximum misspecification of the matrix '**R** is considered to be the national domestic use matrix, $'\mathbf{A}^{*}$. The justification for this choice is that the matrix $'\mathbf{A}^{*}$ is the most basic of all nonsurvey 'methods'. The STPE between the observed regional use table and the UK domestic use matrix was measured at 102.9 - an implied level of error in excess of 200%. Constraining the $'\mathbf{A}^{*}$ matrix to the intermediate row and column sums of $'\mathbf{R}$ by the RAS procedure gives a matrix $'\mathbf{R}^{***}$, which can be considered as reasonably the 'best' nonsurvey estimate of $'\mathbf{R}$. The STPE between $'\mathbf{R}$ and $'\mathbf{R}^{***}$ was calculated at 58.3. In other words, the error generated by the 'worst' nonsurvey method could, potentially, be reduced by half. Within the context of the evidence from Table 8.5, the implication is that this reduction in error is largely due to an improvement in the estimation of regional import propensities; the error that remains, which is significant at a level in excess of 100%, is due principally to the use of the national framework of input-output relationships.

This evidence restores the credibility of the regional input-output specification. It proves that regional economies are heterogeneous and idiosyncratic in structure and that, consequently, at the level of detail afforded by the UK national model, regional input-output production functions cannot be satisfactorily modelled by 'single-step' procedures. Improvements in the homogeneity of the national model should improve matters - although from Figure 8.1 above, a quite substantial improvement would seem necessary. The findings underscore the fact that survey information is an essential ingredient of any regional input-output table. This is not to say that there is no need for a nonsurvey-type methodology, but that such a technique should remain firmly behind a strong survey-based prerequisite. However, the most incisive conclusion from this analysis concerns the 'technology versus trade' debate and its implications for the deployment of survey resources. Clearly, the 'technology-doesn't-matter' school would have the regional model builder working within a national input-output framework, using the available survey resources to assess local consumption propensities. But what Table 8.5 reveals is that, potentially, local consumption propensity can be dealt with within a relatively broad method of approach, and that survey resources should be devoted to the estimation of regional-specific production functions.

However, with knowledge of the 'worst' and 'best' levels of nonsurvey error, the next question is obviously, where do conventional nonsurvey trade estimators stand within this scale? In order to answer this, nonsurvey estimates of $^r\mathbf{R}$ were generated by applying SLQ and CILQ formulae to the national domestic use matrix $^r\mathbf{A}^*$. The STPE between the observed matrix and the standard nonsurvey estimates was measured at 82.4 for SLQ and 86.6 for CILQ. Table 8.6 summarises. It would seem, therefore, that there remains some significant potential for improving the standard 'third-best' methods of estimation. The remainder of this chapter seeks to achieve this objective.

Table 8.6 The Potential for Reducing General Nonsurvey Estimation Error

	STPE
'Worst' Estimate	102.9
CILQ	86.6
SLQ	82.4
'Best' Estimate	58.3

8.4 The Development of the 'Third-Best' Estimation Methodology

This section of the analysis seeks to estimate the parameters of the 'third-best' methodology set out in Chapter 5 from the Scottish and UK data set. The estimation of the interindustry equations precedes that of households. The first section considers the estimation of the parameters for the intermediate row adjustment equation. The second section considers the estimation of the intermediate column adjustment formula. The third section considers household expenditures.

8.41 Estimation of the Row Sum Equation

8.411 Simple Model

In Chapter 5, the row sum equation was specified as:

$$H_9: \quad \frac{r_i}{u_i^{\bullet} k_i} = \alpha q_i^{\beta_0} \quad , \quad 0 < \alpha < 1,\ \beta_0 > 0 \qquad [8.5]$$

where

r_i are the intermediate row sums of the Scottish regional use matrix, $^r\mathbf{R}$;

$u_i^{\bullet}$ are the intermediate sums from the nonsurvey combined use matrix

k_i is the propensity to purchase commodity i domestically

q_i are the simple location quotients.

The parameter α was believed to express an 'average' propensity to consume locally. The parameter β_0, termed the demand elasticity, expresses the extent to which relative increases in regional output supply are 'absorbed' by local industry demand.

The data was assembled from the 99 sector model. The variables of equation [8.5] were transformed into natural logarithms. This gave the linear equation:

$$\ln\left(\frac{r_i}{u_i^* k_i}\right) = \ln(\alpha) + \beta_0 \ln(q_i) + \varepsilon_i \qquad [8.6]$$

where

ε_i is a stochastic parameter

The parameters of equation [8.6] could then be estimated by Ordinary Least-Squares (OLS) regression.

One observation, relating to Public Administration, was excluded from the estimation because, in both UK and Scottish tables, it was not classed as making intermediate sales. Figure 8.5 illustrates the relationship between the dependent - $\ln(r/u^*k)$ - and explanatory - $\ln(q)$ - variables. Table 8.7 below gives the results of the OLS regression.

Figure 8.5 Variables in Equation [8.6]

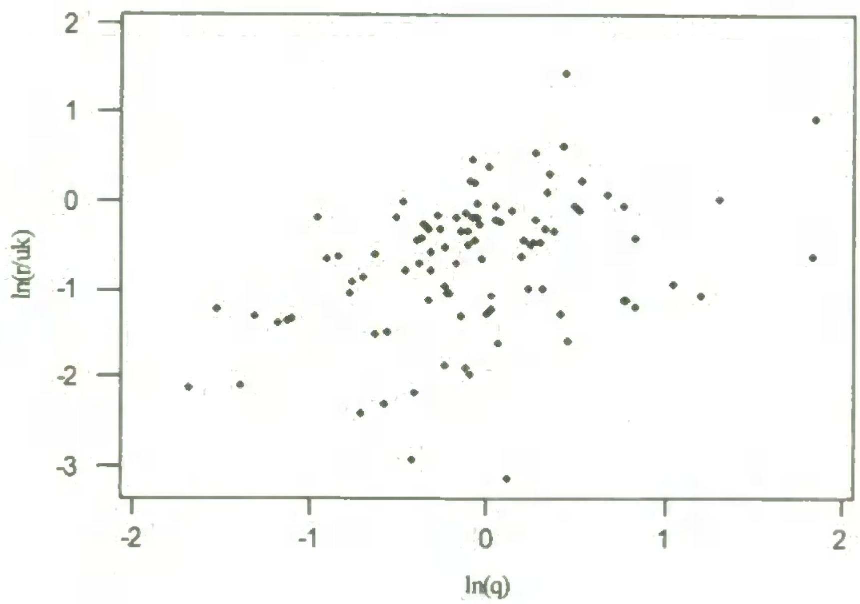


Table 8.7 OLS Regression of Equation [8.6]

Degrees of Freedom	96
Adjusted R^2	0.136****
Standard Error of Regression	0.738
$\ln(\alpha)$	-0.653**** (0.075)
β_0	0.480****(0.119)

**** *significant at 1% or better ($p<0.005$)*

() *standard error of coefficient*

The test of model significance - whether the model has any explanatory power -is an F test. Specifically, the value of R^2 is tested to see whether it is significantly greater than zero[2]. Following Pindyck and Rubenfeld (1976)

$$F_{k-1,n-k} = \frac{R^2}{1-R^2}\frac{n-k}{k-1} \qquad [8.7]$$

where

n is the number of observations;

k is the number of independent variables.

The test of significance for each of the model parameters is a t-test, for example

$$t_{n-2} = \frac{\beta_0 - h}{se(\beta_0)} \qquad [8.8]$$

where

h is the hypothesised value - unless otherwise specified, h=0

se() is the standard error of the estimated parameter.

From adjusted R^2 in Table 8.7, it can be seen that the model explains 13.6% of the variation in the dependent variable. Although this may seem low, the model possesses explanatory power, since the value of R^2 is greater than zero at the 1% level of significance.

---

[2] Strictly speaking, these tests require the estimated equation's residuals to be normally distributed. All estimated equations passed the Anderson-Darling test for normality at 1% significance, although some failed marginally at 5%. Given that sample sizes are relatively large, the validity of the tests will be preserved. See Theil (1971) and D'Augostino and Stevens (1986) for further discussion and detail.

The estimated values of the model parameters were, as anticipated, positive: the estimated value of α was 0.521, and the 99% confidence interval for the scalar was (0.427, 0.634)[3]. Therefore it could be accepted with 99% confidence that 0< α <1. Hypothesis 9 was thus maintained. However, the average local consumption propensity for Scottish industry was calculated from the survey at 0.635 (total Scottish intra-regional intermediate purchases divided by total intermediate purchases from the UK). The belief that the parameter α was equal to the average consumption propensity was therefore, for the moment, rejected.

When the predicted values of the dependent values were transformed back into estimated intermediate row sums of the regional use matrix, $r_i^\bullet$, *i.e.*

$$r_i^\bullet = \left(0.521 q_i^{0.48}\right) \cdot u_i^\bullet k_i \qquad [8.9]$$

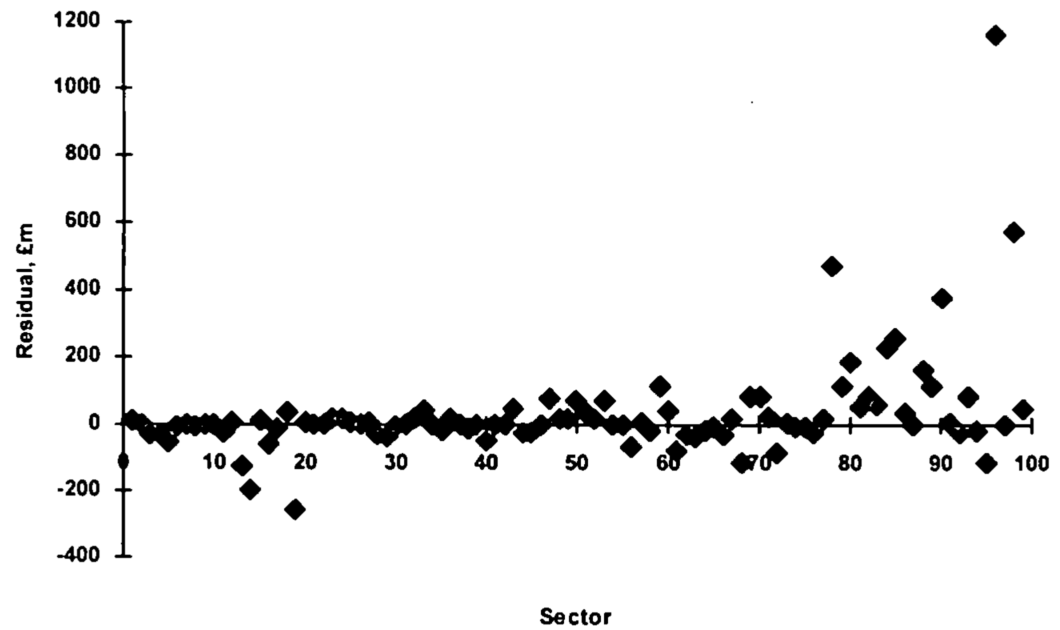
and residual errors, $\varepsilon^\bullet_i$ were computed, *i.e.*

$$\varepsilon_i^\bullet = r_i - r_i^\bullet \qquad [8.10]$$

it was clear that the predictions suffered some degree of systematic bias. Figure 8.6 below illustrates.

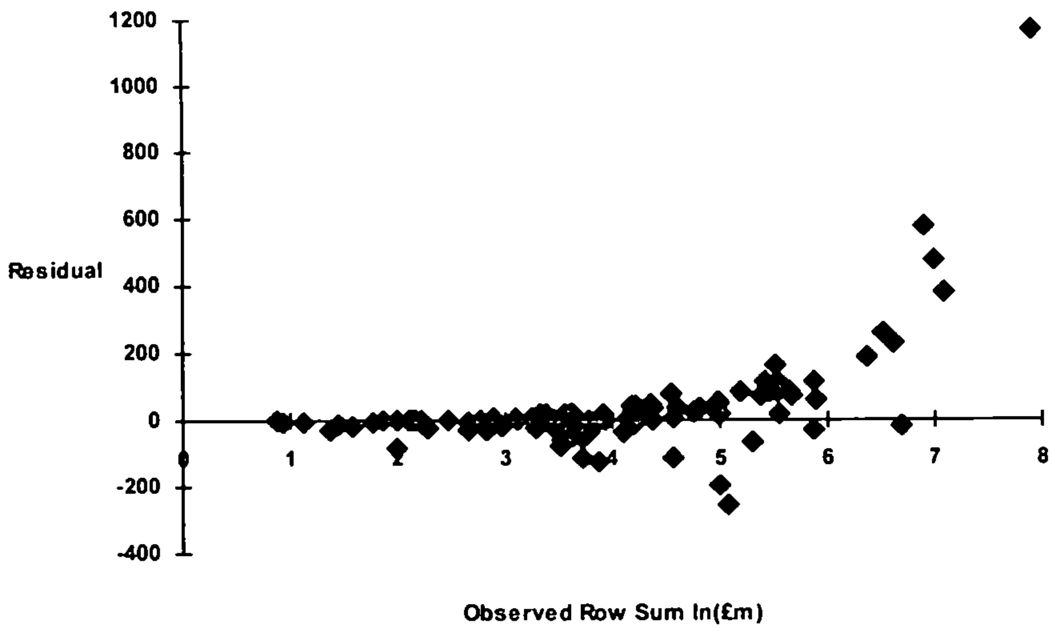
[3] Transformed from the confidence limits of the natural logarithms.

Figure 8.6 Intermediate Row Sum Residuals from Estimated Equation [8.6]



In Figure 8.6, services occupy the higher order sectors. There are two things worthy of comment. Firstly, the variance of the residual values increases as one moves through manufacturers to services. Secondly, the residuals are systematically positive (*i.e.* r_i underpredicted) at the service end of the distribution. The first effect arises because the dependent variable in equation [8.6] is a ratio: a 10% estimation error for a 'large' sector is of course much greater in terms of transactions then a 10% error for a 'small' sector. The residual pattern occurs because the definition of the service sectors is much broader, and the sectors therefore much 'larger', than for manufacturers. Figure 8.7 below confirms the correlation between the observed intermediate row sum (represented on a logarithmic scale) and its residual.

Figure 8.7 Observed and Residual Intermediate Transactions, Equation [8.6]



The evidence from error simulation studies (West, 1981; Jackson 1991) suggests that one should aim to minimise errors in these 'large' sectors. One possible solution would therefore be to apply some form of sectoral weight to the variables prior to the estimation of equation [8.6]. With this in mind, it should be clear that the estimate of α=0.521 is an unweighted estimate - *i.e.* the average *of* the 98 local consumption propensities, rather than the region's average. The average of the local consumption propensities was in fact 0.577 - comfortably within the 95% confidence limit of the estimate of α. Logically, the α from a weighted regression will approximate the region's, or the 'weighted', average local consumption propensity.

As for the systematic underestimation of the service sectors, hypothesis 10 claims that different commodity groups, such as manufacturers and services, will on average have significantly different local consumption propensities. The systematic overestimation of Figure 8.7 could therefore be explained if the propensity to consume services from local supply was much higher than the 'average' of all commodities. The fact that the observed weighted average local consumption propensity of 0.635 lies significantly above unweighted estimate of α=0.521 is a further suggestion that the local consumption of the relatively few large service groups is higher than for the relatively numerous manufacturing sectors. If α, and for that matter β, are substantially different between commodity groups, a sensible strategy to improving the predictive power of [8.6] would be to estimate *separate* relationships for each identifiable group. The next section investigates.

8.412 Testing for Structural Differences: Service and Non-Service Sectors

In Chapter 8.5 it was stated:

H_{10}: The relationship between dependent and independent variables is significantly different between service and non-service sectors.

