MULTI-REGIONAL FREIGHT GENERATION MODELS

by

A. BENHEDDI

A Doctoral thesis
Submitted in partial fulfilment of the requirements of the award of Doctor of Philosophy of the Loughborough University of Technology

Supervisor: Professor N.J.Ashford,
Professor of Transport Planning

" FOR MY PARENTS "

SUMMARY

This study is concerned with the development of models for total and commodity freight by road and by rail originating from and terminating into the different regions of the U.K., as well as from and into 134 composite zones of these regions.

Previous work is reviewed and improvements in freight generation modelling are sought in providing a conceptual framework, investigating a wider range of variables than hitherto and in tackling some statistical shortcomings in previous studies viz. the violation of assumptions that underly the use of ordinary least squares method and the effects of spatial and industrial aggregation.

The main techniques used are regression analysis, principal component analysis and covariance analysis. The accessibility variables, however, are derived from the calibration of the group locational surplus maximizing model.

The models are critically discussed and an attempt to relate the results to location theory is made. Subsequently, the potential for future research and some planning implications are outlined.

A C K N O W L E D GMENTS

I would like to thank Professor D.J. Johns, my director of research, for his support.

Professor N.J. Ashford, my supervisor, provided an invaluable stimulus through his suggestions and his encouraging approach. I wish to record with gratitude, in particular, his suggestions on the organisation of my thesis, as well as, his suggestions of techniques that made me aware of the "arsenal" of multivariate analysis.

My research has been continuously actively encouraged by Dr. D.E. Pitfield. I am greatly indebted to him, for his help in setting about my work and for his most appreciated advice.

In addition, I would like to thank all the staff from the Department of Transport Technology for their very fine cooperation.

Finally, I would like to thank my parents and friends, the Warners, for their moral support. Special thanks to my brother, Zemri, for his encouragements throughout my studies.

CONTENTS

CONTENTS

CONTENTS (Continued)

CONTENTS (Continued)

LIST OF TABLES

The tables are numbered according to chapters (first number) and to the order of appearance in the text (second number)

LIST OF TABLES (Continued)

LIST OF TABLES (Continued)

LIST OF TABLES (Continued)

LIST OF TABLES (Continued)

LIST OF FIGURES

As for the tables, the first number represents the chapter and the second number represents the order of appearance of the figure in the text.

L I S T OF A P P E N D I C E S

MAIN ABBREVIATIONS

S.I.C. Orders	=	these are broad industrial groups in the Standard Industrial Classification for the U.K. (See Appendix II.1).
M.L.H. Subdivisions	=	Minimum List Heading, these are subgroups of the S.I.C. orders.
S.N.D.(I)	=	Standard Normal Deviate for the Generalized Moran I Statistic.
O_i, D_j	=	respectively originating and terminating freight.

VARIABLES AND UNITS OF MEASUREMENT

I Freight Data (1972)

In all models the freight data is in thousand tons

II Socio-Economic Data *

Accessibility	Miles
Agriculture area	thousand acres
Cement deliveries to regions	thousand tons
Coal consumption	thousand tons
Distance	miles
Employment	thousands
Gross domestic profits & surpluses	£ thousand
Gross domestic product	£ thousand
Income from employment	£ thousand
Net capital expenditure	£ thousand
Net output	£ thousand
Number of goods vehicles	number
Petroleum consumption	thousand tons
Production of iron	thousand tonnes
Production of steel	thousand tonnes
Public investment	£ thousand
Population	thousands
Region area	km^2
Retail turnover	£ thousand
Zone area	km^2

* Where the data are time dependent, they refer to 1972 except for employment at the zonal level which is for 1971.

CHAPTER 1

INTRODUCTION

1.1 General Introduction

The explanation and forecast of transport demands through the use of models is justified by the need to set resource planning and allocation on a firmer basis. Although a great deal has been accomplished at the urban level, in particular on passenger demand models, only a limited number of studies have looked at freight shipments. Of these, only two in the U.K., have concerned themselves with nationwide multi-regional shipments of goods (Department of the Environment, 1971; Chisholm and O'Sullivan, 1973). This lack of research is primarily due to the fact that before the 1962 Road Goods Survey was completed, little information on origin/destination[1] of freight within Great Britain was available. The 1972 British Rail freight data, used in this thesis has enabled further research in freight distribution (Pitfield, 1978a, 1978b; Benabi, 1978) and additional work on generation modelling. This area of generation modelling is the central concern in this research.

Generation modelling consists of the development of explanatory models which present causal

---

(1) O_i , D_j thereafter.

relationships between freight flow and independant exogeneous variables. The calibration of such models requires, in addition to the availability of freight data, a certain amount of compatible economic information. It is probably this additional data requirement that has resulted in the relative neglect of generation models compared to distribution models in previous studies.

The potential usefulness of these models was the motive for undertaking this research which required an associated extensive data collection exercise. This utility may be judged from the role of generation models in the development of freight forecasting models. For example, in the distribution phase, the O_i and D_j variables are given by data if the base year is considered whilst in a forecasting situation they are derived from the generation models. The calibration of generation models also provides a framework within which some of the assumptions inherent in interaction models may be tested[(1)]. Furthermore, the forecasts of freight volumes by different modes and at different places

---

(1) For example interaction models used in the distribution phase may assume a closed system. Therefore in previous interregional freight distribution studies the question of whether the British economy may be regarded as closed has been raised and answered formally prior to the modelling exercise. Such an assumption may be checked statistically in the generation phase.

may provide valuable information for the appraisal of projected transport investments. Finally, because generation models seek a more precise and quantified understanding of the maxim "traffic is a function of activity", due regard may be given to the traffic implications of industrial linkages and regional policies. This potential usefulness of freight generation models is subject to the ability to calibrate successfully, models which give a good description of existing freight patterns. However, there are methodological considerations with respect to the calibration process that involve three important issues.

(i) The most accessible data is usually aggregated information. Unfortunately, over time, changes in the internal structure of aggregates may take place, making projections unreliable.

(ii) Even when data is disaggregated there is little evidence as to the stability of the models over time.

(iii) If the models are to be of any use then truly causal relationships must be specified. More often however, analysts have selected variables for their availability and projectability and models have been associative rather causal.

Thus the objectives of further study should be devised with a view to the variety of possible uses of the models and the underlying difficulties.

1.2 Objectives of the Study

The first aim of this thesis is to review all prior reported work at both the urban and regional levels. The major points of interest in previous studies are isolated and areas requiring more attention are underlined.

Because a major weakness in previous work is the limited number of explanatory variables examined, it was thought essential to devote some time to gathering socio-economic data to supplement the 1972 British Rail freight data.

In addition it is intended to investigate the statistical problems of using spatial data. In particular, the assumptions involved in the use of ordinary least squares method are examined and the effect of aggregation evaluated.

The study is also concerned with some understanding of the workings of regional economies. This area is examined in an attempt to describe spatial and industrial linkages, the use of locational variables and tests of locational hypothesis.

1.3 Plan of the Thesis

The thesis is divided into four parts.

Part one includes this introduction, a review of previous work and a description of the data and of the main techniques to be used.

Part two studies the statistical analysis carried out at the regional level for total freight and looks at the commodity breakdown of freight.

Part three describes experiments in modelling total and commodity freight at the zonal level. Emphasis is given, at this level, to improving the models by incorporating an accessibility variable and the corrections to the usual violation of some of the ordinary least squares assumptions.

Part four draws together comparisons of the regional and zonal results and the conclusions of the research. Subsequently, it discusses areas with potential for future inquiry and planning implications of the findings.

CHAPTER 2

REVIEW OF PREVIOUS WORK

2.1 Scope of Review

Existing freight studies may be divided into two groups. The first group concerns itself with urban goods movement and the second with multi-regional freight movement. Because both types of studies are often potentially faced with similar problems, studies of the second type (as this thesis) may benefit from the findings of urban goods movement studies to which more attention has been devoted. Initially, previous work is critically reviewed; subsequently areas requiring further research are outlined and some implications for further work are drawn.