In terms of the parameters of equation [8.6] the hypothesis could be stated:

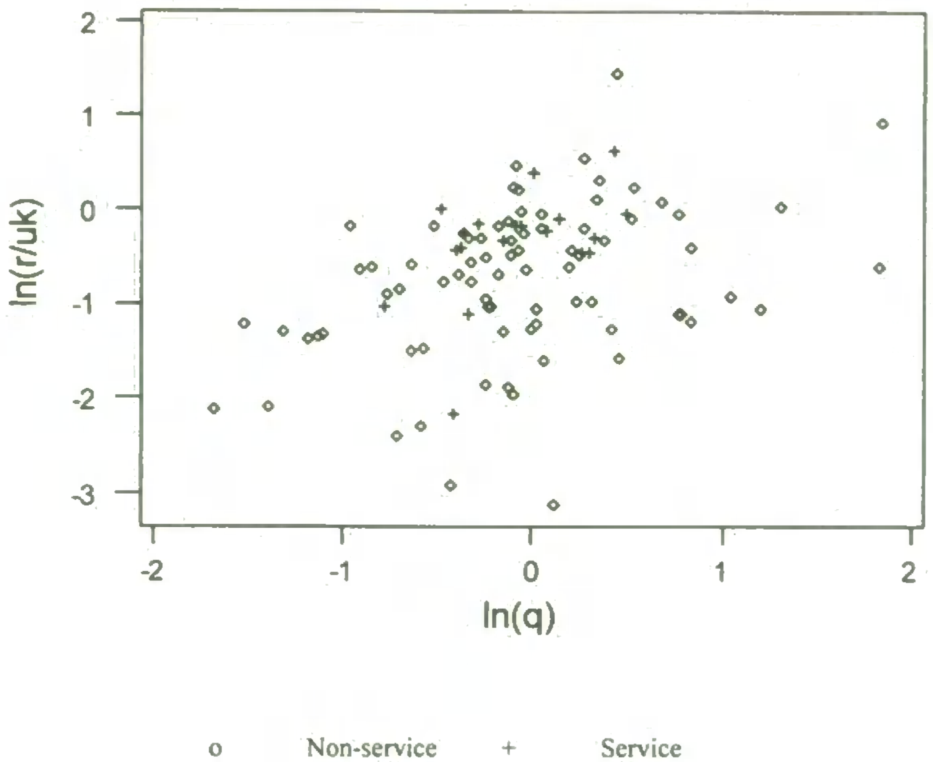
H_{10}: $\alpha_s, \beta_s > \alpha_{ns}, \beta_{ns}$

where

'*s*' denotes Service sectors, '*ns*' Non-services

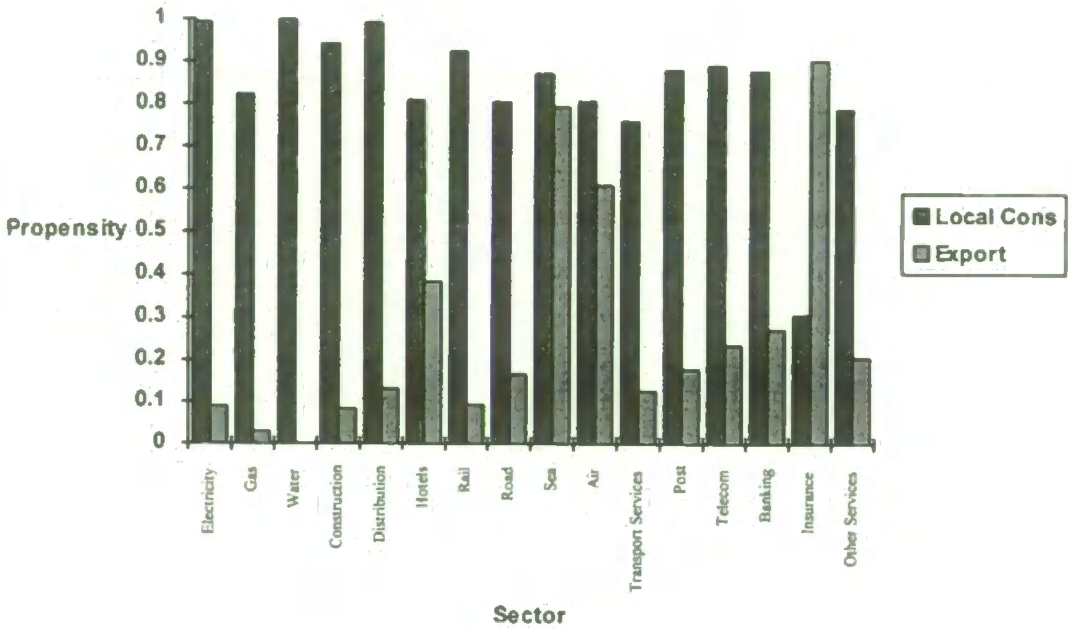
In Figure 8.5 above, the distinction between service and non-service sectors is not apparent. Figure 8.8 provides clarification. In addition to the 'normal' definition of the service sectors (*i.e.* sectors 88 to 98), construction (87), and the utility industries (6 to 8) are counted as activities which service regional industry.

Figure 8.8 Service and Non-service Sector Observations



From Figure 8.8 it can be seen that the service observations are quite closely grouped together. The one exception is the insurance sector which, perhaps predictably, is the only service to have a relatively low local consumption propensity from industry and a high export content. Figure 8.9 below illustrates.

Figure 8.9 Local Consumption and Export Propensities for Scottish Service Sectors, 1989



Hypothesis 10 was tested using dummy variables to pick out the potential structural discontinuities in equation [8.6]. The specification estimated was as follows:

$$\ln\left(\frac{r_i}{u_i^{\bullet} k_i}\right) = \ln(\alpha_{ns}) + \beta_{ns}\ln(q_i) + \beta_{s(\alpha)}D_s + \beta_{s(\beta)}\left[D_s \cdot \ln(q_i)\right] + d + \varepsilon_i \qquad [8.11]$$

where

D_s equals 1 for Service observations, 0 otherwise

d the effect of other dummy variables

Estimates of the service sector parameters α_S, β_S are then given as

$$\alpha_s = e^{\alpha_{ns} + \beta_{s(\alpha)}} \qquad [8.12]$$

$$\beta_s = \beta_{ns} + \beta_{s(\beta)} \qquad [8.13]$$

Clearly, the structural relationship between dependent and independent variables could be said to be different between the two groups if either of the parameters $\beta_{s(\alpha)}$, $\beta_{s(\beta)}$ were significantly different from zero.

In estimating equation [8.11] from the observed data set, the insurance sector, with its high propensity to trade, was classified as belonging to the non-service group. This left 15 observations within the service group, 83 in non-services. A dummy variable was included for observation 76, Clothing and furs, which stands clearly above the principal cluster at about (0.5, 1.5) in Figure 8.8. The intermediate transaction values of r and $u^{*}k$ for this observation were extremely small, thus value of the ratio $r/u^{*}k$ was felt to be relatively unreliable. At this stage, no reason could be found to justify any other observation being assigned dummy variable status. Table 8.8 reports the results of the OLS estimation.

Table 8.8 OLS Estimation of Equation [8.11]

Degrees of Freedom	93
Adjusted R^2	0.244^{****}
Standard Error of Regression	0.690
$\ln(\alpha_{ns})$	$-0.775^{****(0.078)}$
$\beta_{s(\alpha)}$	$0.552^{****(0.184)}$
β_{ns}	$0.443^{****(0.115)}$
$\beta_{s(\beta)}$	$0.157^{(0.490)}$
Dummy for 76	$2.03^{****(0.697)}$

**** *significant at 1% or better (p<0.005)*

() *standard error of coefficient*

The estimated values of the average of the local consumption propensities were 0.461 for the non-service group and 0.800 for the service sectors. Figure 8.9 confirms this as a reasonable estimate for this latter group. The significance of the parameter $\beta_{s(\alpha)}$ at $p<0.005$ indicates that the average of the local consumption propensities for the service group is higher than for non-service commodities. The parameter $\beta_{s(\beta)}$ however is not significantly different from zero, and therefore the elasticity of r/u^*k with respect to a change in relative specialisation is not different between the two groups. Hypothesis 10 is therefore only partially maintained.

However, the above conclusions are somewhat premature, and indeed, may be spurious. Whilst the service/non-service split represents the most logical commodity grouping, from Figure 8.8, it is clearly not the most obvious alternative structural relationship. The service sectors are in fact grouped with the majority of non-service observations: a fairly clear linear relationship however emerges below this principal cluster. Are the conclusions drawn from the results of equation [8.11] therefore spurious? Is it right to identify services as a separate relationship, or is the only proper distinction between the relatively small group of largely non-service observations, and the majority of service/non-service sectors? And do the commodities within this secondary relationship share some identifying characteristic? The next section investigates.

8.413 Testing for Alternative Structural Differences

The commodities which are judged to form part of the 'secondary' linear relationship are given in Table 8.9. Local consumption propensities, export propensities, and location quotients are shown. Figure 8.10 illustrates the full data set in terms of sector numbers.

Figure 8.10 Interindustry Intermediate Row Sum Observations by Sector

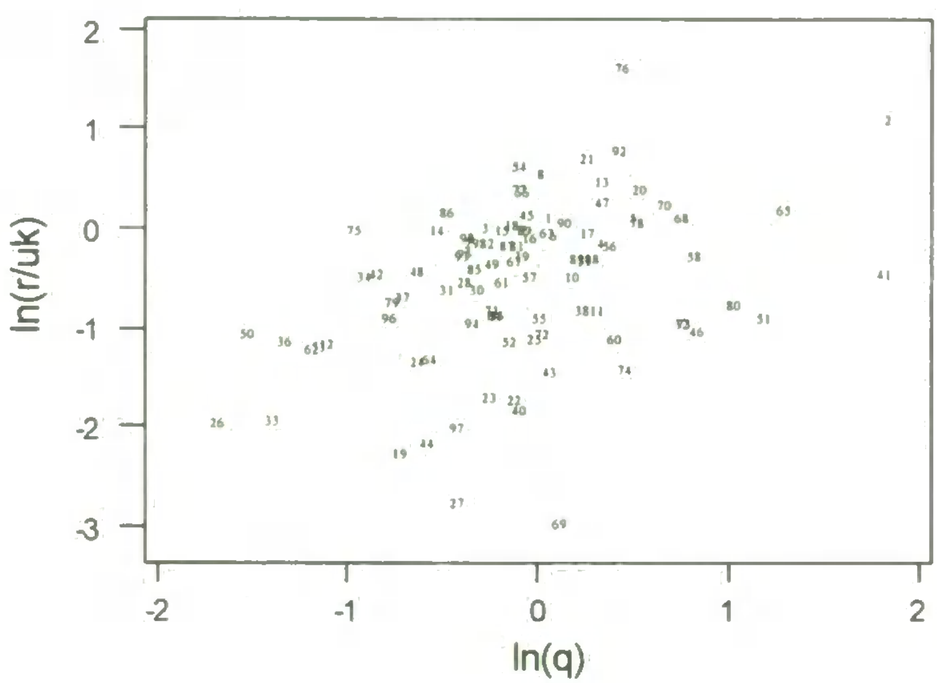


Table 8.9 Sectors within the 'Secondary' Relationship

Sector	Commodity	Local Consumption	Export	Quotient
19	Inorganic chemicals	0.07	0.92	0.49
22	Synthetic resins etc.	0.16	0.84	0.88
23	Paints, dyes etc.	0.15	0.89	0.78
32	Industrial plant and steelwork	0.67	0.86	2.15
40	Other machinery	0.15	0.86	0.91
41	Office machinery and computers	0.42	0.93	6.22
43	Basic electronics	0.36	0.74	1.06
44	Electrical equipment for industry	0.01	0.80	0.56
46	Electronic components	0.14	0.71	2.31
51	Shipbuilding and repairs	0.85	0.72	3.33
60	Bread and biscuits	0.67	0.70	1.51
73	Jute etc.	0.46	0.87	2.18
74	Leather	0.43	0.82	1.58
80	Pulp, paper and board	0.32	0.84	2.83
97	Insurance	0.31	0.90	0.66

The only characteristic obviously common to the commodities in Table 8.9 is that they are all export-orientated. The average export propensity across the 99 sectors was 38%, whereas, for this sub-group it is clearly much higher. If one recalls from the analysis of Chapter 5, those sectors whose output was geared principally towards exports were predicted to form a separate, and potentially problematic, group. However, most of the sectors in Table 8.9 do fall into a single category. The Scottish paper, machinery, chemicals and electronics industries each have a relatively high level of foreign ownership. Table 8.10 below shows the employment split between UK and foreign owned companies in Scottish sectors in 1994[4]. The operations of foreign owned sectors are known to take place on a relatively 'global' scale: that is, they are associated with relatively low levels of local purchasing and high levels of export sales. Lack of autonomy in the purchasing/sales decision of these firms is just one of the reasons why this is observed (see for example Turok, 1993 on the Scottish electronics industry). Unfortunately, the data that would confirm the level of foreign ownership in each individual sector listed in Table 8.9 are not available for reasons of disclosure. The exceptions are 80, Pulp, paper and board (32.2% foreign owned, as in Table 8.10) and 41, Office machinery and computers (68.9% foreign owned). It can only be assumed that the chemicals, machinery and the remaining electrical sectors listed in Table 8.10 are those responsible for the relatively high levels of foreign ownership within the wider definitions of Table 8.9.

Table 8.10 Foreign Owned Employment, Scotland 1994

Sector	Foreign Owned Employment (%)
Pulp, paper 8 board	32.2
Chemicals	39.3
Machinery	27.2
Electronics 8 computers	41.0
Other manufacturing	12.0

Source: ONS Special Analysis from the Census of Production, 1994s

The remaining observations, 51, 60, 73, 74 and 97 could not be justifiably explained as 'inward investor' sectors. The levels of foreign ownership were relatively low in food and drink (5.6%), textiles (6.2%), leather (0%) and 'other' transport (16.9%). Although figures are not to hand, one would assume the Scottish insurance sector is largely UK owned. The manufacturing sectors fit into the 'problem' category identified in Chapter 5. All are regionally specialised commodities (q_i>1), but the regional demand from regional industry is much lower than one would have anticipated (using the notation of Chapter 5 $t_i > q_i$). Table 8.11 below illustrates for these four industries, and for

[4] The reported levels in 1994 are assumed to reflect those in 1989. The sectoral analysis is felt to be

those of a similar degree of specialisation, but which appear, from Figure 8.10, to belong to the principal relationship.

Table 8.11 Demand Components within Specialised Sectors

Outliers	Commodity	*q*	*t*	*r*
51	Shipbuilding and repairs	3.33	0.41	0.85
60	Bread and biscuits	1.51	0.42	0.67
73	Jute etc.	2.18	0.71	0.46
74	Leather	1.58	0.49	0.43
Main Group				
2	Fishing	6.27	2.55	0.99
58	Fruit, vegetables and fish processing	2.28	0.81	0.81
65	Alcohol	3.67	1.25	0.83
68	Woollen and worsted	2.14	1.70	0.55
70	Textile finishing	1.95	1.51	0.72

t equals the ratio of observed local demand to nonsurvey demand

r is the local consumption propensity out of domestic demand

Table 8.11 suggests that the difference between the groups 51-74 and 2-70 is that, for the latter, local demand is more in line with the degree of regional specialisation. However, precisely why this should be is not obvious: there seems little in the nature of the commodities by which one could distinguish the groups *a priori*. The temptation is to conclude from Table 8.11 that 'extra attention' and 'expert' judgement are required in making nonsurvey estimates for the region's most specialised sectors. But of course, the 'extra attention' required by the region's most specialised sectors should probably take the form of a survey based analysis. In the estimation that follows, the sectors 51-74 are treated as outliers from the main relationship, as is insurance.

sufficiently broad for this to hold.