2.2 Discussion of Previous Work

Since the concern of this thesis is to model freight generation, the various models that have been previously used for this purpose are of interest. These vary in terms of the variables they include and in terms of the mathematical form they take. In addition, the modelling process is closely tied to the condition of the data; in particular the level of aggregation adopted affects the results. As a consequence this discussion centres around three issues viz. the level of aggregation, the choice of variables and the mathematical form of the relationships. At the outset it is appropriate to see how previous researchers have dealt with these particular points.

(i) Aggregation

There are at least four dimensions over which freight may be aggregated namely time, space, commodities and modes.

Time: In existing models, demands are expressed as rate per day, week or year. Although in theory it is possible to imagine time as a separate variable, according to Roberts (1977), this is a complication that is difficult to handle in existing models. Nevertheless, if seasonality of flows is important, as in some developing countries and for some commodity groups, separate models ought to be developed. In both types of studies reviewed, time has been dealt with at the data collection phase only. The main concern has been with the selection of an appropriate period for recording traffic. In establishment based studies it has been found that the appropriate duration should vary with the size of the plant; a shorter period is found to be more suitable for large premises (Redding, 1972; Watson, 1975). Surveys of long duration such as the two year survey from which Slavin (1976) took his data might be more informative. However, long duration surveys are costly. It should be noted that the survey period for the 1962 Road Goods Survey (on which the Department of the Environment (1971) and Chisholm and O'Sullivan's works (1973) were based) was four weeks "selected to represent seasonal patterns" Chisholm and O'Sullivan (1973).

Space: Over the last few years, zonal regressions have become discredited for two reasons: loss of information and loss of variation. The latter leads to spuriously inflated correlations. The bulk of the urban goods movement studies, at least those undertaken in the U.K., have considered the actual behavioural unit that is the establishment. These latter models avoid the problems of spatial aggregation. However, as with household regressions in passenger generation models, establishment based models may be criticised for the fact they do not incorporate areal variables which may have some bearing on traffic generated. For example, in a zonal based study, Slavin (1976) found that location indices were relevant and Loelb and Kenneth (1976) successfully used area specialisation measures.

Commodity[1] Aggregation: The great diversity of commodity attributes and the importance of such attributes for the selection of modes and for the mean haul travelled makes the choice of the nature of commodity

---

(1) An industry is defined by its principal product and accordingly the terms commodity and industry are used interchangeably.

aggregation a crucial one. Previous research at the plant level has established variations in generation rates over different industries and shown that higher disaggregation leads to more reliable models. Starkie (1967) found that the results for subsets of manufacturing categories exhibit substantial advantage over those for the aggregate manufacturing class. Leake and Gan (1973) reported that pooled results were statistically poor. Also, the subclasses of industrial groups were found to generate different trip rates. Redding (1972) found a similar variation. Maltby (1973) used covariance analysis to test whether similarities exist between different industries in generation rates. When similarities are found, no accuracy is lost by combining the classes. His analysis suggested a significant difference between heavy and non-heavy industries. At the multiregional level, the superiority of disaggregation is supported by Heyman's finding.[1] She found that regressions based on individual industries performed better than those based on industrial groupings.

Aggregation Over Modes: Given the marked differences between modal attributes, aggregation over modes has been the least used. Chisholm and O'Sullivan (1973) developed separate models for road and for rail and for the aggregate road plus rail. Rail models were generally poor whereas road yielded the best results and reasonable results were obtained for the aggregate

---

(1) See Department of the Environment (1971).

mode. In urban studies, stratification by vehicle type has been performed. Maltby (1973) distinguished "heavy commercial vehicles" and "light commercial vehicles". By means of canonical correlation analysis, Watson (1975) showed the relevance of such a stratification. He found that shipments may be adequately characterised by weight and volume in a combined "variate" and that "this variate correlates highly with truck type and capacity". Another related question is whether goods vehicle trips can be considered as a proxy for goods movement. If the answer is affirmative, then information on the type of shipment and the degree of utilisation of the vehicle is needed.

It appears, there are serious problems in attempting to derive relationships from aggregate data. In fact the underlying assumption is that of homogeneity (for example, between industries or between contiguous areas) thereby allowing them to be group... If aggregate data is used, then, it is appropriate to test whether such an assumption is violated and to evaluate the effects on the results if this is the case.

(ii) Choice of Variables

It is worth examining previous work with respect to the variables used and the relevance of variables overlooked. Freight has usually been related to factors describing the individual establishment or the zone. In urban studies employment and floorspace have often been used. Starkie (1967) found significant relationship between freight generation and both of these variables but suggested employment as the most suitable in terms of the criteria of stability, projectability and availability. Leake and Gray (1979) found that "variables relating to floor and site area did not figure quite so prominently as employment variables" in relationships derived from a stepwise regression on three types of variables viz. area variables, employment variables, vehicles operated variables. However, Leake and Gan's results were different. The number of employees proved to be unsatisfactory while floorspace and site area were found to be useful variables. Maltby concluded that the most satisfactory variable varied with the type of traffic. Watson found that floorspace performed less well than employment but component floorspace (e.g. production floorspace, storage floorspace) gave satisfactory results. Zonal based urban studies and multi-regional freight studies have considered employ- but not floorspace. Use of total employment and employment by industry proved useful (Slavin, 1976; Department of the Environment, 1971; Chisholm and O'Sullivan, 1973).

It would seem that if either floorspace or employment give statistically reliable models, then employment should be preferred because employment is more sensitive to changes in production than floorspace.

Other variables that were mostly components of floorspace and of employment have been examined (Maltby, 1973; Watson, 1975; Leake and Gray, 1979) as well as variables such as distance from a given location (Redding, 1972), credit availability (Watson, 1975). As the inclusion of these variables added very little to the significance of the relationship, a number of authors suggested that simple regression was most appropriate (Redding, 1972, G.L.C. and Metra Group Ltd., 1970; Maltby, 1973). While these authors pointed out the fact that one variable was dominant, the multiple regression exercise out at the multi-regional level were limited by severe multi-collinearity between employment measures across industries.

Previous research may be criticised for the fact it overlooked other relevant variables by limiting itself to employment and floorspace. Firstly, in using these variables, homogeneity has been assumed in the production functions of industries producing like commodities. In fact, if the degree of mechanisation varies, then estimation of activity based on employment, for example, is misleading (Watson, 1975). Therefore, measures such as output or use of materials may be superior. Secondly, as noted, previous models have not, in most cases, accounted for structural effects such as location. It may be appropriate to distinguish between plants located in the central areas and those in the peripheral areas. For example, unlike establishments located in central places, firms sited in inaccessible peripheral areas may base their activities on higher stock holdings; thus a variable such as warehousing facilities ought to be considered. In fact, some attempt have been made along these lines. For instance, Watson (1975), Leake and Grey (1979) used variables to account for storage facilities. Chisholm and O'Sullivan (1973) investigated the relevance of the centre-periphery model as a factor affecting freight patterns. Clearly there are many, theoretically usable variables, the performance of which has not been tested. Of particular interest is the failure of previous models to incorporate accessibility. This justifies the wider range of variables to be examined in this thesis(1).

A question related to the choice of variables is that of the distinction between O_i/D_j. The multi-regional studies, cited have shown that the explanatory variables used, best explain D_j's. Most establishment based studies concerned themselves with trip-ends estimation. Watson (1975) investigated the question of whether inbound and outbound truck trips might require different models. He found that employment best explained

(1) The reasons for investigating the possible effects of foreign trade on inland freight patterns are given in Appendix I.

inbounds trips. The reason may be that demand for input (as represented by inbound trips) is closely related to production (as measured by employment) whilst dispatch of production (distribution) is more related to storage and inventory policies for which floorspace may be a good proxy. In this study, therefore, separate models are developed for O_i's and D_j's.

(iii) Mathematical Form of the Relationships

In establishment based studies, both linear and non-linear form have been found useful. Starkie (1967) suggested that the basic distribution of vehicle trip against employment or floorspace was non-linear. This non-linearity has been explained by economies of scale. This is understandable since large establishments can organise larger average loads per trip than smaller establishments in the same industry. Following Starkie's work, almost all researchers in the field tested log-transforms. Redding (1972) found no apparent advantage in using transformed variables. Maltby (1973) however found significant log-relationships. Leake and Gan (1973) tested four basic forms viz. linear, parabolic, geometric and exponential. Only the linear and the parabolic equations proved to be significant. In multi-regional studies however only the linear form has been considered and has proved useful.

In the existing body of research it would seem that the chief criterion of comparison between alternative mathematical forms has been the R^2. It is regrettable that full use of statistical techniques[(1)] has not been made. Although transformations may, sometimes, be justified on a priori grounds, an analysis of the regression residuals or an analysis of variance provide two solid bases on which the superiority of a given form may be judged.

2.3 Implications for this Research

The foregoing discussion leads to several conclusions relevant to the conduct of a further study.

(i) Aggregation

In this respect multi-regional studies have had to use available data, notably the given zonal breakdown and employment at the S.I.C. orders level (see Appendix II.1). These studies have assumed homogeneity within the S.I.C. groups and have not considered explicitly the problems of aggregation. Although analysts of establishment goods movement have devoted some attention to the problem of aggregation over industries, they have tended to look at specific levels of aggregation (often the S.I.C. order level). This study proposes to look at different levels of

---

(1) For treatment of non-linearities in freight models see Starkie (1971).

aggregation and to consider a wide range of industries. As regards spatial aggregation two levels are considered and the effects of spatial aggregation are statistically evaluated. The effects of modal aggregation are also examined.

(ii) Choice of Variables

The variables tested to date have been of a limited nature. This research evaluates the relevance of variables used in previous work and widens the range of explanatory variables. In particular more attention is devoted to relating the models to locational variables such as accessibility.

(iii) Mathematical Forms of the Models

While analysts of vehicle goods - movement have kept an open view of alternative functional forms, previous research at the regional level has presented solely the linear form. Here, the linear form is considered as the working hypothesis. Nevertheless, a systematic scrutiny of the residual plots as well as statistical tests of O.L.S. assumptions are undertaken. In general, a transformation is carried out only when statistically justified.

CHAPTER 3

DATA

3.1 Freight Data

The basic freight data used in this study were provided by British Rail. They consisted of flows by commodity group and mode. This section provides a brief description; a fuller description can be found in British Rail Board (1975) and in Pitfield (1977).

Spatial Framework: The U.K. mainland has been divided up into 134 traffic areas. Figure 3.1 shows the zones for which the data are provided and Appendix II.1 lists the centroids of the zones.

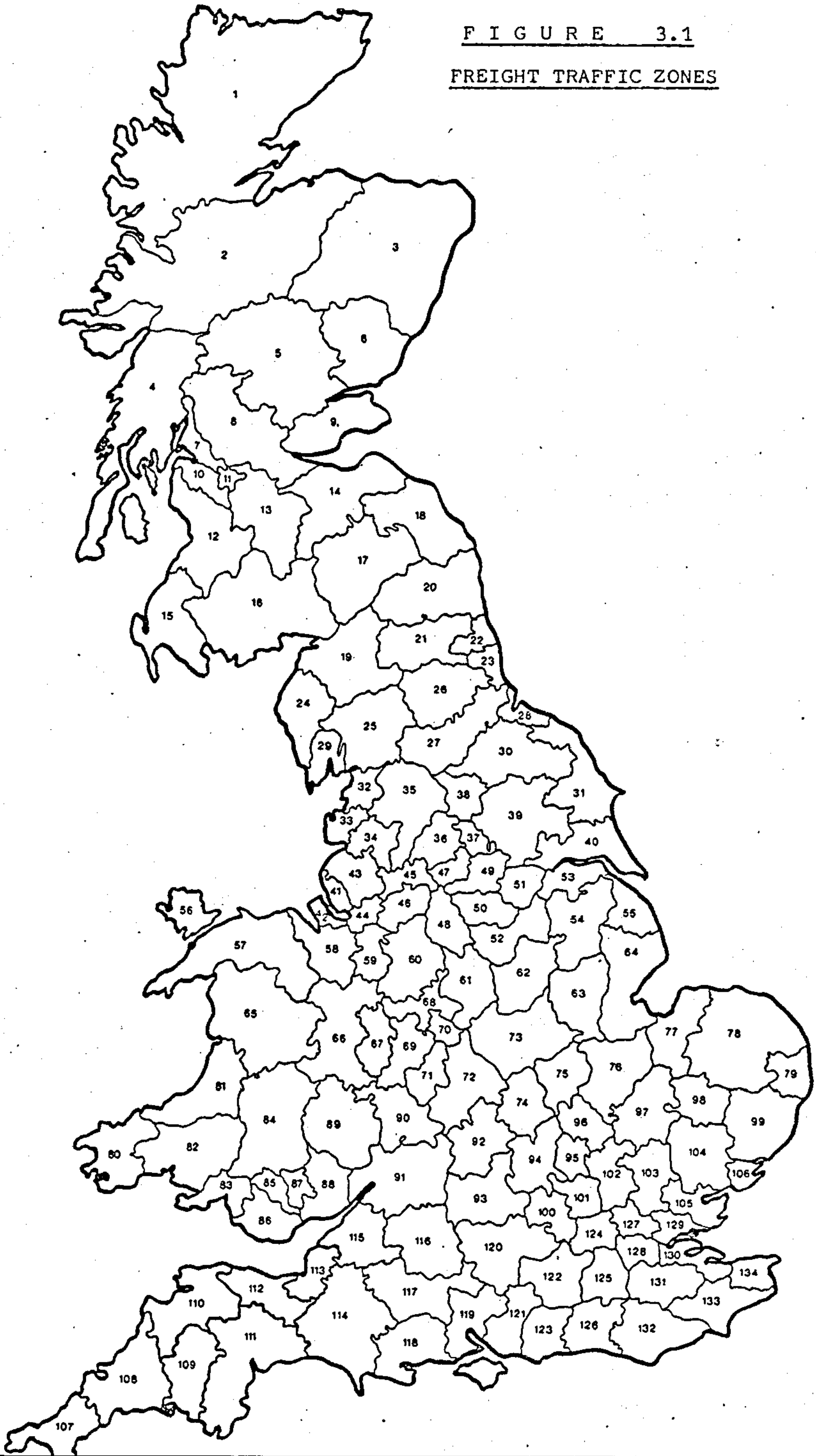
Commodity System: The freight flows were given in 30 commodity groups. The objective when defining these groups was to "reduce heterogeneity" (British Rail Board, 1975). The degree to which such an objective has been achieved may be judged from Appendix II.2.

Rail Data: Rail freight data were derived from continuous recordings of rail traffic.

Road Data: The road freight data were based on the 1967/68 Road Goods Survey up-dated to 1972, and converted to the spatial and commodity framework above.

FIGURE 3.1

FREIGHT TRAFFIC ZONES



Other Modes: This research does not consider pipeline or coastal shipping for which data are not reliable (British Rail Board, 1975; Pitfield, (1977). Besides, the overall share of these latter modes is not important, as seen from Table 2.1.

TABLE 2.1

TOTAL FREIGHT BY MODE, 1972

Mode	Total Tons, Millions
Road	1672.1
Rail	195.0
Coastal Shipping	44.7
Pipeline	51.2
All Modes	1963.0

The Distance Matrix

The road goods survey data was used to derive zone centroids as road freight dominates goods movement in Britain. The derivation was based on the traffic flows between each pair of areas and a comparison of average length of haul with road distance between towns in the area (see References cited above).

3.2 Economic Data

Much attention was given to data collection in this area. However, the ability to preserve the spatial detail which exists in the freight data was severely constrained by the limited number of independant variables at a compatible level. Consequently it was decided:

(i) To devote some time to gathering data at the zonal level;

(ii) To sacrifice the detail of the original freight data for a more extensive exploratory analysis at the level of the standard regions.

3.2.1 Zonal Economic Data

Apart from population data, other economic data are not readily available at this spatially detailed level. As employment has proved to be a useful variable in explaining freight generated it was decided to try gather employment data for the most detail industrial classification viz. the MLH subdivisions. This would enable the commodity groups of the freight data to be replicated in the employment. Inquiries revealed two potential sources namely the employment exchange and the 1971 Census data.