The two observations which lay below both the principal and secondary relationship, 69, Cotton and spinning, and 27 Chemical products and man-made fibres each had extremely low local consumption propensities (0.07 and 0.06 respectively). In the chemicals sector this could, possibly, be attributed to a high level of foreign ownership, although the observation was classed as an outlier to the identified inward investor group. There was no apparent reason for observation 69's low local consumption propensity, and this too was assigned dummy variable status.

The equation to be estimated was therefore:

$$\ln\left(\frac{r_i}{u_i^{*} k_i}\right) = \ln(\alpha_{ns}) + \beta_{ns}\ln(q_i) + \beta_{s(\alpha)}D_s + \beta_{s(\beta)}\left[D_s \cdot \ln(q_i)\right] + \beta_{f(\alpha)}D_f + \beta_{f(\beta)}\left[D_f \cdot \ln(q_i)\right] + d + \varepsilon_i \qquad [8.14]$$

where

f denotes the foreign owned observations

D_f equals 1 for the foreign owned observations, zero otherwise.

Table 8.12 reports the results of the OLS estimation of equation [8.14].

Table 8.12 OLS Estimation of Equation [8.14]

Degrees of Freedom	88
Adjusted R^2	0.734****
Standard Error of Regression	0.410
$\ln(\alpha_{ns})$	-0.435$^{****(0.052)}$
$\beta_{s(\alpha)}$	0.241$^{*(0.125)}$
$\beta_{f(\alpha)}$	-1.365$^{****(0.147)}$
β_{ns}	0.660$^{****(0.08)}$
$\beta_{s(\beta)}$	0.092$^{(0.339)}$
$\beta_{f(\beta)}$	0.075$^{(0.186)}$
Dummy for 27	-2.213$^{****(0.413)}$
Dummy for 69	-2.778$^{****(0.413)}$
Dummy for 76	1.60$^{****(0.416)}$
Dummy for 51, 60, 73, 74 8 97	-1.325$^{****(0.197)}$

**** *significant at 1% or better (p<0.005)*

* *significant at 15% (0.05≤p<0.075)*

() *standard error of coefficient*

Table 8.9 reveals that there were significant structural differences between the identified groups of commodities. The average of the local consumption propensities for the main group of commodities (*ns*) was estimated at 0.647. For the foreign owned sectors, this was estimated much lower at 0.165, as anticipated. The average for the service sectors was estimated at 0.824, which was significantly different from the non-service group at the 15% level (p=0.056 on the marginal difference being equal to zero). Services are therefore justifiably identified as a separate class of commodities (H_{10}). Again, there were no significant differences in the demand response to a change in relative specialisation between the three groups. The successful identification of the three separate groups is reflected by the substantial rise in the model's explanatory power (adjusted-R^2), although this is partly because equation [8.14] utilises more 'outlier' dummy variables than either of its predecessors.

The possibility of improving the model further by giving the largest (and therefore 'most important') sectors greater importance within the process of estimation was then investigated. Attaching relative importance to observations could be achieved by weighting the observations within equation [8.11] by a variable which is suitably reflective of sector size. The intermediate row sum from the Scottish combined use matrix (*i.e.* u_i) was selected for this role, although there were a number of equally feasible alternatives (*i.e.* output, employment *etc.*). The results from the Weighted Least Squares (WLS) estimation of equation [8.14] are given in Table 8.13 below. These are preferred to the previous estimations.

Table 8.13 WLS Estimation of Equation [8.14]

Degrees of Freedom	88
Adjusted R^2	0.801^{****}
$\ln(\alpha_{ns})$	$-0.397^{****(0.044)}$
$\beta_{s(\alpha)}$	$0.291^{****(0.085)}$
$\beta_{f(\alpha)}$	$-1.436^{****(0.105)}$
β_{ns}	$0.607^{****(0.080)}$
$\beta_{s(\beta)}$	$0.093^{(0.205)}$
$\beta_{f(\beta)}$	$0.123^{(0.126)}$
Dummy for 27	$-2.274^{****(0.511)}$
Dummy for 69	$-2.810^{****(0.421)}$
Dummy for 76	$1.581^{****(0.508)}$
Dummy for 51, 60, 73, 74 8 97	$-1.367^{****(0.302)}$

**** *significant at 1% or better ($p<0.005$)*

() *standard error of coefficient*

Table 8.13 illustrates that the WLS regression has only a marginal impact on the estimated model parameters, although the significance of service sectors' average local consumption propensity is increased. The estimated propensities are 0.160, 0.672, and 0.899 for foreign owned, non-service, and service sectors respectively. The estimated elasticity of the dependent variable with respect to regional specialisation falls slightly to 0.607, and a single elasticity is still observed between the three groups. Figure 8.11 below, which compares the transaction value residuals of equation [8.6] (OLS) and [8.14] (WLS), illustrates that there is some success in suppressing the associations between sector size and residual error.

Figure 8.11 Transformed Residuals, OLS and WLS Estimation of [8.14]

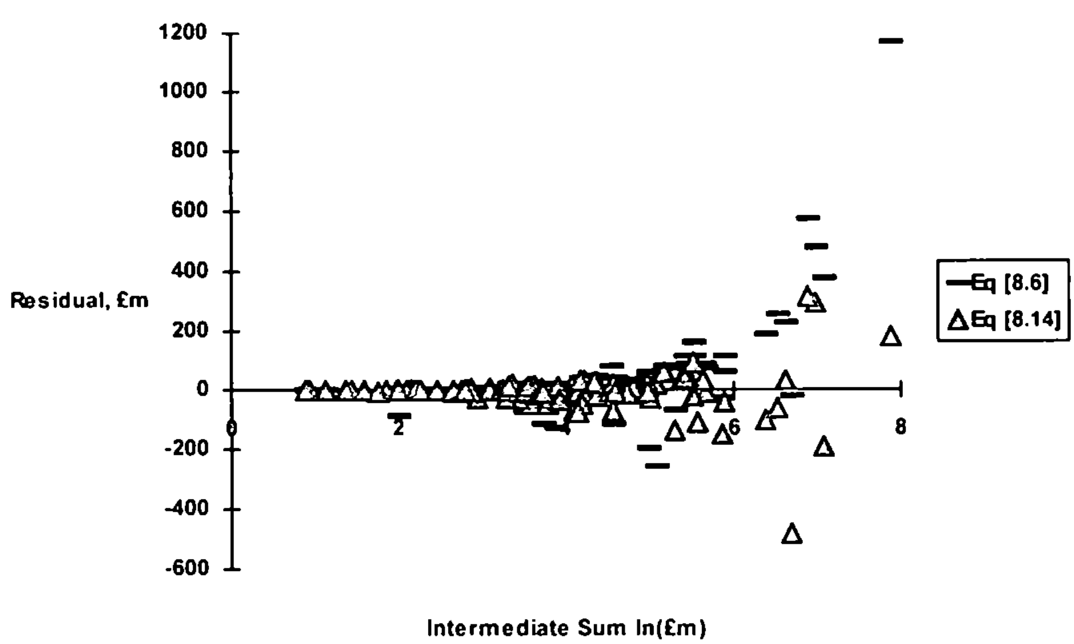
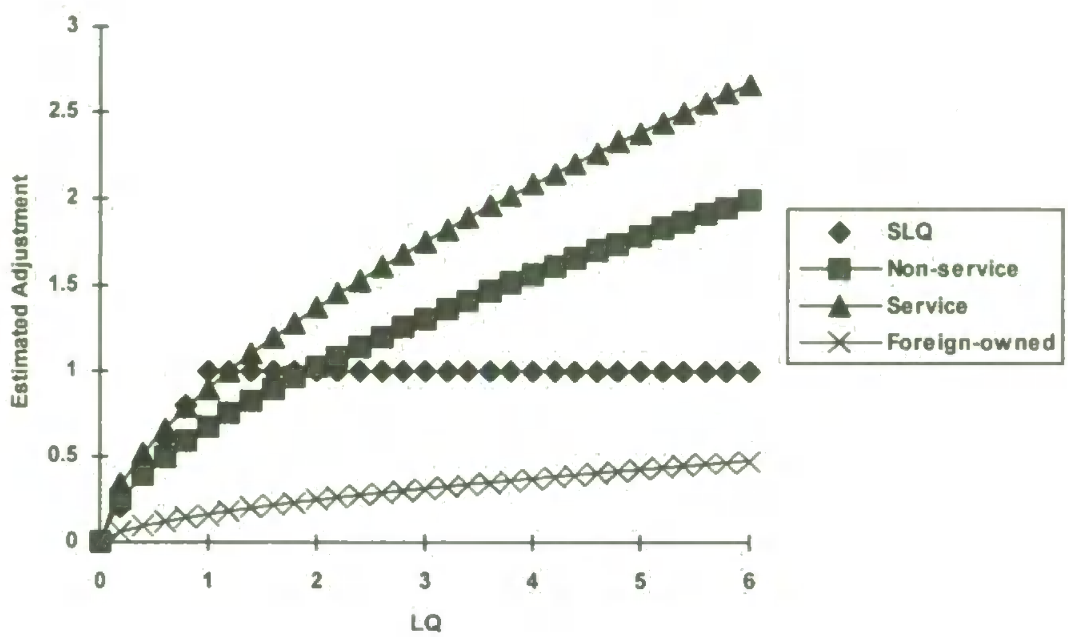


Figure 8.12 provides a comparison between the SLQ and the three core estimating equations of Table 8.13. Note that the estimated equations are non-linear functions, which allow the adjustment process to 'float' above unity.

Figure 8.12 Intermediate Rows: Estimated Functions vs. SLQ



8.42 Estimation of the Column Sum Equation

If one recalls from Chapter 5, total industry purchases (u_j) and the domestic purchase propensities for each industry (k_j) are assumed to be known. In order to provide an estimate of r_j, the intermediate column sum of the regional use matrix, an estimate of the local consumption propensity, l_j ($l_j = r_j/u_jk_j$) is all that is required. Chapter 5 specified the local consumption propensity model as a Stevens-type logit equation:

$$H_{11}: \quad l_j = e^{-1/Z_j} \qquad [8.15a]$$

where

$$Z_j = \alpha q_j^{\beta_1} \quad , 0 < e^{-1/\alpha} < 1 \ , \beta_1 > 0 \qquad [8.15b]$$

As before, the parameter α was believed to capture the average local consumption propensity of the region. The parameter β_1 expresses the extent to which regional suppliers respond to changes in the relative demands of industry (*i.e.* changes in relative specialisation). A positive value for β_1 would be consistent with the view that input supply tends to meet the particular demands of a region's industry (*i.e.* locational economies); a negative value implies a cross-industry relationship where input supply does not respond to regional demand, and regional specialisations therefore 'outgrow' local supply.

Manipulation of equations [8.15a] and [8.15b] gave the linear equation

$$\ln\left[\frac{-1}{\ln(l_j)}\right] = \ln(\alpha) + \beta_1 \ln(q_j) + \varepsilon_j \qquad [8.16]$$

the parameters of which could be estimated by OLS regression. Again, the observation relating to Public Administration was excluded from the analysis because it made no intermediate purchases. This left 98 observations. Figure 8.13 below illustrates the relationship between the dependent and independent variables, listed by sector. Table 8.14 gives the OLS regression of equation [8.16].

Figure 8.13 Variables in Equation [8.16]

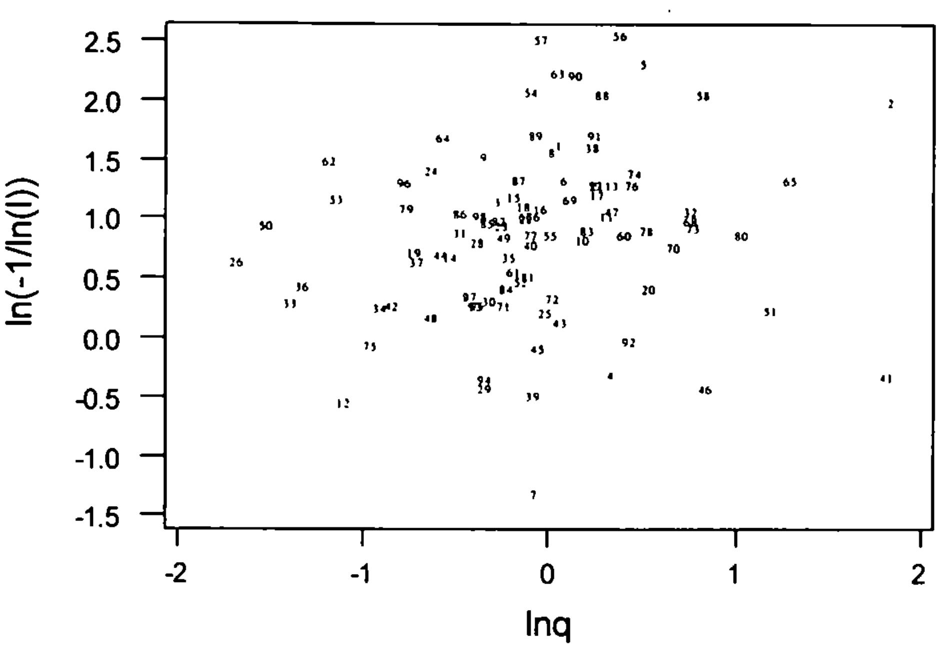


Table 8.14 OLS Regression of Equation [8.16]

Degrees of Freedom	96
Adjusted R^2	0.021^{+}
Standard Error of Regression	0.708
$\ln(\alpha)$	$0.776^{****\,(0.072)}$
β_1	$0.200^{+(0.114)}$

**** *significant at 1% or better (p<0.005)*

\+ *significant at 20% (0.075≤p<0.100)*

() *standard error of coefficient*

The model appears to be quite a poor one, managing to explain only 2% of the variation in the dependent variable, although this is different from zero at the 20% level of significance. The estimate of the parameter $\ln(\alpha)$ is different from zero at the 1% level of significance. The estimate of α is therefore 2.173, which gives an average propensity to consume from regional supply of 0.631 ($e^{-1/\alpha}$). This is consistent with the observed average propensity of 0.635 (the 95% confidence limits for the estimate of $e^{-1/\alpha}$ are 0.58-0.68). The value of β_1 is positive and different from zero at the 20% level of significance. Thus, from Table 8.14, there is only weak evidence to support hypothesis 11.

However, Figure 8.13 illustrates that there are a number of outlying observations. The trading relationships amongst the oil and gas sectors (4, 5 and 7) are complicated - as one might expect. Sectors 56, Slaughtering, and 57 Milk, are particularly well linked with Scottish agriculture. The other two outliers are Scotland's principal foreign owned sectors: 41, Computers and 46, Electronic components. These were identified in the previous section within a larger group of inward investor sectors who bought and/or sold outside the local market. Sector 51, Shipping and repairs, which was also identified in the estimation of the row sums was a possible outlier in the column sums equation. These eight sectors were assigned dummy variable representation. Table 8.15 below shows the results from the OLS estimation of equation [8.16] with dummy variables included.