The Employment Exchange Data

Data were available for 24 S.I.C. orders, for 1972 for each of the employment exchange areas across the country. These data were rejected on the following grounds. Unlike the British Rail zones, the employment exchange areas do not always correspond to administrative areas and consequently there is no firm common base for matching the sets of areas. The use of overlay maps at the Department of the Environment, map library revealed serious difficulty in merging the two sets of zones. For example, even if three to four British Rail zones are aggregated, this does not necessarily correspond to an aggregation of employment exchange areas.

The 1971 Census Data

Employment data disaggregated down to MLH level are provided in the 1971 Census data for all local authorities. The data, which consist of 10 per cent sample for all areas across the U.K., are described in HMSO (1975). On the basis of Appendix III3, it has been possible to group local authorities area into B.R. zones. Note that the data used are those recorded at the work place.

3.2.2 The Regional Data

Although basically the level of aggregation adopted is a matter of data availability, the adoption of the regional level for analysis is interesting for at least three reasons:-

(i) Because a large number of socio-economic measures are published for the regions it is possible to test for their relevance in explaining freight movements;

(ii) It is possible to gain some insight into the question of spatial aggregation by comparing analyses for different levels of aggregation;

(iii) A knowledge of regional traffic forecasts, which are often used as bases for planning, may be useful. In fact similar levels have been used in other countries. For example, only eight major regions have been considered in order to establish forecasts for 1985 in France, although that country is devided into 85 traffic zones for other purposes (Commissariat General du Plan, 1972).

Problems are however associated with the analysis at the regional level. These are the following:-

(i) With only 10 standard regions as observations, there is a limit to the number of explanatory variables that may be included in the models;

(ii) Aggregating to the regional level increases the chances of grouping heterogeneous zones(1). Since regression analysis works on the variation between groups (here the regions), if heterogeneity within groups is great, there may be considerable variation left unexplained (Fleet and Robertson, 1968; Kassof and Deutschman, 1969).

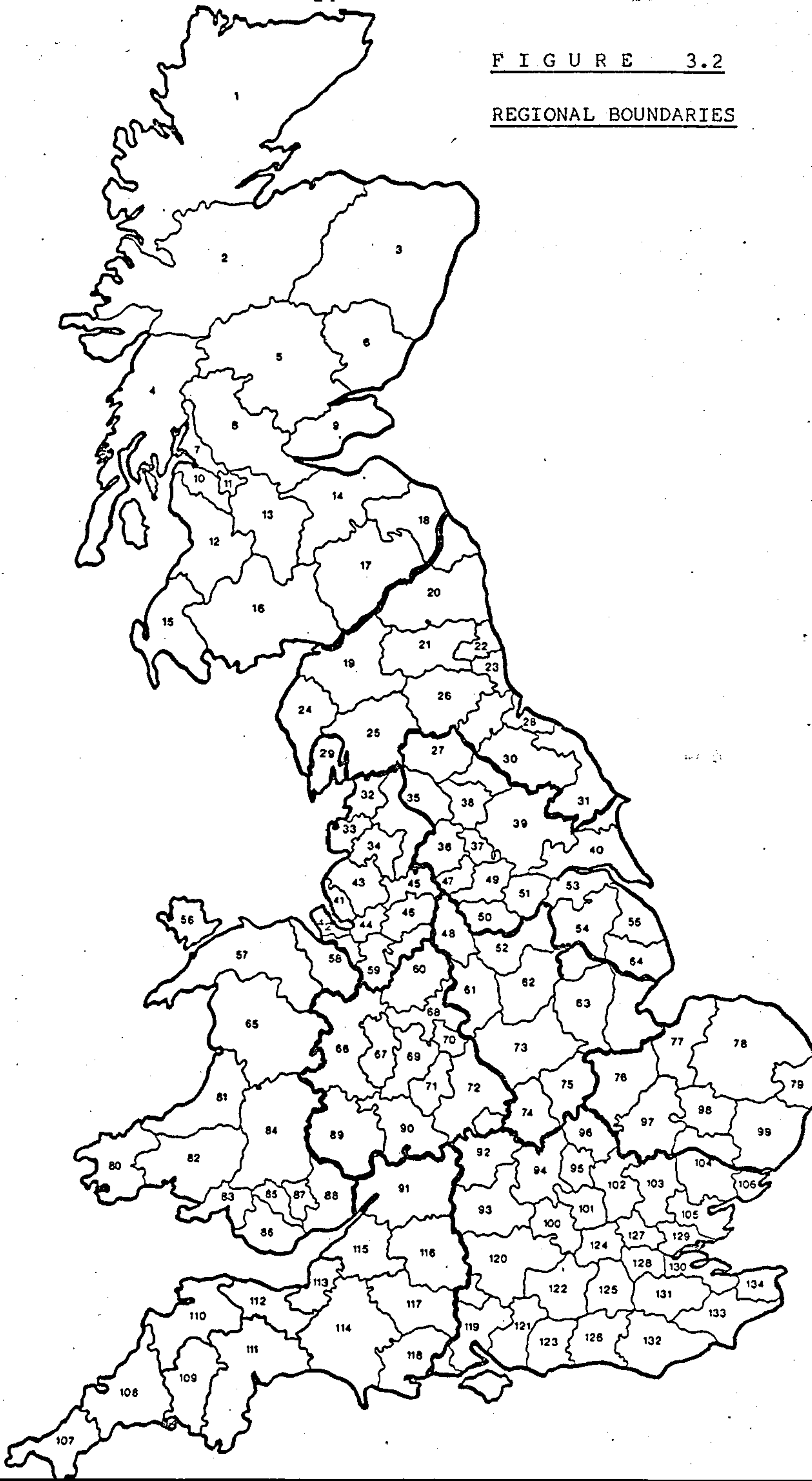
The work with the regional data is constrained by missing information for some industries in some regions as well as the difficulty experienced in matching the B.R. commodity groups to the appropriate industrial groups. As that part of the work is viewed as exploratory, only 14 commodity groups were used for the analysis (see Appendix II.3). Appendix II.3 also lists the sources for the regional economic data.

---

(1) Figure 3.2 shows the regional boundaries. Appendix III.2 gives the relations used for conversion of zonal data to regional data.

FIGURE 3.2

REGIONAL BOUNDARIES



CHAPTER 4

TECHNIQUES TO BE USED

4.1 Regression Analysis

One of the most practicable analytical tools to use to derive an explanation of freight in terms of some determinants is regression analysis. This technique has been successfully used in past studies. As Watson (1975) remarks "It is sufficiently flexible to allow the testing of single variables or combinations of variables; to investigate the relative significant factors, and to examine a variety of model forms".

4.1.1 Approach

At the most aggregate level, given a postulated linear relation between freight generation and the level of economic activity, one may write:

$$O_i^T = b_o + \sum_{j=1}^{k} b_j X_i^j + e_i \qquad (4.1)$$

where, O_i is originating freight in zone i, and T is the superscript denoting total freight aggregation:

b_o , b_j are parameters to be estimated in such a way as they satisfy the least squares criterion,

X_i^j , $j = 1$ to k is a set of k surrogates for economic activity in zone i and e_i' is the error term for observation i.

A similar relationship may be written for D_j when disaggregated by mode m, (4.1) becomes:

$$ {}^{m}O_i^{T} = b_o + \sum_{j=1}^{k} {}^{m}b_j \, X_i^{j} + e_i \qquad (4.2) $$

where, m = 1,2,3, road, road plus rail, rail.

Disaggregated by mode m and commodity c, (4.1) may be written as:

$$ {}^{m}O_i^{c} = b_o + \sum_{j=1}^{k} {}^{m}b_j^{c} \, X_i^{j} + e_i \qquad (4.3) $$

and similarly for terminating traffic D_j.

4.1.2 Main Assumptions of Multiple Regression Analysis

(1) The explanatory variables X_i^j must be linearly independent to avoid the effects of multicollinearity. These effects lead to a loss of precision characterized by large sampling variance. Practically these lead to inefficient tests on b_o and b_j. For example, variables that are expected to have a real impact might appear to be non-significant.