Table 8.15 OLS estimation of Equation [8.16] with Sectoral Dummies

Degrees of Freedom	89
Adjusted R^2	0.376****
Standard error of regression	0.564
$\ln(\alpha)$	0.827$^{****(0.061)}$
β_1	0.344$^{****(0.100)}$
Dummy for foreign ownership 41, 46	-1.783$^{****(0.429)}$
Dummies for oil and gas sectors:	
4	-1.367$^{***(0.569)}$
5	1.211$^{**(0.571)}$
7	-2.221$^{***(0.5675)}$
Dummy for 56	1.493$^{***(0.570)}$
Dummy for 57	1.577$^{***(0.568)}$
Dummy for 51	-1.110$^{*(0.583)}$

**** *significant at 1% or better (p<0.005)*

*** *significant at 5% (0.005≤p<0.025)*

** *significant at 10% (0.025≤p<0.05)*

* *significant at 15% (0.05≤p<0.075)*

() *standard error of coefficient*

Once the outlying observations are controlled for, the significance of the location quotient parameter increases. It seems therefore that there is evidence to suggest that there is a positive relationship between relative regional specialisation and the propensity to source locally. Local suppliers, it would seem, seek to supply the products and service demand by the region's most characteristic industries. Cross-industry formulae, which assume the opposite association would appear to be an inappropriate specification.

It can be seen that the responsiveness of regional suppliers to meeting the demands arising from increases in regional specialisation is less than local demand's responsiveness to increases in the local availability of products.

The criteria of targeting of the 'most important' sectors suggests, once again, that WLS should be the appropriate method of estimation. The weighting variable was again taken from the intermediate sum of the combined use matrix - u_j, and a weighting of $u_j^{0.5}$ was found to be most appropriate. The WLS regression results are presented in Table 8.16 below.

Table 8.16 WLS Estimation of Equation [8.16] with Sectoral Dummies

Degrees of Freedom	89
Adjusted R^2	0.545^{****}
$\ln(\alpha)$	$0.885^{****(0.057)}$
β_I	$0.276^{***(0.097)}$
Dummy for foreign ownership 41, 46	$-1.790^{****(0.286)}$
Dummies for oil and gas sectors:	
4	$-1.404^{****(0.421)}$
5	$1.186^{****(0.313)}$
7	$-2.284^{****(0.449)}$
Dummy for 56	$1.461^{****(0.394)}$
Dummy for 57	$1.517^{****(0.483)}$
Dummy for 51	$-1.087^{*(0.541)}$

**** *significant at 1% or better (p<0.005)*

*** *significant at 5% (0.005≤p<0.025)*

* *significant at 15% (0.05≤p<0.075)*

() *standard error of coefficient*

The change in the estimates of the model parameters between Tables 8.15 and 8.16 is relatively marginal. The value of the average local consumption propensity is estimated at 0.662 (95% confidence limits: 0.625, 0.696), and the significance of the quotient's parameter has fallen slightly. However, the conclusions remain the same.

8.43 Estimation of the Household Local Consumption Expenditures

The interindustry data is felt to be relatively robust - certainly at the aggregate intermediate row-column level. However, little is known about the estimation method and subsequent reliability of the Scottish household consumption transactions. Thus, whilst the analysis of this chapter should be treated with general caution, the warning applies doubly to this section.

In Chapter 5, it was assumed that figures for consumers' expenditures irrespective of source could be assembled, and national overseas import propensities could be applied. Extending the analysis of error components to the household vector, the STPE associated with the use of national patterns was 57.2 (approximately 115% error); the error associated with the use of national overseas import propensities was 13.7 (30% error); and the error associated with the use of a location quotient trade estimator was 34.6% (70% error). Hence the order of specification importance is maintained, although the level of reported error is notably lower. This may reflect a greater use of nonsurvey methods. Nevertheless, taking the figures at face value, one can conclude that the specification of a regional pattern of consumption expenditures and a regional import function are of importance, whilst the estimation of regional overseas propensities are rather less of a priority.

Attention is therefore focused upon improving the estimation of the regional import function. In Chapter 5 it was stated:

H$_{12}$: $c_i = e^{-1/\pi_i}$ [8.17a]

where

$\pi_i = \delta q_i^{\beta_2}$ $, 0 < e^{-1/\delta} < 1 , \beta_2 > 0$ [8.17b]

Moreover, separate structural relationships were hypothesised between service (s) and non-service (ns) sectors:

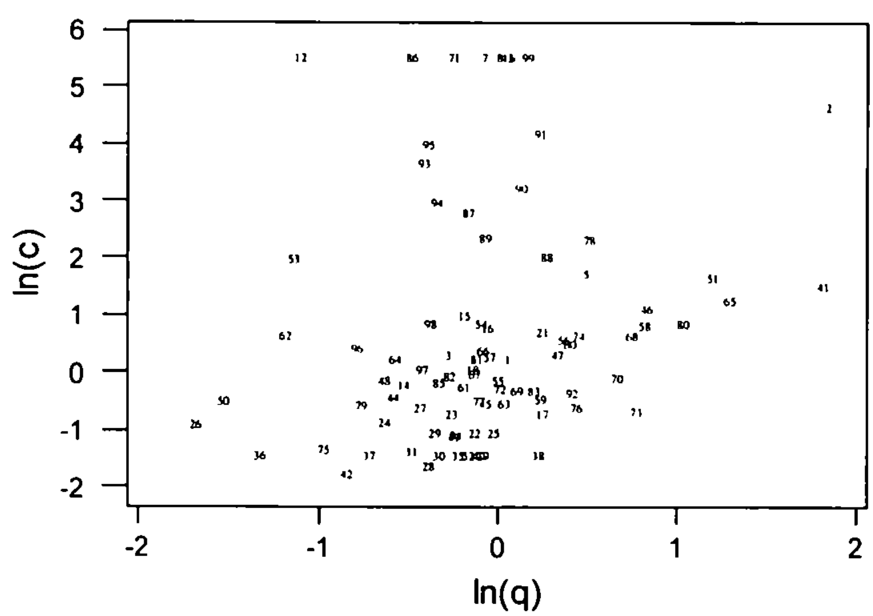
H$_{13}$: $\delta_s, \beta_{2s} > \delta_{ns}, \beta_{2ns}$ [8.18]

As before, [8.17a] and [8.17b] can be rearranged to form the linear equation:

$$\ln\left[\frac{-1}{\ln(c_i)}\right] = \ln(\delta) + \beta_2 \ln(q_i) + \varepsilon_i \quad [8.19]$$

Ten of the 99 observations were undefined because of zero domestic expenditure. However, several further dependant observations were undefined because c_i was either 0 or 1. Seven sectors where c_i =0 were assigned a value c_i =0.005; eight sectors where c_i =1 were assigned c_i =0.995. Figure 8.14 below illustrates the relationship between dependent and independent variables.

Figure 8.14 Variables in Equation [8.19]



The result of the OLS regression of [8.19] is reported in Table 8.17 below.

Table 8.17 OLS Estimation of Equation [8.19]

Degrees of Freedom	87
Adjusted R^2	0.026*
Standard Error of Regression	2.02
$\ln(\delta)$	0.519*** (0.216)
β_2	0.631* (0.347)

*** *significant at 5% (0.005≤p<0.025)*

* *significant at 15% (0.05≤p<0.075)*

() *standard error of coefficient*

Hence relative specialisation had a weak positive effect on local consumption propensity as a whole. However, the results of Table 8.17 can clearly be improved upon by identifying outliers and distinguishing between sub-groups of commodities.

The undefined observations, in particular those assigned a value of 0.995, create something of a problem in the way they 'sit' at the top of the distribution. The values of expenditure for observations 12, Other non-ferrous metals (£1.6m); 71, Textile finishing (£1.42m); 43, Basic electrical (£5.2m) were small enough to class as 'outliers' and ignore. The expenditure upon the remaining sectors however was much greater. All were within the service classification: 6, Electricity (£588m); 7, Gas (£327m); 8, Water (£60m); 86, Construction (£86m); 99, Public administration (£2028m). Figure 8.15 below reveals that most of the service sectors had a high (>0.8) local consumption propensity. The exceptions were 92, Air transport; 96, Banking and finance;

97, Insurance; 98, Other services. Quite feasibly, these two groups could be distinguished *a priori*, or from 'expert' judgement - 'other services' perhaps being the only surprise. The part of hypothesis 13 that relates to δ could therefore really only be partially maintained, since a good portion of the service sector observations are located within the principal cluster In the estimation that follows, these imported services are treated as non-services. Sector 2, Fishing (£62m) - which, predictably, had a very high local sourcing content - was allocated a dummy variable. Sector 53, Other vehicles was excluded from the analysis at only £4.0m domestic expenditure. Of the 85 observations that remained, 13 were counted as service, one was allocated a dummy variable, which left 72 in the non-service classification. The equation to be estimated was:

$$\ln\left[\frac{-1}{\ln(c_i)}\right] = \ln(\delta_{ns}) + \beta_{2ns}\ln(q_i) + \beta_{s(\delta)}D_s + \beta_{s(\beta_2)}\left[D_s \cdot \ln(q_i)\right] + d + \varepsilon_i \qquad [8.20]$$

Table 8.18 below reports the results of the OLS estimation of [8.20].

Figure 8.15 Local Sourcing Propensities out of Consumers' Expenditure for Services, Scotland 1989

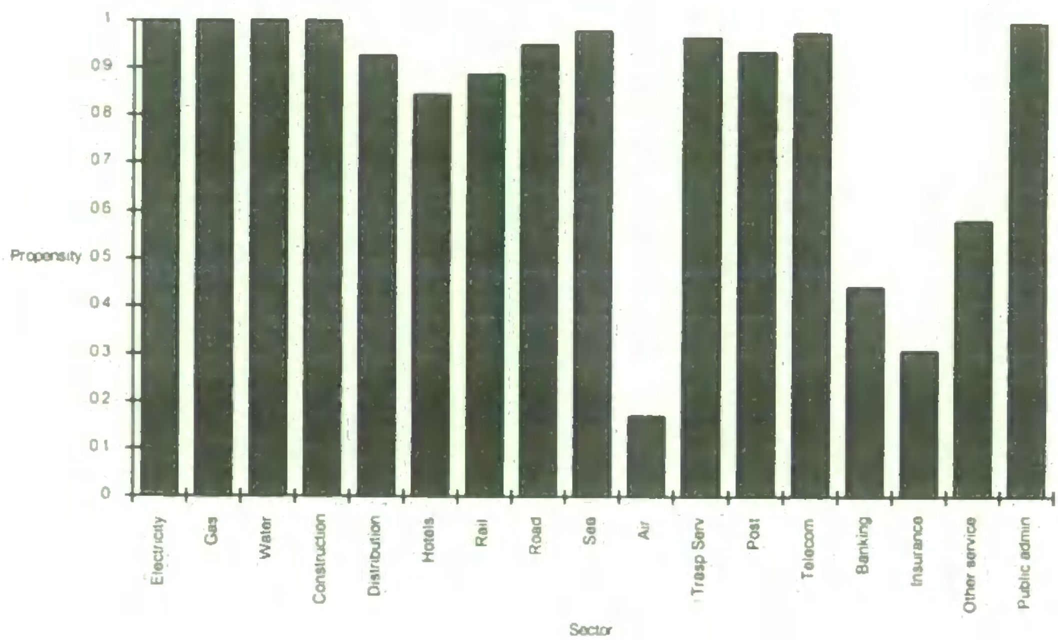


Table 8.18 OLS Estimation of Equation [8.20]

Degrees of Freedom	80
Adjusted R^2	0.773****
Standard Error of Regression	0.893
$\ln(\delta_{ns})$	-0.364**** (0.107)
$\beta_{s(\delta)}$	4.165****(0.283)
β_{2n}	0.776****(0.171)
$\beta_{s(\beta 2)}$	-0.986(1.028)
Dummy for 2	3.337****(0.957)

**** *significant at 1% or better (p<0.005)*

The estimated value for the extent to which local sourcing responds to a change in the relative specialisation of local supply was, as anticipated, positive. Its value was not significantly different for the service group. The estimated values of the average local consumption propensities were 0.237 -

95% interval, (0.172, 0.308) for non-services and 0.980 (0.952, 0.990). As anticipated, or rather by design, they were significantly different. Hypothesis 13 could not be accepted as a general rule - a number of service sectors, such as utilities, are purchased wholly from local supply, whilst others, such as insurance and banking, are characterised by much more open markets, and consequently belong to the more general group of traded commodities. On this, and indeed other points, the Scottish evidence should serve as a useful reference for other studies. The observed values for the average local consumption propensities for these groups were 0.381 and 0.938 respectively, which lay outside the confidence limits for the estimated values of δ. Weighted Lest Squares was therefore applied. A number of different weights were tried. The squared root of the observed domestic consumption purchase was preferred. Table 8.19 below gives the results of the WLS estimation.

Table 8.19 WLS Estimation of Equation [8.20]

Degrees of Freedom	81
Adjusted R^2	0.769^{****}
$\ln(\delta_{ns})$	$-0.161^{(0.133)}$
$\beta_{s(\delta)}$	$3.789^{****(0.238)}$
β_{2n}	$0.559^{***(0.210)}$
Dummy for 2	$3.532^{***(1.224)}$

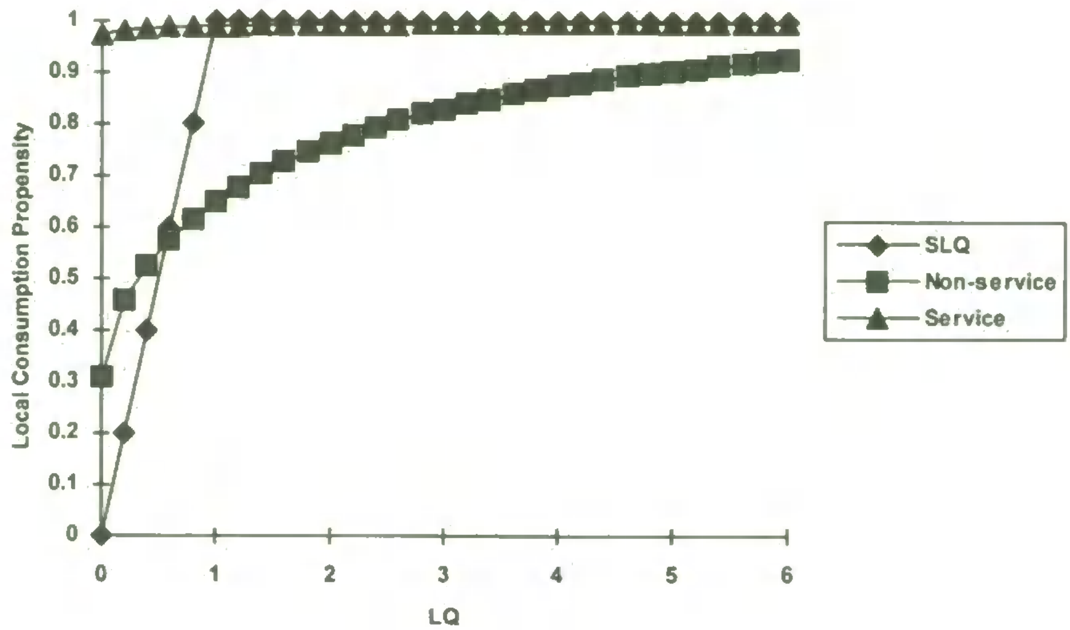
**** *significant at 1% or better (p<0.005)*

*** *significant at 5% (0.005≤p<0.025)*

The estimated local consumption propensities were now 0.310 (0.216, 0.406) and 0.973 (0.945, 0.987), which were slightly closer to their hypothesised observed values. The estimated value, and significance (p = 0.009) of the slope parameter fell slightly, although remained significantly positive. Hypothesis 12 is therefore maintained; as previously mentioned however, hypothesis 13 is not generally supported as there are a number of 'open market' services. Notably, the estimate of 0.559 for the slope parameter compares with the Stevens *et al.* (1983) RPP estimate for US industries of 0.510.

Figure 8.16 below compares the SLQ function to the two core equations. The figure shows that, for services, the function does not fall below 0.95 across the range of regional specialisation. For non-services, the difference between the SLQ function is quite marked with regional imports predicted for even highly specialised sectors.