(2) The values of the error terms should be independent of one another to avoid spatial autocorrelation. This may be due to the omission of a variable or an incorrect specification of the model (Cliff and Ord, 1973). The consequences of autocorrelation are notably the inapplicability of the t and F statistics.

(3) The remaining assumptions about the error terms may be summarized symbolically by:

$$e_i \sim N(0, \sigma^2),$$

i.e. the error terms are normally distributed, have mean zero and constant variance. The most pertinent assumption is that of constant variance (homoscedasticity) since it is often violated in cross-sectional data. Its violation occurs when differences in the structural background for different observations are not specified. In multi-regional freight studies for example, an attempt has been made to specify the structural background of different zones (port-zones and non-port zones and regional effects, (Department of the Environment, 1971)).

Unlike previous studies, it is intended here, to investigate the likely violation of the foregoing assumptions. As regard the assumption of independence of the explanatory variables, the correlation matrices are examined and when necessary recourse is made to principal component analysis (see Section 4.3, below). The assumptions about the error terms are dealt with as follows: (i) the assumption of homoscedasticity and spatial independence are considered as the most severe and are tested respectively by the spearman correlation coefficient and by the generalized MORAN's I Statistic (Cliff and Ord, 1973); (ii) the assumptions of normality and randomness are checked by means of examination of the residuals.

4.2 Covariance Analysis

4.2.1 Relevance of Covariance Analysis

This technique is used to investigate two questions. (1) To check whether industry groups are sufficiently homogeneous to warrant their aggregation. Covariance analysis takes, here, the form of "comparison of regression". If, for example, the same regression holds for iron, ores, scrap, alloys and steel, these industries can be aggregated, without loss of precision into a broader group of metal industries. (2) To test whether groups of zones (urban, rural, connurbations) could be isolated. This is related to the question of whether the omission of variables that account for the structural background of the zones leads to over or under predictions in given zones. In this case covariance analysis takes the form of regression analysis with dummy variables.

4.2.2 Analytical Procedure

The modelling strategy for each of the two alternative forms of covariance analysis are now discussed.

1. Comparison of Regression

The purpose of this approach is to examine whether the linear regressions of freight on employment are the same for different industry groups "i". These may differ in residual variances, in slopes and/or in intercepts. Thus, following closely - Snecedor and Cochran (1969), the computations are carried out in the following steps.

(i) Compare the residual variances. The test takes two possible forms. If the number of classes i = 2, a two-tailed F test may be used. If the number of classes i > 2, the Bartlett test for homogeneity of variances may be used. If heterogeneous variances are found, this may be pertinent.

(ii) Assuming homogeneous variances and letting B_i be the true slope, the question is to decide whether the different slopes $b_1 = b_2 = b_3 \ldots = B_i$.

(iii) Assuming identical slopes it remains to test the null hypothesis that the different intercepts are the same. That is $b_{o1} = b_{o2} = b_{o3} \ldots = B_{oi}$.

For (ii) and (iii), F tests between appropriate mean squares are used[(1)].

2. <u>The Dummy Variable Approach</u>

Here the problem is of introducing into the generation model the qualitative factor zone type. The question is, in this study, to decide upon the significance of zone type when the quantitative variable is controlled.

The first step is to draft a number of classes based on the type of zones (for example industrial centres, port zones, etc.). Each class is then represented by a dummy variable.

Next, the possible existence of class effect is judged in terms of different intercepts. The model is (assuming four types of zones and one explanatory variable).

$$y = b_o + b_1 X + b_2 D_2 + b_3 D_3 + b_4 D_4 \qquad (4.6)$$

where, D_i = 1 for observation in class i, i = 2,3,4*
= 0 otherwise

---

(1) The procedure is fully described in Snecedor and Cochran (1969), Chapters 10 and 14.

* One dummy variable is omitted for reason of determinacy. It is then considered as reference class and information concerning it, is easily calculated.

Equation (4.6) assumes a common slope (which is the coefficient of the X variable). This assumption can be overcome by introducing interaction terms as shown in Equation (4.7) below.

$$y = b_o + b_1X + b_2D_2 + b_3D_3 + b_4D_4 + b_5(D_2X) + b_6(D_3X) + b_7(D_4X) \qquad (4.7)$$

A simple test of significance of the coefficient in (4.7) enables the existence of class effect to be judged (Johnston, 1972). However, analysts often, prefer to summarize the results in a covariance table as shown on Page 32 .

TABLE 4.1

COVARIANCE TABLE (Source: Nie, 1975)

Source of variation	Sum of squares (SS)	d.F	F
(I)SS due to D and X	$SS_y \; R^2_{D,X}$	k_1+k_2	$\frac{(I)/dF(I)}{(V)/dF(V)}$
(II)SS due to D, adjusted for X(differential intercept)	$SS_y \; R^2_{D,X}-R^2_X$	k_1	$\frac{(II)/dF(II)}{(V)/dF(V)}$
(III)SS due to X, adjusted for D	$SS_y \; R^2_{D,X}-R^2_D$	k_2	$\frac{(III)/dF(III)}{(V)/dF(V)}$
(IV)SS due to X interaction (differential slopes)	$SS_y \; R^2_{D,X,DX}-R^2_{D,X}$	k_1k_2	$\frac{(IV)/dF(IV)}{(V)/dF(V)}$
(V)SS residuals	$SS_y \; 1-R^2_{D,X,DX}$	N-K-1	

where,

D = represents dummy variables and X the quantitative variables;

k_1 = is the number of dummy variables and k_2 the number of X variables;

K = $k_1+k_2+k_1k_2$ (degrees of freedom for Equation (4.7) as shown on Page 31;

N = sample size.

The quantities in Column 2 are derived from outputs of regressions (4.6), (4.7), a regression on the X variables and a regression on the dummy variables alone respectively $R^2_{D,X}$, $R^2_{D,X}$, $R^2_{D,X,DX}$, R^2_X and R^2_D (Nie, 1975).

Element in the numerator, in Column 4, represent the mean squares of the respective entry. The denominator is the mean square for the residual.

4.3 Principal Component Analysis

4.3.1 Relevance of Principal Component Analysis

This technique transforms a given set of data into a smaller new set of composite variables that are orthogonal (uncorrelated with one another). For the analysis at the standard region level, in particular, the number of observations is limited to 10 whereas the number of exogeneous variables that may be used are numerous and moreover highly correlated so as the lack of degrees of freedom and multi-collinearity would constitute serious constraints to multiple regression. Therefore, both its data reduction and the statistically desirable property (orthogonality) of the variables it creates make this technique most suitable for the data in hand.

This technique is preferred to other factor analysis methods, for unlike the latter it does not require assumptions about the distribution of the data (Lawley and Maxwell, 1971). The factor analysis model (4.8) assumes a given regularity in the data (this is expressed through the existence of the unique factor U_j).

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + \dots a_{jn}F_n + d_jU_j \qquad (4.8)$$

where, F_i are the hypothetical factors

a_{ji} the loadings

U_j the unique factor

d_j the loading on the unique factor

On the other hand, the principal component model (4.9) is an exact mathematical transformation of the original variables Z_i and no question of a hypothesis arises.

$$Z_j = a_{j1}C_1 + a_{j2}C_2 + \dots a_{jn}Cn \qquad (4.9)$$

where, a_{ji} are the loadings and C_i the components.

4.3.2 Identification of the Components

Two questions are fundamental in this sort of analysis:

(i) How many components should be retained?

The most commonly used criterion is Kaiser's test (Child, 1971). This test requires that only components with associate eigenvalues $\lambda_i \geqslant 1$ can be retained. It ensures that only components accounting for at least the amount of the total variance of a single variable is treated as significant.

(ii) How are the components interpreted?

The loadings are similar to correlation coefficients. It has thus been suggested to test for their significance in an alike manner. The common test is based on the standard errors of the Pearson correlation coefficient ($S_{(r_{xi\ xj})}$). However, the Burt-Bank's formula is retained, here, for it is more rigorous (Child, 1971). In fact it consists of an adjustment of the standard error of the correlation coefficient as seen from (5).

$$S_{(amj)} = \left\{ S_{(r_{xi\ xj})} \right\} \cdot \sqrt{\frac{n}{n+1-m}} \qquad (5)$$

where,

$S_{(amj)}$ is the adjusted standard error of the loadings

$S_{(r_{xi\ xj})}$ is the standard error of a correlation (as reported in the significance levels for Pearson product-moment correlation coefficient)

n is the number of variables

m is a subscript for the component (position in the extraction).

Thus to be considered a loading must be greater than or equal to $S_{(amj)}$ for the appropriate sample size.

4.3.3 Regression on Principal Components

Theoretically the best way to use principal components in regression is yet unresolved (Johnston, 1972). Empirically there are three ways in which it has been used. In the first case, the set of potential explanatory variables is so excessive relative to the number of observations that regression on the smaller set of principal components is preferred to preserve degrees of freedom. A variant of this approach has been to retain a number of specific, important original variables for the final regression, along with the principal components (Pidot, 1969). The third procedure has been used when the chief concern has been multi-collinearity. The procedure is then to use the components in a regression and transform back to obtain estimates of the coefficients of the original variables.

For the exercise at the standard region level, with only 10 observations and high multi-collinearity, the first approach is used. For the zonal models the variables that load most highly on each factor are used in the regressions. This is justified on the grounds that the zonal models are viewed as reliable models for forecasts while the regional models are considered as explanatory. In forecasting context, the planner would be interested in the performance of a single variable that may represent a set of variables.

PART II

REGIONAL ANALYSIS

This part of the analysis is concerned with the building of total and commodity freight models at the level of the standard region[1]. As the high multi-collinearity between the explanatory variables constituted a major problem, it was considered convenient to:

(i) present the results derived from regression on the original variables;

(ii) discuss the correlation matrices so as to illustrate the problem of multi-collinearity and justify the use of principal component analysis;

(iii) describe the results obtained from regressions on principal components.

This leads to the following structure of

---

(1) Analysis at this level of spatial aggregation is of an exploratory nature and it is argued (see Part III) that if forecasts are needed for the regions then they ought to be derived from forecasts based on the zonal models. As the main purpose of this part is explanation rather than reliability for forecasts, the statistical tests on the assumptions of O.L.S. are not computed. Nevertheless comments on the visual patterns of the residual are provided.

this part of the thesis. Chapter 5 deals with total freight analysis. In Chapter 6, the results from regression of commodity tonnage on various economic measures are described. Chapter 7 contains the results obtained from regressions of commodity tonnage on principal components and a discussion of the main findings of this part of the analysis.

CHAPTER 5

TOTAL FREIGHT MODELS

5.1 Regression on Original Variables

The approach was:

(i) the explanatory variables whose intercorrelations were relatively low were used in a stepwise regression;

(ii) with the remaining variables, simple regression models that were theoretically the most plausible were developed in a first step. Then additional variables were inserted using the correlation matrix as a screening device.

The best multiple regression models are reported in Table 5.1. They were derived from stepwise regressions of total freight by road, by rail and by road plus rail on the following measures: petroleum consumption, coal consumption, cement deliveries to regions, production of steel and production of pig-iron.

Road Models

The two included variables, namely petroleum consumption and cement deliveries, accounted for 96 and 97 per cent of the variance explained in originating and terminating traffic respectively. In both equations

TABLE 5.1

STEPWISE REGRESSION RESULTS FOR TOTAL FREIGHT ON SELECTED X's

Standard Regions

F(.01) = 9.55
F(.001)=21.7

Dependent Variables	Constant b_o	Coefficients	R^2	F
Road				
O_i	13372.92	$20.43X_{10} + 55.97X_5$	0.96	82.03
D_j	6419.14	$22.59X_{10} + 56.58X_5$	0.97	98.62
Road + Rail				
O_i	36948.26	$13.08X_3 + 26.74X_5$	0.92	44.65
D_j	49057.79	$18.56X_3 + 10.623X_5$	0.91	38.39
Rail				
O_i	6780.56	$2.38X_3 + 2.35X_{19}$	0.63	10.48
D_j	5778.78	$2.76X_{19} + 3.79X_{18}$	0.64	11.56

X_3 coal consumption

X_{10} petroleum consumption

X_5 cement deliveries for regions

X_{19} steel production

X_{18} iron production

the cement deliveries variable first entered the regression. It accounted for 57 and 51 per cent of the variance explained in D_j and O_i respectively. Both equations are significant at the .01 level.

<u>Road Plus Rail</u>

The models included coal consumption and cement deliveries as explanatory variables. The high level of explanation achieved is due mainly to cement deliveries for the O_i equation (58 per cent) and to coal consumption for the D_j relationship (53 per cent). The models are significant at the .01 level.

<u>Rail Models</u>

The models for rail achieved a lower level of explanation. The R^2 obtained are 0.63 and 0.64 for O_i and D_j respectively. In this case, only one variable is common to O_i and D_j namely steel production. This variable contributed to only 27 and 24 per cent of the variation explained. The variables that contributed to the remaining shares are coal consumption and pig iron production respectively for O_i and D_j.

These findings support the two hypotheses which were previously suggested: (i) freight movements can be modelled by measures of consumption and production of materials; (ii) it is not always the same variables that best explain O_i and D_j.

The remaining models described in this section are simple regression models. In each of the following models inclusion of additional explanatory variables added very little to the significance of the relationship. Table 5.2 shows the results obtained for total freight by road and total freight by both road and rail regressed on three independent variables, viz. Regional gross domestic product, total employment and resident population.

Regional G.D.P.

This variable explained a great amount of the variation in total freight. It achieved a minimum $R^2 = 0.87$ for O_i by road and explained up to 94 per cent of the variation in D_j by road. Its failure to explain rail freight is understandable since rail serves only a limited number of economic sectors whilst G.D.P. is a measure of the overall domestic production.

Employment

This variable achieved a similar pattern of explanation to G.D.P. However, in terms of variation explained, total employment performed better than G.D.P. As seen from Table 5.2, the minimum R^2 yielded is 0.89 for O_i by road and rail and the maximum $R^2 = 0.96$ is for D_j by road.

T A B L E 5.2: Total Freight Generated, Regressed on: Regional G.D.P.(£000); Total Employment (000) and ~~Total~~ Population (000)

Standard Regions
F(.001) = 25.4

Dependent Variables	Independent Variables	Constant b_o	Coefficient b	R^2	F
Total tonnage by Road plus Rail					
O_i	G.D.P.	92043.140	17.778	0.87	58.493
D_j	G.D.P.	85434.197	19.062	0.91	84.658
O_i	total employment	80513.163	46.978	0.89	71.880
D_j	total employment	72883.917	50.455	0.94	130.524
O_i	total population	69632.063	21.542	0.91	81.833
D_j	total population	61533.657	23.064	0.94	148.053
Total tonnage by Road					
O_i	G.D.P.	69482.610	18.383	0.93	116.824
D_j	G.D.P.	66260.399	19.017	0.94	131.112
O_i	total employment	57784.778	48.478	0.95	163.720
D_j	total employment	54029.266	50.207	0.95	204.443
O_i	total population	46718.147	22.200	0.96	205.713
D_j	total population	42723.180	22.962	0.96	251.337

Resident Population

As judged from R^2, this variable yielded better results than the first two. The pattern of results with respect to modes and O_i/D_j differences is however similar.

Other Variables

Foreign Trade Variables

The need to test for the relevance of these variables is discussed in Appendix I. The best models obtained from regression of total freight by different modes on imports and exports are shown in Table 5.3.

TABLE 5.3

TOTAL FREIGHT REGRESSED ON IMPORTS

Standard Regions
F(0.05) = 5.32

Dependent Variables	b_o	b	R^2	F
Road				
O_i	99601.55	3.26	0.52	8.84
Road+Rail				
O_i	119170.44	3.25	0.51	8.26

The relationships are significant at the 5 per cent level. The poor fit is as expected (see Appendix I). The fact that terminating movements regressed on exports did not prove a significant fit even at the 5 per cent level supports the conjecture put forward in Appendix I, that due to their composition British exports are unlikely to have an impact on inland movement in an analysis in terms of tonnage.