Figure 8.16 Household Local Consumption Propensities: SLQ vs. Estimated Functions



Hypothesis 14 stated that, given the nature of retailing, the local consumption propensities out of consumers' expenditure may be lower than for industry input purchases:

$$\mathbf{H_{14}}: \delta < \alpha \qquad [8.21]$$

The observed average local sourcing propensity for industries was 0.64; for consumers it was 0.68. Hypothesis 14 was therefore rejected at a general level of observation. However, there is some suggestion of a lower propensity for consumers' expenditure on local non-services, which is compensated for by a slightly higher local consumption propensity on services. The observed figures are 38% for consumers on non-services, compared to 57% for the same defined set of non-service industries. Consumers' local spending on services was observed at 94%, compared to 91% for the same definition of service industries. However, given the frailties of the consumers' expenditure data, no firm conclusions can be drawn.

8.5 Moving to a Generalised Second-Best Estimating Methodology

The parameters of the estimating equations for Scotland will be different for other regions. This section considers how they could be specified by applying a 'survey-based-nonsurvey' approach.

The specified equations are not intended for use as a purely nonsurvey procedure. As it has been shown for the interindustry matrix, with *perfect* specification (*i.e.* RAS), the levels of associated estimation error are unacceptably high. Indeed the notion of a single-step nonsurvey procedure is strongly rejected: the input-output model must be marketed and implemented as an evolving regional economic database. The equations are therefore intended for use within a project which has already generated survey information. Their use is to provide estimates of the least important input-output relationships - however judged - for which there is no survey based information. Therefore, given that regional survey data will have been generated, it seems logical to suggest that these data be used to provide estimates for the model parameters. Certainly an estimate of the region's average local consumption propensity, and feasibly estimates for different commodity groups such as service and non-service sectors, could be provided from sample information. This is the general principle behind 'second-best' estimation. One would of course have to be aware of the possibility of sample bias. For example, if the sample consisted only of foreign owned firms, one would have to question the general applicability of the estimated purchasing propensities. However, this merely serves to emphasise that judgement, and above all common sense, should take a central role in the estimation procedure.

This is particularly the case in the identification of 'outlying' sectors (*i.e.* those picked up by the sectoral dummies). As it has been argued, any idiosyncratic and outlying sectors should probably have been the subject of survey investigation: thus the need to account for such observations may well have been eliminated by the time the second-best estimation phase is reached. However, after the core estimates for second-best sectors have been produced, it should be important to filter them through expert opinion in order to identify any likely outliers and errors. Judgement is an essential tool.

Judgement will be required in the specification of values for the elasticity of demand/supply with respect to regional specialisation, as sample survey specification seems difficult. If sufficient sample information exists, possibly, an estimate could be derived by regression. However, in this unlikely event, more general advice is required. There would seem a fairly good case for suggesting that each should remain positive and relatively inelastic (*i.e.* $0<\beta_0$, β_1, $\beta_2<1$) for most regions. Notably, the demand elasticity of 0.61 and consumers' local sourcing elasticity of 0.56 estimated for Scotland compares with the RPP elasticity of 0.51 as estimated for US regions by Stevens *et al.*

(1983). It is less certain that the elasticity of demand (intermediate rows) should, in general, be greater than the elasticity of supply (intermediate columns). Furthermore, if, as the scale of regional size diminishes, one observes increasingly less intermediation, the relevant elasticities may fall, and indeed, at some point of regional size, reach zero. However, these issues could really only be resolved by extending the analysis of this chapter to models which represent a range of regional sizes, and this is of course not possible.

A schematic representation of the proposed 'new' hybrid procedures is given in the next chapter.

8.6 The Relative Performance of the New Estimation Methodology

Whilst it seems certain that the analysis of this chapter has resulted in a third best methodology which is theoretically superior to the classic location quotient approaches, the question is, does this improvement in specification and the move to a second-best information set represent an improvement in simulation performance? In Chapter 5, it was stated:

H_{15}: The 'new' estimation methodology performs significantly better than its nonsurvey predecessors.

Recalling the evidence of Table 8.6 (reproduced below), it was clear that, whilst there was some potential for making a practical improvement to nonsurvey estimates, the scale of this improvement was never going to be dramatic. Even if the RAS data set could be reproduced to perfection by nonsurvey means, the level of error in the estimated regional interindustry matrix would still lie in excess of 100%.

Table 8.6

	STPE
'Worst' Estimate	102.9
CILQ	86.6
SLQ	82.4
'Best' (RAS) Estimate	58.3

The location quotient error in estimating regional consumption expenditures was measured at 34.6 (70%).

The equations of Tables 8.13, 8.16 and 8.18 clearly include a number of dummy variables which account for unusual observations, or for a particular group such as foreign owned sectors. The effect of these dummy variables is carried through into the generation of predicted values because this represents the addition of second-best, survey-based-nonsurvey, information and 'expert opinion' (*i.e.* the ability of local economists to identify the region's idiosyncrasies).

Predicted values were generated from the equations represented in Tables 8.13 and 8.16. Two simulations were undertaken. The first used only the row sum estimates to generate the Scottish regional use matrix from the nonsurvey domestic use matrix, '$\mathbf{A}^*$'. The second simulation used both row and column sum estimates to constrain $'A^*$ using the RAS algorithm. Table 8.20 below records the results of the exercise.

Table 8.20 Performance of the New Estimation Technique: Interindustry Transactions

	STPE
Rows only	73.3
RAS constrained	66.3

Given the potential for reducing error, the estimating equations can claim some success in achieving their aim. The random error associated with the CILQ was around 180%. With a perfect specification of the RAS data set, this could be reduced to around 120%. Table 8.20 illustrates that the row and column sum estimating equations succeed in reducing the level of error to around 130%.

At a broader, and perhaps more practical level, the column sum equation should give reasonable estimates of Type I output multipliers (following the Burford and Katz analysis). The STPE between the observed and estimated intermediate column sums was calculated at 12.8 - a random error of around 25% upon which the estimates cannot be rejected as equal to the observed values. In comparison, the SLQ generates an STPE error in the column sums of 31.7, or 60% - sufficient to reject the hypothesis of equivalence between observation and estimate.

If it is felt that there is insufficient information upon which to specify the column sum constraints (possibly the u_j are not available), the estimation methodology could revert to a rows-only approach, in which case it can be seen that nearly 50% of the potential for error reduction has been achieved.

Given the results of Table 8.5, one can conclude that the reduction in error afforded by the 'new' second-best process results from a superior estimation of the regional import function. The remaining error would appear to be largely the result of national 'technology': therefore, survey work *must* be devoted - at the very least in equal proportion - to the estimation of regional-specific production functions. The Stevens 'trade-only' hypothesis is therefore firmly rejected within the new hybrid approach.

Figures 8.17 and 8.18 below take a closer look at the estimation errors for each sector across the elements of its row and column (as opposed to Figure 8.11 which looks at residuals in intermediate rows). The illustrations compare residual errors to sector size, as measured by the intermediate transactions sum. Both charts are represented on a log scale and illustrate a reasonably constant level of estimation error across sectors. There is, as one would expect, a strong positive correlation between residual error and sector size. The studies of Jackson (1991), and to an extent West (1981) indicate that the sales and purchasing patterns of these large sectors are 'important' to get right. This presents the beginning of a broad strategy for targeting sectors for survey: gross intermediate purchases for manufacturers will, in general, be available from the Census of Production. So, for example, the Scottish drinks industry

(65) would have been identified as one of the main purchasers, survey information on its pattern of backward linkage would have reduced the level of associated random error in the hybrid table by over 10%[5]. Notably, the large sectors by transaction sales in the 99 sector model are the extremely heterogeneous groups such as 'other' services (98) and banking and finance (96). It would seem that a step towards an improved estimation - certainly from the perspective of the second-best approach - would be to work with more identifiable *i.e.* homogeneous, service sectors. Indeed, the UK and Scottish input-output table definitions have moved this way (ONS, 1995; Scottish Office, 1996).

[5] Notably, in the Scottish model, the drinks sector is given particular disaggregation. Note, of course, that the error in the *remainder* of the matrix would change relatively little at e=65.6.

Figure 8.17 Residuals in Intermediate Row Elements (Log Scale)

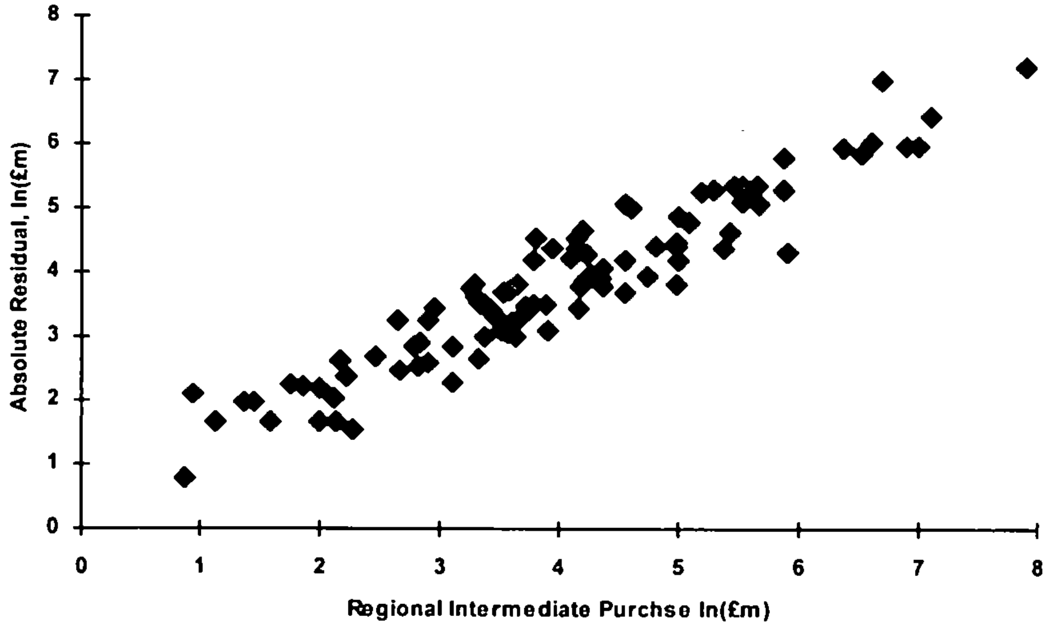
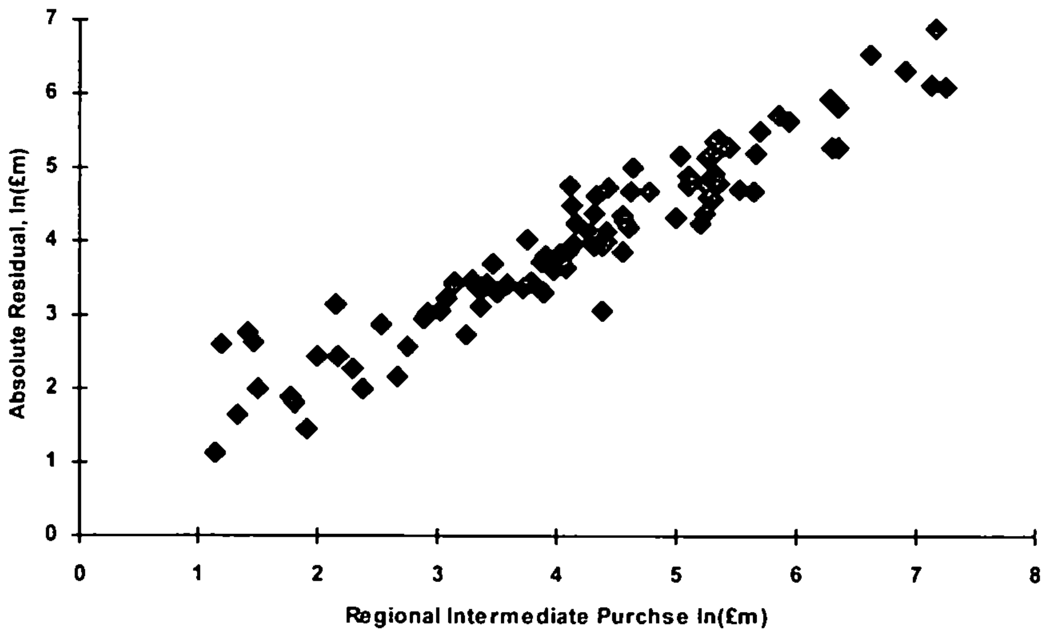


Figure 8.18 Residuals in Intermediate Column Elements (Log Scale)



As for household transactions, the reduction in estimation error is of a degree that is similar to the reduction of error in the interindustry matrix. The STPE is reduced from 34.6 to 16.4, implying a reduction in the level of error from 70% to around 30%. Thus, whilst the location quotient estimates can be automatically rejected as equal to the observed values, those of the second-best estimating equation can not. Again, the it is the heterogeneous 'other services' sector that was the principal contributor to estimation error. Removing this sector from the calculation of STPE reduced its value to 12.9 (25% random error). Once again, improved homogeneity within this sector would provide a platform for error reduction.

In conclusion, both estimation methodologies are considered to represent a significant improvement on the performance of classic quotient approaches. From a purely performance related perspective, hypothesis 15 is therefore maintained. On a conceptual level, restoring the logical preference-order of approach within hybrid procedure should encourage the input-output table to be viewed not as a 'black-box' approach to analysis, but as a framework for assembling a regional economic database. If this affords an increase in the survey content of the average regional input-output model, then it would represent a real improvement to regional economic analysis.

The summary of these results is reserved for the concluding chapter.

CHAPTER 9

SUMMARY, CONCLUSIONS

AND

RECOMMENDATIONS FOR FUTURE RESEARCH

9.0 Summary, Conclusions, and Recommendations for Future Research

9.1 Summary and Conclusions

In Chapter 1, one of the most likely causes of the local economic 'data problem' was identified as the absence of a coherent framework for the collection and assembly of relevant information. A more unified and strategic approach to data generation would, it was believed, ease the data problem, afford a higher level of analytical sophistication, and yield long-run cost and efficiency savings within the process of data collection. The process of generating regional input-output tables was identified as an approach that had the potential to offer such rewards.

In Chapter 2 the regional input-output specification was seen as analytically superior to a number of alternatives. This superiority arises simply because input-output works upon a much larger set of model parameters. But it is important to understand that analytical superiority is not borne out of mathematical wizardry: the parameters have to be specified through a regional economic database in order to reap the rewards. It was shown that the basic Leontief specification was weak in its representation and treatment of demographic features of the economic system, and in its static, ultra-Keynesian perspective. However, solutions to these weaknesses are afforded by building *upon and around* the basic data set. The specification of the regional input-output table therefore plays an integral role in the development of higher levels of analytical sophistication.

With this in mind, Chapter 3 explored the methods of generating the data for the basic specification. The survey-based approach embraces the idea of the regional input-output table as a local economic database. The nonsurvey approach recognises only the relative analytical sophistication of the specification and attempts to provide a 'workable' model whilst avoiding the costs associated with fieldwork. Hybrid procedures recognise the importance of a regional-specific data base, yet note that the collection and collation of such information is invariably subject to diminishing marginal returns: their object is to provide an optimal use of survey and nonsurvey data.

One of the issues at the root of general hybrid procedure is the identification of 'important' elements: this facilitates the effective deployment of survey resources. Techniques that have attempted to deliver a 'shopping list' of coefficients for estimation are not well placed within the conventional nonsurvey-to-survey ordering of popular hybrid approaches. The nonsurvey foundation of such tables will be a poor identifier of the specific set of regional 'inverse important' coefficients. If it was indeed a good guide, what would be the need for survey?

Whilst these techniques have found appropriate uses in the final stages of model construction, and updating existing input-output tables, this identification failure is a fairly good indication that the conventional nonsurvey-to-survey ordering of approach of the popular hybrid paradigm is misplaced.

Nevertheless, some 'holistic guidance' on the deployment of survey resources and the development of nonsurvey techniques is at hand. Firstly, simulation studies have demonstrated the importance of the household sector within the Leontief analytic function. Secondly, studies have attempted to illustrate that differences in regional and national 'technology' are insignificant. Within the hybrid paradigm, this implies that survey resources should be devoted to the estimation of local consumption propensities. The evidence provides a rationale for focusing upon developing nonsurvey import estimation methods.

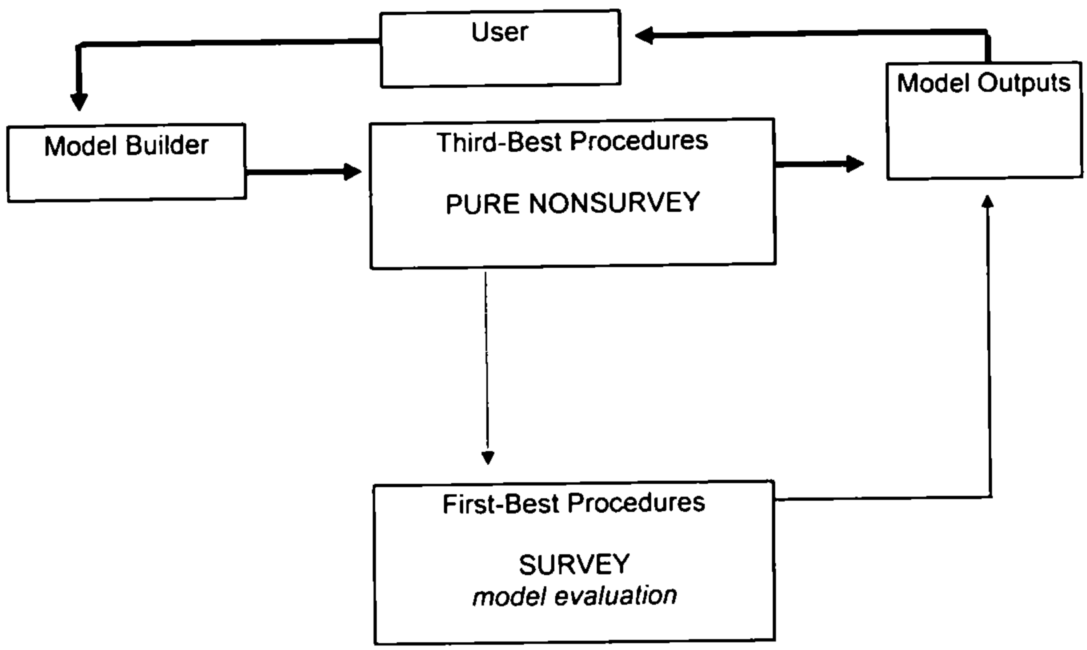
Chapter 4 considered the evidence that has formed the basis for the rejection of the initial 'classic' set of nonsurvey estimators and has been the platform for the development of such techniques. Direct comparison studies were considered to have been poorly executed. Their failure to apply any *a priori* logic to the task in hand would appear to have contributed significantly to subsequent poor practice in regional input-output model building and the development of estimation procedures.

The scenario simulation approach has taken a more general line to the evaluation of input-output models. The conclusion from these studies that the household sector is relatively important provided useful guidance for the development of the 'new' procedures. However, the theoretical evidence supporting the hypothesis that 'trade-only' matters in estimation was firmly rejected. *A priori*, the specification of the regional total use functions seemed at least as important as the regional trade function - but the issue of relative importance had to be resolved from an empirical analysis.

The attempts that have been made to develop nonsurvey approaches were considered poor, given that virtually all had concentrated on import estimation and ignored the (likely) importance of specifying regional production functions. Within the area of import estimation alone, poor practice was found. There was a general failure to question the nature of the estimation problem and the logic of techniques prior to development - a habit seemingly inherited from the direct comparison studies. Particularly irritating was the propensity to 'fiddle' with location quotient formulae - making minor detail-adjustments. This practice rather gives the impression that the basic nonsurvey formulae are well founded, when in fact, the one positive thing to emerge from direct comparison studies has been that the flaws in these estimators are almost certainly fundamental.

All of this led to question *why* the classic nonsurvey methodologies are at the foundation of popular hybrid practice? In a logical preference-order, survey work would be approached first, whilst pure nonsurvey methods would remain a distant third. Placing first-best procedures behind the third-best approach, with the notion that one can pluck at a few choice coefficients, is unlikely to install the image of a regional economic database within anybody's mind. It also creates a dangerous loop-hole in procedure. Model builders find that it is respectable to use nonsurvey procedures, but who is to say at what point a nonsurvey model achieves 'GRIT status'? Whilst it is impossible to police regional input-output models, the guidelines for construction should at least encourage good practice. GRIT would seem only to provide a safe-haven for the very worst facets of input-output practice. Figure 9.1 below will be used by contrast to illustrate the relative strengths of the new hybrid procedure.

Figure 9.1 Schematic Representation of GRIT



Having exposed these problems in Chapter 4, Chapter 5 laid the foundations for an empirical analysis that set out to explain the nature of nonsurvey estimation error. Chapter 6 considered the problem of matrix distance analysis and derived critical values for an appropriate statistic. Chapter 7 justified the selection of data and gave details of the transformation procedures that were required to achieve consistency between tables.

Chapter 8 then sought to identify and explain the nature of nonsurvey estimation error by empirical means. The main findings were as follows:

The error associated with the specification of total use functions and overseas import propensities was found to be positively related to matrix order (H_1 and H_3 supported). This supports the view that heterogeneity within the national input-output sectors contributes to misspecification at the regional level. It was clear that the national input-output table for 1989, defined at 99 sectors, had not reached the point at which the Scottish regional production functions were represented individually. Furthermore there was no evidence as to the level of disaggregation required to reach this point.

The error associated with the use of conventional trade estimators was found to be unrelated to matrix order (H_4 supported). This was taken to indicate that the methods in question suffered some generally deficiency.

The error associated with total use and regional import specification was found to be significant (H_6 maintained). The error associated with the use of national overseas import propensities was markedly the lowest source of misspecification; the significance of this error however could not be determined (H_7 partially maintained).

For Scotland, the specification of the regional production functions was found to be the most important estimation objective. However, this finding probably does not represent the general rule. Limited evidence from the Peterborough study of 1968 suggested that the degree of misspecification in regional total use and regional import functions rose as the scale of regional size diminished (H_2, H_5). Moreover there seemed to be some evidence to suggest that the correct specification of the regional import function becomes relatively more important as regional scale is reduced (H_8). However, whilst the findings were in line with expectations, the evidence was simply too limited to form any firm conclusions (H_2, H_5, H_8 weakly supported). Even so, since both total use and region import misspecification remained significant, any attempt to develop nonsurvey procedure should attempt to account for both functions. Trade-only approaches to nonsurvey development, such as Stevens *et al.*, 1983 and Flegg *et al.*, 1995 are therefore unjustified.

An analysis of the potential for reducing nonsurvey estimation error within the interindustry matrix however revealed that, whilst regional import specification could be significantly improved upon, the regional production functions remained largely intractable within their national counterparts. Heterogeneity was re-emphasised as a significant cause of estimation error. This finding served to underscore the fundamental importance of survey based regional input-output information, and enabled the notion of a 'single-step' nonsurvey technique to be firmly rejected. It followed that improvements in the homogeneity of the national input-output tables would improve any second-best estimation methodology. However, the criticism of the 'trade-only' approach to the development of nonsurvey techniques that was delivered in Chapter 4 becomes slightly unstuck with this finding, for it would appear that the development of broad methods of approach to import estimation is the 'correct' line of strategy to follow. However, this is only because the level of error associated with total use estimation remains *intractable* to a broad method of account. From a *survey perspective*, therefore, the Stevens 'trade-only' hypothesis was firmly rejected: survey resources *must* be devoted to the estimation of regional-specific production functions.

Given the relatively marginal potential for improving the third-best approach, the parameters of the specified household regional expenditure and interindustry intermediate row and column equations for Scotland were estimated by regression from the empirical data. These equations develop upon the Stevens *et al.* (1989) specifications. The scale parameter in the regressions was found to reflect the region's average propensity to consume locally from domestic supply. Relative regional specialisation was found to have a positive effect on the degree of local intermediation (H_9 and H_{11} maintained). In particular, the finding that the elasticity of supply with respect to changes in the relative specialisation of demand was positive allowed the rejection of the cross-industry specification of regional trade determination.

Local sourcing was found to be significantly higher for service as opposed to non-service products (H_{10} maintained). Scottish sectors with a high degree of foreign ownership appeared particularly poorly embedded within the local economy. The extent to which they were sourced from local supply was significantly lower than for indigenous sectors - a finding which has been noted elsewhere (*i.e.* Turok, 1993). A constant demand elasticity with respect to changes in relative regional supply was however found between these groups. The elasticity compared with that of the Stevens *et al.* (1989) study.

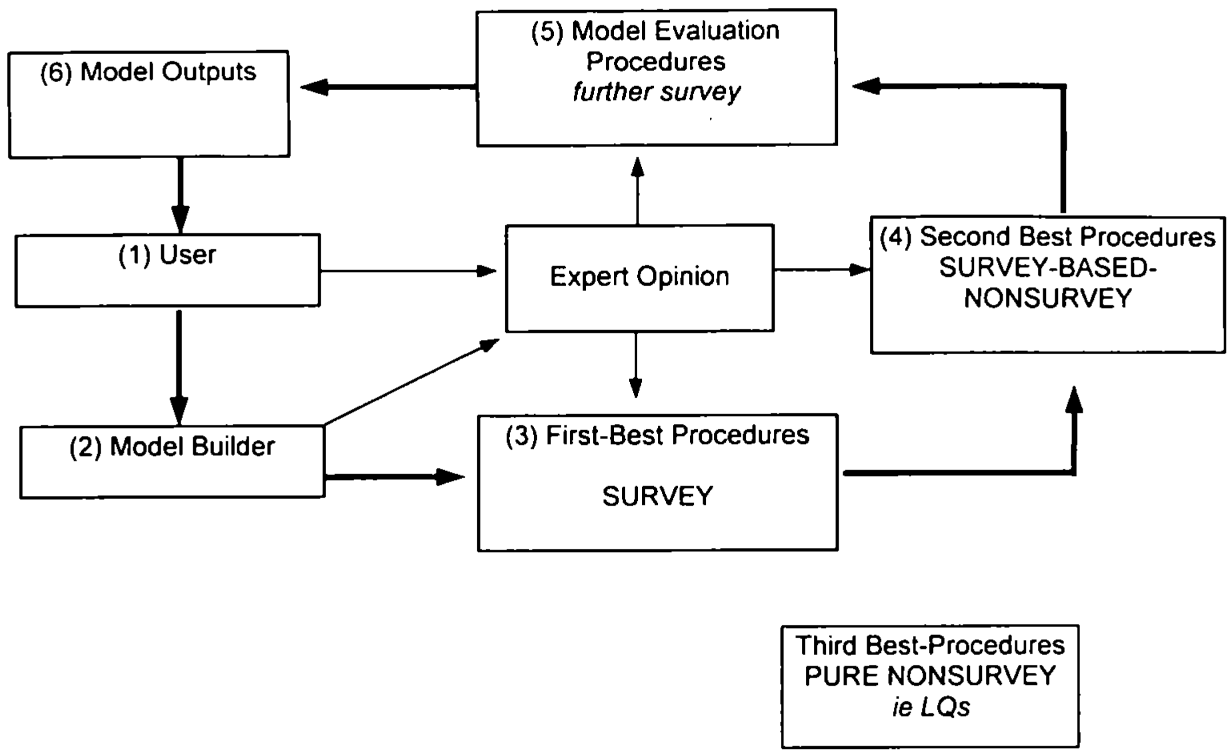
The results for the estimation of the local expenditure content of households were similar, although these data are notably more fragile. Relative specialisation was found to have a positive influence on the propensity to purchase locally made goods. An underlying average local propensity was observed (H_{12} maintained). Whilst a number of the service industries - notably utilities - had significantly higher local purchase propensities, there were a number of more 'open market' services, *i.e.* banking, finance and insurance, which had propensities more in line with the non-service group. A constant slope parameter between service and non-service groups was observed (H_{13} partially maintained). The observed propensity for consumers to purchase local produce was slightly higher than for industries (H_{14} rejected) although there was a hint that the propensity for consumers to import manufactured (service) goods was higher (lower) than for industries.

The final estimating equations could be specified for other regions using survey-based information ('second-best') and expert opinion. The assessment of the second-best technique's performance suggested that the interindustry equations compared favourably with the fully specified RAS data set and the household equation was superior to a location quotient import estimation (H_{15} maintained). However, the error that remained in both simulations re-emphasised earlier conclusions that: (*i*) from the perspective of second-best estimation, improvements in the national model's homogeneity were required; (*ii*) from a survey perspective, the results confirmed the need for a significant survey content in any regional input-output specification, but in particular, resources should be devoted to the estimation of regional-specific production and total expenditure functions. Purchasing data from the Census of Production would represent a reasonable beginning to the formation of a survey strategy.

Of course, the principal limitation of this study is that it has been based upon a single regional model, but this was unavoidable. Studies of other areas - particularly for smaller regions - would help to determine the generality of the findings. This is a potential area for future research.

Figure 9.2 below presents a general representation of the 'new' hybrid procedure.

Figure 9.2 Schematic Representation of the 'New' Hybrid Procedure



The stages follow an anti-clockwise direction that begins with the user issuing instructions and allocating resources to the model builder. The course is cyclical because the input-output table is an evolving database.

Comparison of Figure 9.2 with the representation of GRIT in Figure 9.1 reveals that the 'new' hybrid procedure follows a much more rational ordering of approach. The broken line between nonsurvey and survey phases in Figure 9.1 is considered weak for a number of reasons. Firstly, it is questionable whether the pure nonsurvey model provides any meaningful regional economic information. As a consequence, it will not act as a positive input into the survey phase: for example it will not identify the set of 'important coefficients'. Whilst it may help in forming some general surveying strategy in the form of identifying specialised, non-specialised, large and small sectors, this could almost certainly be determined by simpler means (*i.e.* Census of Production data, considering location quotients *etc.*). Moreover, it is highly likely that the commissioning body will have some idea of the survey strategy it wishes to follow. So what is the point of the third-best model?

If its purpose is to provide some 'rough cut' figures for the EDA to use in the short term, then this is considered a very dangerous strategy. Delivering sectoral multipliers at the beginning of the project only demonstrates that there *are* quick fix, zero cost alternatives: why should the EDA invest in data collection if the figures it requires can be produced in an afternoon? Is the model builder *sure* that the £*x* thousand pounds of investment will produce better results in a year's time? And how to justify the fact that estimated nonsurvey multipliers are - as they inevitably will be - much higher than their eventual survey counterparts? For the purposes of their impact analyses, the EDA may well decide it prefers the story told by the nonsurvey model! In short, the pure nonsurvey phase provides no positive input to the subsequent phase of analysis and, potentially, weakens the case for an investment in survey based procedures.

In Table 9.1 below, the steps of Figure 9.2 are expanded upon.