Number of Vehicles Licences

This variable was considered as a proxy for the supply of transport facilities. The results obtained are unusual. As seen from Table 5.4, the linear form yielded a reasonable R^2 but the standard error of the estimate was relatively high. Examination of the residual plots revealed an abnormal pattern. Reference to statistical work on regression residuals suggested that a quadratic or a transformation of the y variables was needed (Draper and Smith, 1966). The semi-log transformation yielded the best results. In terms of correlation the non-linear form has achieved a lower R^2 but the standard errors were considerably lower.

Unsuccessful Variables

The variables population density, employment density, warehouse floorspace and region area did not prove useful. The first three variables were examined with the purpose of testing whether the concept of centrality is effective at the regional level. The last was examined on the grounds that some relationship

TABLE 5.4

REGRESSION OF TOTAL FREIGHT BY ROAD ON THE NUMBER OF VEHICLES LICENCES

Standard Regions
F(0.01) = 11.26

(1) $y = b_o + bX$, (2) $\text{Log } y = b_o + bX$

	b_o	b	R^2	F	Standard error b_o	Standard error b
(1)						
O_i	62530.545	631.782	0.92	94.31	14069.437	65.024
D_j	59173.528	652.918	0.92	100.82	14063.439	65.024
(2)						
O_i	11.399	0.002	0.71	20.38	0.143	0.091
D_j	11.419	0.002	0.71	19.86	0.141	0.034

might exist between regional specialisation and region area (Brown, 1972). Because Brown's analysis suggested a non-linear form, transformation were performed. Log, semi-log and second degree polynomial forms were unsuccessfully tested with variable regions area.

5.2 Correlation Matrix

Table 5.5 describes the intercorrelations between the explanatory variables initially considered as characteristic. The pattern may be summarized as follows:

TABLE 5.5

CORRELATION MATRIX: AGGREGATE EXPLANATORY VARIABLES

(Standard Regions)

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19
Total Employment	1	0.99	0.72	0.12	0.92	0.95	0.90	0.64	0.74	0.28	0.99	0.99	0.96	0.96	0.98	-0.28	0.40	-0.30	-0.31
G.D.P.		1	0.70	0.12	0.98	0.96	0.91	0.65	0.75	0.38	0.98	0.90	0.97	0.97	0.89	-0.34	0.47	-0.38	-0.32
Coal Consumption			1	-0.02	0.63	0.68	0.45	0.40	0.41	0.63	0.77	0.48	0.59	0.58	0.04	0.19	0.53	-0.49	-0.12
Region area				1	0.13	0.08	0.26	0.03	0.20	-0.20	0.15	-0.02	0.09	0.10	0.00	0.40	0.10	0.06	0.07
Cement Deliveries					1	0.99	0.91	0.66	0.76	0.18	0.97	0.24	0.98	0.98	0.31	-0.39	0.50	-0.33	-0.35
Number of Vehicles						1	0.91	0.64	0.73	0.20	0.98	0.30	0.97	0.97	0.40	-0.39	0.96	-0.38	-0.35
Population							1	0.56	0.78	0.52	0.86	0.36	0.93	0.93	0.65	-0.49	0.60	-0.22	-0.18
Exports								1	0.91	0.25	0.63	0.37	0.72	0.69	0.40	-0.31	0.70	0.22	0.25
Imports									1	0.45	0.71	0.46	0.83	0.81	0.63	-0.41	0.82	0.13	0.25
Petrol Consumption										1	0.24	0.50	0.40	0.21	0.66	0.27	-0.06	-0.28	-0.19
Retail Turnover											1	0.64	0.94	0.94	0.78	-0.21	0.92	-0.39	-0.32
Net Output in Manufacturing Industry												1	0.89	0.64	0.97	0.08	0.40	-0.25	0.03
Net Capital Exp. in Manufact. Ind.													1	0.99	0.90	-0.49	0.99	-0.30	-0.25
Gross Domestic Profits & Surpluses														1	0.90	-0.49	0.93	-0.32	-0.29
Income from Employment															1	0.06	0.46	-0.15	0.00
Public Investment																1	-0.53	0.05	0.14
Agriculture Investment																	1	-0.30	-0.28
Production of Iron																		1	0.68
Production of steel																			1

(i) the great majority of first order correlation coefficients are greater than 0.60;

(ii) the majority of variables are highly correlated with one another. Only three variables namely production of steel, production of iron and agriculture investment seem unrelated to the other variables. While these latter variables are measures of activities whose production is concentrated in few regions, the majority of other variables are in fact measures of the same general influence. For example total employment (X1), total population (X7) or gross domestic product (X2) can all be considered as surrogates for the general level of economic activity.

The conclusions that may be drawn from Table 5.5 are:

1. The similarity obtained in the regression results for the three variables G.D.P., total employment and total population is explained by the fact that the variables correlated highly with one another and therefore would be expected to model freight generation in a similar way.

2. The high frequency of strong intercorrelations between explanatory variables explains the failure of multiple regression models.

3. The pattern of intercorrelations indicates that the space economy of Britain is characterised by some localised activities.

Due to these conclusions and because the study is concerned with the workings of regional economies the use of principal component analysis was considered, in order to isolate major dimensions in the regional economies and to then test their association with total freight generation.

5.3 Regression of Principal Components

5.3.1 Derivation and Identification of the Components

A reduced matrix or variables expressed in the same unit (£000) is subjected to principal component analysis. These are retail turnover, net output for manufacturing industries, gross trading profits, net capital expenditure for manufacturing industries, income from employment and agriculture investment. The reasons for confining the analysis to this set of variables are that (i) Johnston (1972) points out that it is preferable to use variables measured in the same unit (ii) within the study as a whole, this reduction in the number of variables does not seem to involve a loss of information since the variables not used here were previously tested in the conventional regression analysis.

Results

The results are shown in Table 5.6. The reduced matrix of X variables yielded only one component which meets the criterion $\lambda_i \geqslant 1$. In Table 5.6, two components are reported. The second is retained as its eigenvalue is very close to one and because it is easily identifiable.

Component C1

This principal component, which on its own accounts for 82 per cent of the variance in the original data set has a relatively clear meaning. The variables that load most highly on this component are X2 and X3 respectively, net output and net capital expenditure in manufacturing industries. The third highest association is with (X4) gross trading profits and surplus. Additionally, it is worth noting that X6 which measures agriculture investment does not load significantly on this component. On these grounds, component C1 may, safely, be identified as a measure of industrial performance.

The pattern of component scores[1] over the regions supports this identification. The highest scores are found in the South East and the North West

---

(1) Analysts suggest to check whether the distribution of component scores sustain the identification (for example, Robson, 1971; Baxter, 1976).

TABLE 5.6

COMPONENTS DERIVED FROM AGGREGATE MEASURES, STANDARD REGIONS

Variables		Component C1 (Loadings)	Component C2 (Loadings)
Retail Turnover	(X1)	0.914	0.391
Net Output in Manuf.Ind.	(X2)	0.969	0.191
Net Capital Expenditure	(X3)	0.962	-0.076
Gross Trading Profits & Surpluses	(X4)	0.939	0.329
Income from Employment	(X5)	0.924	0.369
Agriculture Investment	(X6)	0.167	0.978
Eigen Values		4.93	0.93
PCT of Variance		82.2	15.5
CUM PCT		82.2	98.8

Extreme Values of Principal Components Scores	C1	C2
HIGHEST :	South-East, North-West	Scotland, South-West
LOWEST :	East Anglia, South-West	North-West, West Midlands

both industrialised areas; the lowest scores are recorded for East Anglia and the South-West both of which include large rural areas.

Component C2

This component, which accounts for 15.5 per cent of the total variance, appears to be a logical complement to the previous measure. The single significant loading relates component C2 to variable (X6), agricultural investment. This suggests the identification of component C2 as a measure of agricultural activity. The pattern of scores is consistent with the above identification. In fact, the highest scores appear in the South-West and Scotland whilst the lowest in the West Midlands and the North-West.