Table 9.1 **Steps Outlining the Alternative Hybrid Approach**

Step (Fig 9.2)	
(1) & (2)	Discussion with funding bodies to identify survey sectors and resource requirements.
	Design of sample, design of survey questionnaire.
(3)	Implementation of first-best procedures. Construction of model from purchasing and sales estimates given on survey returns.
	Assemble published and estimated data on gross output for each sector, and other published information such as employment, employment income, GDP, consumers' spending.
(4)	Implementation of Second-best procedures. For each commodity, where a good, broad sample of industry purchases is considered to have been obtained, apply the sample estimate of commodity use and trade to standard nonsurvey estimates.
	Where the estimates in (iii) are considered to be lacking in generality, use the estimating equations and expert opinion to provide intermediate row and/or column sum values, and/or consumers' local consumption propensities. Wherever possible parameters for the equations (i.e. average local consumption propensity) should be specified from sample data. Use the row/column estimates to constrain standard nonsurvey estimates.
	Patch sample survey estimates of (iii) in with survey-based-nonsurvey estimates of (v) and (vii) using subjective weighting.
(5)	Balance and evaluate model using standard procedures. Where important cells are identified which are considered to have a weak survey base, carry out further analysis, or prioritise for future attention.
(6)	Deliver model outputs. On the basis of the evaluation, form strategy for project in the coming period.

With respect to GRIT, this approach is considered to have the following advantages:

- The initial interface between model builder has the right focus: the survey. There is no mention of short-cut methods, or quick fix answers.

- The input-output table is marketed as an *evolving* local economic database that will offer benefits in terms of co-ordinating and rationalising information collection, assembly, and analysis.

- It is sold on the analytical strength that these data provide, not on the analytical 'strength' of magic-box mathematics.

- The link between the first-best method and second-best method is a positive one: the first-best methods have an input into the next stage of estimation.

- Note that the role of expert opinion is fundamental at each stage. Whilst the second-best methods have estimating equations at their core, the expert plays an important role in their specification: in identifying potentially idiosyncratic sectors, defining service and non-service groups *etceteras*.

- If approached in this way, the implementation of the second-best approach will mark an improvement in the final model's holistic representation of the regional economy.

- The important phase of model evaluation, (important coefficient selection *etc.*) sits more naturally within the process as a whole: there is no debate as to where these methods belong.

- Finally, the model outputs appear only after the full estimation process has been completed. The 'loop-hole' in GRIT, where pure nonsurvey tables could be produced with respectability, is tightened. Although one cannot prevent 'model sharks', the new procedure guards against them: and this it achieves by maintaining a logical preference order of approach to estimation, and keeping third-best methods firmly out in the cold.

9.2 Recommendations for Future Research

It is therefore recommended that, following this research:

(*i*) The input-output model should be marketed as an evolving economic database which affords the benefits associated with a co-ordinated and rationalised approach to collection, assembly and analysis of local economic information.

(*ii*) The input-output project should follow the first-to-second best method of estimation implied by this study.

(*iii*) The model should be marketed with a view to future extension. In particular, given the importance of demography within the regional economy and its poor representation within the basic Leontief specification, the model should be extended to this effect once the basic interindustry framework is in place.

(iv) Predictions on the behaviour of error functions, and the response of equation parameters to diminishing regional size have been made within this study. If sample data for Devon and Cornwall is generated, then this would offer an opportunity to assess the generality of these results.

References

Armstrong, H., (1993) The Local Income and Employment Impact of Lancaster University, *Urban Studies*, **30**, pp1653-68

Aurousseau, (1921) The Distribution of Population: A Constructive Problem, *Geography Review*, **11**, pp563-92

Bacharach, M., (1970) *Bi-proportional Matrices and Input-Output Change*, CUP, Cambridge

Batey, P., (1991) Household Structure and Extended Input-Output Models: a New Formulation, in Dewhurst, Hewings, Jensen (eds.), *Flexibility and Extensions in Input-Output Models*, Avebury Press, Aldershot

Batey, P., Madden, M., (1981) Demographic-Economic Forecasting Within an Activity-Commodity Framework: Some Theoretical Considerations and Empirical Results, *Environment and Planning A*, **13**, pp1067-83

Batey, P., Madden, M., Scholefield, G., (1993) Socio-economic Impact Assessment of Large-scale Projects Using Input-Output Analysis: A Case Study of an Airport, *Regional Studies*, **27**, pp179-91

Berenson, M., Levine D., (1986) *Basic Business Statistics: Concepts and Applications*, Prentice-Hall, New Jersey

Beyers, W., (1976) Empirical Identification of Key Sectors: Some Further Evidence, *Environment and Planning A*, **17**, pp73-99

Bishop, K., Simpson, C., (1972) Components of Change Analysis: Problems of Alternative Approaches to Industrial Structure, *Regional Studies*, **6**

Bishop, P., (1996) Buyer Supplier Linkages in the Defence Industry: the Case of Deveonport Dockyard, *Area*, **28** pp78-88

Black, (1981) Injection leakages, trade repercussions and the regional income multiplier. *Scottish Journal of Political Economy*, **28**, pp227-235

Boomsma, P., Oostehaven, J., (1992) A Double Entry Method for the Construction of Biregional Input-Output Tables, *Journal of Regional Science*, **32**, pp269-84

Borque, P., *et al.* (1967) *The Washington Input-Output Study*, University of Washington

Brand, S., (1997) On the Appropriate Use of Location Quotients in Generating Regional Input-Output Tables: a Comment, *Regional Studies*, **31**, pp791-4

Bromley, D., (1972) An Alternative to Input-Output Models: a Methodological Hypothesis, *Land Economics*, **48**, pp125-33

Brown, A., *et al.* (1967) Appendix to the Green Paper on the Development of Areas: Regional Multipliers, *National Institute Economic Review*, **51**

Brown, D., Giarratani, F., (1979) Input-Output as a Simple Econometric Model: A Comment, *Review of Economics and Statistics*, **61**, pp621-23

Bullard, C., Sebald, A., (1977) Effects of Parametric Uncertainty and Technological Change in Input-Output Models, *Review of Economics and Statistics*, **59**, pp75-81

Bulmer-Thomas, V., (1982) *Input-Output in Developing Countries*, Wiley, New York

Burford, R., Katz J., (1977) Regional Input-Output Multipliers Without a Full Input-Output Table, *Annals of Regional Science*, **9**, pp21-38

Burford, R., Katz, J. (1978) Regional Input-Output Multipliers Without a Full Input-Output Table: Reply, *Annals of Regional Science*, **12**, pp99-102, 105-6

Burford, R., Katz, J. (1981) A Method for Estimation of Input-Output-Type Multipliers When No Input-Output Model Exists, *Journal of Regional Science*, **21**, pp151-61

Burford R., Katz, J. (1985) Short-cut Formulas for Output, Income and Employment Multipliers, *Annals of Regional Science*, **19**, pp61-76

Business Strategies Limited, (1997) *Future Skills Needs in Gwent*, Appendix D, Gwent TEC

Butterford, M., Mules, T., (1980) A Testing Routine for Evaluating the Cell-by-Cell Accuracy in Short-Cut Regional Input-Output Tables, *Journal of Regional Science*, **20**, pp293-310

Cambridge Econometrics, (1993) *Regional Economic Prospects: Cambridge Econometrics*, Cambridge

Chenery, H., and Watanabe, T., (1958) International Structures of Production, *Econometrica*, **26**, pp487-521

Conway, R., (1980) Changes in Regional Input-Output Coefficients and Regional Forecasting, *Regional Science and Urban Economics*, **10**, pp153-71

Conway, R., (1991) An Empirical Comparison of Regional Multipliers, in Dewhurst, Hewings, Jensen (eds.), *Flexibility and Extensions in Input-Output Models*, Avebury Press, Aldershot

Czamanski, S., Malizia, E., (1969) Applicability and Limitations in the Use of National Input-Output Tables for Regional Studies, *Papers of the Regional Science Association*, **23**, pp 65-77

D'Augostino, R., Stevens, M., (eds.). (1986). *Goodness-of-Fit Techniques*, Marcel Dekker.

DCDI, (1996), *An Introduction to Devon and Cornwall*, November

De Kanter, J., Morrison, W., (1978) The Merseyside Input-Output Study and its Applications in Structure Planning, in Batey (ed.) *Theory and Method in Urban and Regional Analysis*, Pion, London, pp120-49

Dewhurst, J. Ll., (1992) Using the RAS Technique as a Test of Hybrid Methods of Regional Input-Output Table Updating, *Regional Studies*, **26**, pp81-91

Dewhurst, J. Ll., (1993) Decomposition of Changes in Input-Output Tables, *Economic Systems Research*, **5**, pp41-53

Dewhurst, J. Ll., West, G., (1991) Conjoining Regional and Inter-regional Input-Output Models with Econometric Models, in Dewhurst, Hewings, Jensen (eds.), *Flexibility and Extensions in Input-Output Models*, Avebury Press, Aldershot

Drake, R., (1976) A Short-cut to Estimates of Regional Input-Output Multipliers: Methodology and Evaluation, *International Regional Science Review*, **1**, pp1-17

Eskelinen, H., Suorsa, M., (1980) A Note on Estimating Interindustry Flows, *Journal of Regional Science*, **20**, pp261-6

Evans, W., (1954) The Effect of Structural Matrix Errors on Interindustry Relations Estimates, *Econometrica*, **22**, pp461-80

Flegg A., Webber C., and Elliot M., (1995) On the Appropriate Use of Location Quotients in Generating Regional Input-Output Tables, *Regional Studies*, **29**, 547-561

Fraser of Allender, (1978) *Scottish Input-Output Tables for 1978*, Fraser of Allender Institute, Glasgow

Garhart, R. (1985) The Role of Error Structure in Simulations of Regional Input-Output Analysis, *Journal of Regional Science*, **25**, pp353-66

Garhart, R., Moloney, R., O'Leary, E., Donnellan, T., (1996) An Input-Output Model of Cork and Kerry: A Preliminary Report, *Paper to the 27th Annual Conference of the RSAI British and Irish Section*, Edinburgh, September

Geary, R., (1973) A Method of Estimating the Elements of an Interindustry Matrix Knowing the Row and Column Totals, *Economic and Social Review*, **4**, pp477-85

Gerking, S., (1976) Input-Output as a Simple Econometric Model, *Review of Economics and Statistics*, **58**, pp274-82

Gerking, S., (1979) Reconciling Reconciliation Procedures in Regional Input-Output Analysis, *International Regional Science Review*, **4**, pp23-36

Ghosh, A., (1958) Input-Output Approach to an Allocative System, *Econometrica*, **25**, pp58-64

Giarratani, F., Garhart, R., (1991) Simulation Techniques in the Evaluation of Regional Input-Output Models: a Survey, in Dewhurst, Hewings, Jensen (eds.) *Flexibility and Extensions in Input-Output Models*, Avebury Press, Aldershot

Greenstreet, D., (1989) A conceptual framework for construction of hybrid regional input-output models, *Socio-Economic Planning Sciences*, **23**, pp283-89

Greytak, D., (1969) A Statistical Analysis of Regional Export Estimating Techniques, *Journal of Regional Science*, **9**, pp387-95

Gripaios, P., Gripaios, R., Trinder, R., Trinder, D., (1991) Exporting from the South West, *South West Economy Research Papers*

Hanseman, D., (1982) Stochastic Input-Output Analysis: a Simulation Study, *Environment and Planning A*, **14**, pp1425-35

Hanseman, D., Gustafson, E., (1981) Stochastic Input-Output Analysis: a Comment, *Review of Economics and Statistics*, **63**, pp468-70

Harrigan, F., McGilvray, J., McNicoll I., (1980a) A comparison of national and regional technical structures, *Economic Journal*, 90, pp795-810

Harrigan, F., McGilvray, J., McNicoll, I., (1980b) Simulating the Structure of a Regional Economy, *Environment and Planning A*, **12**, pp927-36

Harrigan, F., McGregor, P., Dourmashkin, N., Parman, R., Swales, JK., Yin, Y., (1991b) AMOS: A Micro-macro Model of Scotland, *Economic Modelling*, **8**, pp424-79

Harrigan, F., McGregor, P., Dourmashkin, N., Swales JK., Yin, Y., (1991a) The Sensitivity of Output Multipliers to Alternative Factor Market Assumptions: a Computable General Equilibrium Analysis, in Dewhurst, Hewings, Jensen (eds.), *Flexibility and Extensions in Input-Output Models*, Avebury Press, Aldershot

Hazari, B., (1970) Empirical Identification of Key Sectors in the Indian Economy, *Review of Economics and Statistics*, **52**, pp301-05

Hewings, G., (1971) Regional Input-Output models in the UK: Some Problems and Prospects for the Use of Nonsurvey Techniques, *Regional Studies*, **5**, pp11-22

Hewings, G., (1974) The Effect of Aggregation on the Empirical Identification of Key Sectors in a Regional Economy: A Partial Evaluation of Alternative Techniques, *Environment and Planning A*, **6**, pp439-53

Hewings, G., (1977) Evaluating the Possibilities for Exchanging Regional Input-Output Coefficients, *Environment and Planning A*, **9**, pp927-44

Hewings, G., (1986) Problems of Integration in the Modelling of Regional Systems, in Batey and Madden (eds.), *Integrated Analysis of Regional Systems*, Pion, London, pp37-53

Hewings, G., (1996) communication

Hewings, G., Jensen, R., (1985) Short-cut 'Input-Output' Multipliers: a Requiem, *Environment and Planning A*, **17**, pp747-59

Hewings, G., Jensen, R., (1986) Regional, Interregional and Multiregional Input-Output Analysis, in Nijkamp (ed.) *Handbook of Regional and Urban Economics Volume I*

Hewings, G., Romanos, M., (1981) Simulating Less Developed Regional Economies Under Conditions of Limited Information, *Geographical Analysis*, **13**, pp373-90

Hill, S., Munday, M., (1994) *The Regional Distribution of Foreign Manufacturing Investment in the UK*, MacMillan, London

Hill, S., Roberts, A., (1996) *Welsh Input-Output Tables for 1994*, University of Wales Press, Cardiff

Hirschman, A., (1958) *The Strategy of Economic Development*, Yale

Ireson, R., Tomkis, C., (1978) *Inter-regional Input-Output Tables for Wales and the Rest of the UK*, Welsh Council, Cardiff

Isard W., (1960) *Methods of Regional Analysis: An Introduction to Regional Science*, MIT Press, Massachusetts

Isard, W., (1953) Regional Commodity Balances and Interregional Commodity Flows, *American Economic Review*, **43**, pp167-80

Isard, W., (1956) *Location and Space Economy*, Wiley, New York

Isard, W., Keunne, R., (1953) The Impact of Steel Upon the Greater New York-Philadelphia Urban Industrial Region, *Review of Economics and Statistics*, **35**, pp293-301

Isard, W., Romanoff, E., (1968), *The Printing and Publishing Industries of Boston SMSA, 1963: and Comparison with the Corresponding Philadelphia Industries*, Regional Science Research Institute, Massachusetts

Israilevich, P., (1986) *Biproportional forecasting of input-output tables*. Ph.D. dissertation. Philadelphia: University of Pennsylvania

Israilevich, P., Hewings, G., Schindler, G., Mahidhara, R., (1996) The Choice of an Input-Output Table Embedded in Regional Econometric Input-Output Models, *Papers in Regional Science*, **75**, pp103-120

Isserman, A., (1977) The Location Quotient Approach to Measuring Regional Economic Impacts, *Journal of American Institute of Planners*, **43**, pp33-41

Isserman, A., (1980) Estimating Export Activity in a Regional Economy: A Theoretical and Empirical Analysis of Alternative Methods, *International Regional Science Review*, **5**, pp155-84

Jackson, R., (1986) The Full-Distribution Approach to Aggregate Representation in the Input-Output Modelling Framework, *Journal of Regional Science*, **26**, pp515-32

Jackson, R., (1991) The Relative Importance of Input Coefficients and Transactions in Input-Output Structure, in Dewhurst, Hewings, Jensen (eds.), *Flexibility and Extensions in Input-Output Models*, Avebury Press, Aldershot

Jensen, R., (1980) The Concept of Accuracy in Regional Input-Output Models, *International Regional Science Review*, **5**, pp139-54

Jensen, R., (1990) Construction and use of regional input-output models: progress and prospects, *International Regional Science Review*, **13**, pp9-25

Jensen, R., Dewhurst, J Ll., West, G., Hewings, G., (1991) On the Concept of Fundamental Economic Structure, in Dewhurst, Hewings, Jensen (eds.) *Flexibility and Extensions in Input-Output Models*, Avebury Press, Aldershot

Jensen, R., MacDonald, S., (1982) Technique and Technology in Regional Input-Output, *Annals of Regional Science*, **16**, pp27-45

Jensen, R., Mandeville, T., Karunaratne, N., (1979) *Regional Economic Planning: Generation of Regional Input-Output Analysis*, Croom Helm, London

Jensen, R., McGuarr, D., (1977) Reconciliation Techniques in Input-Output Analysis: Some Comparisons and Implications, *Urban Studies*, **14**, pp327-37

Jensen, R., West, G., (1980) The Effect of Relative Coefficient Size on Input-Output Multipliers, *Environment and Planning A*, **12**, pp659-60

Jensen, R., West, G., Hewings, G., (1988) The Study of Regional Economic Structure Using Input-Output Tables, *Regional Studies*, **22**, pp209-20

Jilek, J., (1971) The Selection of Most Important Coefficients, *Economic Bulletin for Europe*, **23**, pp86-105

Johns, P., Leat, M., (1986) *An Input-Output Analysis of the Cornish Economy*, Report to Cornwall County Council Planning Department, North of Scotland College of Agriculture, Aberdeen

Johnston, R., McKinney, M., Stark, T., (1996) Regional Price Level Variations and Real Household Incomes in the United Kingdom, 1979/80-1993, *Regional Studies*, **30**, pp567-578

Jones, L., (1976) The Measurement of Hirchmanian Linkages, *Quarterly Journal of Economics*, **90**, pp323-33

Keurre, J., Weller, B., (1989) Forecasting the Local Economy Using Time-Series and Shift-Share Techniques, *Environment and Planning A*, pp753-70

Keynes, J., (1936) *The General Theory of Employment, Interest and Money*, MacMillan, London.

Knudsen, D., Fotheringham, A., (1986) Matrix Comparison, Goodness-of-Fit, and Spatial Interaction Modelling, *International Regional Science Review*, **10**, pp.127-148

Kokat, R., (1966) Some Conceptual Problems in the Implementation of the Maryland Interindustry Study, *Papers of the Regional Science Association*, **17**, pp95-104

Kullback, S., Leibler, R., (1951) On Information and Sufficiency, *Annals of Mathematical Statistics*, **22**, pp78-86

Lahr, M., (1993) A Review of the Literature Supporting the Hybrid Approach to Constructing Regional Input-Output Models, *Economics Systems Research*, **5**, pp277-93

Leontief, W., (1936) Quantitative Input and Output Relationships in the Economic System of the United States, *Review of Economics and Statistics*, **18**, pp105-125

Leontief, W., The Dynamic Inverse, in Carter, Brody (eds.), *Contributions to Input-Output Analysis*, Amsterdam

Lui, A., Grainger, J., Jaffry, S., (1995) Alternative Approaches for Modelling a Regional Economy in an Input-Output Framework, *Paper presented to the ESRC Urban and Regional Economics Conference*, Ambleside, Lancaster

Madden, M., Batey, P., (1986) A Demographic-Economic Model of a Metropolis, in Woods, Rees (eds.), *Population Structures and Models*, Allen and Unwin, London

Madden, M., Batey, P., Thomson, J., (1996) Decomposing Input-Output Tables to Establish the Effects of Temporal Changes in Demographic and Economic Characteristics, *Economic Systems Research*, **8**, pp209-225

Martin, R., (1997) Regional Unemployment Disparities and their Dynamics, *Regional Studies*, **31**, pp237-252

Mayer, W., Pleeter, S., (1975) A Theoretical Justification for the Use of Location Quotients, *Regional Science and Urban Economics*, **5**, pp343-55

McCann, P., Dewhurst, J. Ll., (1996) Regional Size, Industrial Location and Input-Output Expenditure Coefficients, *Paper to the 27th Annual Conference of the RSAI British and Irish Section*, Edinburgh, September

McDowall, S., (1973) *Discussion paper on the results of the Input-Output study of Sutherland for 1970*, University of St. Andrews, Edinburgh

McGilvray, J., (1977) Linkages, Key Sectors and Development Strategy, in Leontief (ed.) *Structure, System and Economic Policy*, Cambridge

McGregor, P., McVittie, E., Swales, JK., Yin, Y., (1996) A CGE Analysis of the Employment Effects of Regional Capital Subsidies, *Paper to the 27th Annual Conference of the RSAI British and Irish Section*, Edinburgh, September

McMenamin, R., Haring, J., (1974) An appraisal of nonsurvey techniques for estimating regional input-output models, *Journal of Regional Science*, **14**, pp191-205

McNicoll, I., (1976) *The Shetland Economy*, Fraser of Allender Research Monograph No. 2

McNicoll, I., (1982) Ex-Post Appraisal of an Input-Output Forecast, *Urban Studies*, **19**, pp397-404

Miernyk, W., (1976) Comments on Recent Developments in Regional Input-Output Analysis, *International Regional Science Review*, **1**, pp47-55

Miernyk, W., *et al.* (1967) *Impact of the Space Program on A Local Economy*, West Virginia University Press

Miller, R., (1957) The Impact of the Aluminium Industry on the Pacific Northwest: A Regional Input-Output Analysis, *Review of Economics and Statistics*, **39**, pp200-209

Miller, R., Blair, P., (1985) *Input-Output Analysis*, Prentice-Hall, New Jersey

Moore, F., Peterson, J., (1955) Regional Analysis: An Interindustry Model of Utah, *Review of Economics and Statistics*, **37**, pp368-83

Morrison, W., (1973) The development of an urban interindustry model. *Environment and Planning A*, **5**, pp369-383, 433-460, 545-554

Morrison, W., Smith, P., (1974) Nonsurvey Input-Output Techniques at the Small Area Level: an Evaluation, *Journal of Regional Science*, **14**, pp1-14

Morrison, W., Smith, P., (1977) Input-Output Methods in Urban and Regional Planning: A Practical Guide, *Progress in Planning*, **7**, pp 59-151

Moses, L., (1955) On the Stability of Interregional Trading Patterns and Input-Output Analysis, *American Economic Review*, **45**, pp802-3

Moses, L., (1958) Location and the Theory of Production, *Quarterly Journal of Economics*, **72**, pp259-72

Nevin, E., Round, J., Roe, A., (1966) *The Structure of the Welsh Economy*, University of Wales Press, Cardiff

Norcliffe, G., (1983) Using Location Quotients to Estimate the Economic Base and Trade Flows, *Regional Studies*, **17**, pp161-8

ONS, (1978) *1973 UK Input-Output Tables*, ONS

ONS, (1992) Input-Output Balance for the United Kingdom, 1989, *Economic Trends*, No. 467, September

ONS, (1994) *Input-Output Tables for the United Kingdom, 1990*, HMSO, London

Oppenheim, N., (1992) *Questionnaire Design, Interviews and Attitude Measurement*, Pinter, London

Park, S., Mohtadi, M., Kubursi, A., (1981) Errors in Regional Nonsurvey Input-Output Models: Analytical Simulation Results, *Journal of Regional Science*, **21**, pp321-37

Phibbs P., Holsman, A., (1982) Estimating Input-Output Multipliers - a New Hybrid Approach, *Environment and Planning A*, **14**, pp335-42

Phibbs, P., Holsman, A., (1980) A Short-Cut Method for Computing Final Demand Multipliers for Small Regions, *Environment and Planning A*, **12**, pp1001-8

Pindyck, R., Rubenfeld, D., (1976) *Econometric Models and Economic Forecasts*, McGraw-Hill, London

Proops, J., Pullen, M., Booth, I., Fishman, L., (1981) *An Input-Output Study of North Staffordshire*, University of Keele

Quandt, R., (1958) Probabilistic Errors in the Leontief System, *Naval Research Logistics Quarterly*, **5**, pp155-70

Quandt, R., (1959) On the Solution of Probabilistic Leontief Systems, *Journal of Policy Modelling*, **10**, pp179-94

Ralston, S., Hastings, S., Brucker, S., (1986) Improving Regional Input-Output Models: Evidence Against Uniform Regional Purchase Coefficients Across Rows, *Annals of Regional Science*, **20**, pp65-80

Rasmussen, P., (1957) *Studies in Intersectoral Relations*, Amsterdam

Richardson, H., (1985) Input-Output and Economic Base Multipliers: Looking Backward and Forward, *Journal of Regional Science*, **25**, pp607-661

Roberts, A., (1996) *The Impact of Foreign Manufacturing Investment on Wales*, PhD Thesis, University of Wales

Round, J., (1972) Regional Input-Output Models in the UK: A Reappraisal of Some Techniques, *Regional Studies*, **6**, pp 1-9

Round, J., (1978) An Interregional Input-Output Approach to the Evaluation of Nonsurvey Methods, *Journal of Regional Science*, **18**, pp 179-194

Round, J., (1983) Nonsurvey Techniques: A Critical Review of the Theory and the Evidence, *International Regional Science Review*, **8**, pp189-212

Round, J., (1986) Social Accounting for Regional Economic Systems, in Batey and Madden (eds.) *Integrated Analysis of Regional Systems*, Pion, London, pp90-106

Sawyer, C., Miller, R., (1983) Experiments in Regionalisation of a National Input-Output Table, *Environment and Planning A*, **15**, pp1501-20

Schaffer, W., (1976) *On the Use of Input-Output Models for Regional Planning, Studies in Applied Regional Science*, Vol. 1, Martinus Nijhoff, Netherlands

Schaffer, W., Chu, K., (1969) Nonsurvey techniques for constructing regional interindustry models, *Papers of the Regional Science Association*, **23**, pp 83-101

Schintke, J., Stäglin, R., (1988) Important Input Costs in Market Transaction Tables and Production Flows Tables, in Ciaschini (ed.) *Input-Output Analysis: Current Developments*, Chapman and Hall, London

Schultz, S., (1976) Intersectoral Comparisons as an Approach to the Identification of Key Sectors, in Polenske, Skolka (eds.) *Advances in Regional Input-Output Analysis*, Ballinger, Massachusetts

Scottish Office, (1994) *Scottish Input-Output Tables for 1989*, HMSO, Edinburgh

Scottish Office, (1995) An Extended Input-Output System for Analysis of the Scottish Economy, EAS Research Paper, No. 39, December

Scottish Office, (1996) *Input-Output Balances for Scotland, 1993*, HMSO, Edinburgh

Shen, T., (1960) An Input-Output Table with Regional Weights, *Papers of the Regional Science Association*, **6**, pp113-19

Sherman, J., Morrison, W., (1950) Adjustment of an Inverse Matrix Corresponding to a Change in One Element of a Given Matrix, *Annals of Mathematical Statistics*, **21**, pp124-7

Simpson, D., Tsukui, J., (1965) The Fundamental Economic Structure of Input-Output Tables : An International Comparison, *Review of Economics and Statistics*, **47**, pp434-46

Smith, P., Morrison, W., (1974) *Simulating the urban economy, experiments with input-output techniques*, Pion, London.

Sonis, M., Hewings, G. (1992) Coefficient Change in Input-Output Models: Theory and Applications, *Economic Systems Research*, **4**, pp143-157

Sonis, M., Hewings, G., (1989) Error and Sensitivity Analysis: a New Approach, in Miller, Polenske, Rose (eds.) *Frontiers of Input-Output Analysis*, OUP, Oxford

Sonis, M., Hewings, G., (1991) Fields of Influence and Extended Input-Output Analysis: a Theoretical Account, in Dewhurst, Hewings, Jensen (eds.) *Flexibility and Extensions in Input-Output Models*, Avebury Press, Aldershot

Steele, D., (1972) A Numbers Game (or The Return of Regional Multipliers), *Regional Studies*, **6**, pp 115-130

Stevens B., Treyz G., Lahr M., (1989) On the Comparative Accuracy of RPC Estimating Techniques, in Miller, Polenske, Rose (eds.) *Frontiers of Input-Output Analysis*, OUP, Oxford

Stevens, B., Trainer, G., (1976) Error Generation in Regional Input-Output Analysis and its Implications for Nonsurvey Models, in Pleeter (ed.), *Economic Impact Analysis: Methodology and Applications*, Boston

Stevens, B., Trainer, G., (1980) The Generation of Errors in Regional Input-Output Models, Regional Science Research Institute, Working Paper A1-76

Stevens, B., Treyz, G., Ehrlich, D., Bower, J., (1983) A New Technique for the Construction of Nonsurvey Regional Input-Output Models and Comparisons with Two Survey-Based Models, *International Regional Science Review*, **8**, pp271-86

Stone, J., (1961) *Input-Output and National Accounts*, Organisation for European Economic Co-operation, Paris

Szyrmer, J., (1989) *Trade-Off Between Error and Information in the RAS Procedure*, in Miller, Polenske, Rose (eds.), *Frontiers of Input-Output Analysis*, Oxford University Press, Oxford, pp.258-278

Szyrmer, J., (1992) Input-Output Coefficients and Multipliers for a Total Flow Perspective, *Environment and Planning A*, **24**, pp921-37

Theil, H., (1971) *Principles of Econometrics*, Wiley, New York

Tiebout, C., (1967) Input-Output and the Firm: A Technique for Using National and Regional Tables, *Review of Economics and Statistics*, **49**, pp260-62

Turok, I., (1993) Inward Investment and Local Linkages: How Deeply Embedded is 'Silicon Glen'?, *Regional Studies*, **27**, pp 401-17

Wagstaff, H., (1973) Employment multipliers in Scotland, *Scottish Journal of Political Economy*, **20**, pp239-262

Walras, L., (1874) *Elements of Pure Economics*, Jaffé (transl., 1954), Homewood, Illinois

Weber, A., (1929) *Theory of Location Industries*, Frederick (transl.), University of Chicago Press, Chicago

West, G., (1980) Generation of regional input-output tables (GRIT): an Introspection, *Economic Analysis and Policy*, **10**, pp71-86

West, G., (1981) An Efficient Approach to the Estimation of Regional Input-Output Multipliers, *Environment and Planning A*, **13**, pp857-67

West, G., (1982) Sensitivity and Key Sector Analysis in Input-Output Models, *Australian Economic Papers*, **21**, pp365-78

West, G., (1986) A Stochastic Analysis of an Input-Output Model, *Econometrica*, **54**, pp363-74

West, G., (1990) Regional Trade Estimation: A Hybrid Approach, *International Regional Science Review*, **13**, pp103-18

West, G., (1993) *Input-Output Analysis for Practitioners*, University of Queensland, Australia

West, G., (1994) A Comparison of Input-Output, Integrated and CGE Impact Models at the Regional Level, *Paper to the 25th Annual Conference of the RSAI British and Irish Section*, Dublin, September

West, G., Jensen, R., (1980) Some Reflections on Input-Output Multipliers, *Annals of Regional Science*, **14**, pp77-89

Willis, K., (1987), Spatially disaggregated input-output tables: an evaluation and comparison of survey and nonsurvey results, *Environment and Planning A*, **19**, pp107-116

Wilson, T., (1968) The Regional Multiplier: A Critique, *Oxford Economic Papers,* **20** pp374-393.

Xu, S., Madden, M., (1991) The Concept of Important Coefficients in Input-Output Models, in Dewhurst, Hewings, Jensen (eds.), *Flexibility and Extensions in Input-Output Models*, Avebury Press, Aldershot