The two components isolated a major dimension in the regional economies namely the differentiation between levels of industrial development. However, the fact that the component representing industry accounts for over 82 per cent of the variation in the original data suggests that at the standard region level the dichotomy (predominantly rural or industrial) is not as great in the U.K. as it is in other countries (see also, Commission for European Communities, 1975).

5.3.2 Regression on Principal Components

With only two components, conventional multiple regression was performed. The coefficient of C2 (measure of agricultural activity) was negative and non-significant. This was not surprising given the impact observed for component C2 in the preceding section. In terms of total freight and at the standard region level therefore it may be stressed that agriculture has relatively little impact.

A simple regression of total freight generation on component C1, the measure of industrial development, yielded models shown in Table 5.7 below.

TABLE 5.7

REGRESSION OF TOTAL FREIGHT ON COMPONENT C1

Standard Regions
F(.001) = 25.4

Dependent Variables	Constant, b_o	Coefficient, b	R^2	F
Road & Rail				
O_i	84049.37	14.16	0.89	69.93
D_j	7655.41	15.20	0.94	121.81
Road				
O_i	61457.56	14.61	0.95	153.10
D_j	57876.72	15.13	0.95	184.64

As seen from Table 5.7, road freight achieved an $R^2 = 0.95$ for both O_i and D_j equations. For road and rail, 89 and 94 per cent of the variance was explained in, respectively, originating and terminating movements. The F ratios are all significant at the 0.001 level.

Summary of Findings

Satisfactory models were obtained with the basic variables population, total employment and G.D.P. The sign and magnitude of the coefficients in the models are plausible. The b coefficients yielded by total employment were of a larger magnitude suggesting a strong relationship between economic activity and freight movements. G.D.P. and total population yielded lower b coefficients than those observed for total employment. This is understandable as G.D.P., for example, represents the standard of living rather than the actual economic activity of the regions since it includes private and governmental redistribution of revenues, which, originally may have been created in other regions. The principal component C1 behaved similarly to these variables.

The poor results obtained for rail are largely due to the specialisation of this mode. This suggests that a disaggregation of commodity is the only way, rail may be successfully modelled. Such a conclusion is supported by the improvement of rail models when using variables such as production of steel or that of pig-iron (as substantial tonnage of these commodities is moved by rail, see Table A.5, Appendix IV).

CHAPTER 6

COMMODITY ANALYSIS (I), USE OF ORIGINAL VARIABLES

In the multi-regional freight studies previously cited, the freight flows were categorized into 15 commodity groups and the explanatory variables consisted of employment in each of the 24 S.I.C. orders. The models were derived from stepwise regressions of each commodity group on the 24 employment variables.

The authors reported very little success. They underlined, in particular, the undesirable high multi-collinearity between employment variables.

The approach, here, is twofold.

(1) Following the practice in establishment based studies, derive simple regression models relating commodity O_i, D_j to employment or net output in the apparently most applicable industry.

(2) The correlations matrices of employment, net output and net capital expenditure variables are subjected to principal component analysis so that useful, orthogonal groupings are derived. These are subsequently used in a stepwise regression.

6.1 Regression on Original Variables

Table 6.1 summarises the types of equations discussed in this section.

TABLE 6.1

TYPES OF REGRESSIONS REPORTED IN SECTION 6.1

DEPENDENT VARIABLES	INDEPENDENT VARIABLES
(1) O_i, D_j by Road + Rail, Road, Rail for 13 Commodity Groups	Resident population, employment in the respective industry and net output in the respective industry.
(2) O_i, D_j by Road + Rail, Road, Rail	
for - Steel	Steel production
- Cement & Building Materials	Cement delivery in regions
- Fuel Oil & Petroleum	Petroleum consumption
- Coal & Coke Product	Coal consumption

6.1.1 Resident Population

Road and Rail Models

The results are set out in Table 6.2. For either O_i or D_j, population performed well only with half the commodity groups considered. The successful relationships are obtained for commodities whose

TABLE 6.2: TONNAGE OF COMMODITIES BY ROAD PLUS RAIL REGRESSION ON RESIDENT POPULATION

Standard Regions
F(.001) = 25.4
F(.01) = 11.2

Dependent Variables	Constant b_o	Coefficient b	R^2	F
O_i				
Food, drinks & tobacco	4817.513	2.129	0.80	33.63
Sand, clays & stones	13764.931	3.749	0.86	52.12
Fertilizers	746.041	0.221	0.68	17.62
Cement & building materials	2945.903	4.947	0.96	233.99
Fuel oil & petroleum products	834.016	1.535	0.78	28.94
Vehicles & CKD	-236.911	0.327	0.61	12.87
Electrical equipment	-163.550	0.295	0.90	77.28
D_j				
Food, drinks & tobacco	5092.162	2.079	0.81	35.28
Sand, clays, stones	12694.365	3.943	0.83	40.10
Chemicals	1398.545	0.437	0.52	8.94
Plastics	88.779	0.056	0.58	11.09
Fertilizers	609.472	0.244	0.75	24.63
Cement & building materials	2458.749	5.049	0.99	890.82
Fuel oil & petroleum products	1764.866	1.375	0.80	32.44
Vehicle & CKD	-555.488	0.381	0.71	20.04
Electrical equipment	-191.654	0.301	0.90	76.84

shipments are a close function variation in population such as food, drinks and tobacco, cement and building materials. The relationships for this type of goods are significant at the 0.01 level. The amount of variance explained varies from a minimum of 61 per cent for O_i vehicles to 99 per cent for D_j cement and building materials. As with previous investigation the poorest results are obtained for minerals.

Road Freight Models

The results for road are reported in Table 6.3. The results obtained show a similar pattern to that discussed above. The marked differences are the decreases in the R^2 for O_i sand, clays and stones and for O_i vehicles.

Rail Freight Models

The regression results for this mode are generally poor. As seen from Table 6.4, the best equations are for O_i and D_j of the electrical equipment commodity which achieve respectively $R^2 = 0.89$ and $R^2 = 0.91$. Reasonable results are obtained for terminating traffic of commodities sand, clays and stones, chemicals and furniture, textiles and shoes.

6.1.2 Net Output

Net output variables are examined to see if they provide a better fit for commodity freight.

TABLE 6.3: TONNAGE OF COMMODITIES BY ROAD RESGRESSED ON RESIDENT POPULATION

Standard Regions
$F(.001) = 25.4$
$F(.01) = 11.2$

Dependent Variables	Constant b_o	Coefficient b	R^2	F
O_i				
Food, drinks & tobacco	4822.045	2.102	0.80	33.91
Sand, clays & stones	13877.190	2.013	0.52	8.72
Fertilizers	613.359	0.228	0.73	22.55
Cement & building materials	2624.902	4.917	0.96	843.24
Electrical equipment	-152.957	0.292	0.90	75.80
Fuel oil & petroleum products	416.844	1.263	0.75	24.18
D_j				
Food, drinks & tobacco	5009.241	2.058	0.81	35.50
Sand, clays & stones	12620.038	3.710	0.83	38.73
Chemicals	1402.489	0.425	0.51	8.49
Plastics	89.796	0.055	0.57	10.99
Fertilizers	513.771	0.245	0.78	29.05
Cement & building materials	2178.319	5.012	0.99	823.36
Fuel oil & petroleum products	666.085	1.220	0.80	32.29
Vehicles & machinery	-533.826	0.362	0.71	19.86
Electrical equipment	-183.447	0.298	0.90	75.09

TABLE 6.4: SELECTED RESULTS FOR COMMODITY TONNAGE BY RAIL REGRESSED ON RESIDENT POPULATION

Standard Regions
$F(0.05) = 5.32$
$F(0.01) = 11.26$

Dependent Variables	Constant b_o	Coefficient b	R^2	F
D_j				
Sand, clays, stones	74.979	0.232	0.52	11.018
Chemicals	-3.949	0.017	0.59	11.969
O_i				
Fuel oil & petroleum	417.256	0.271	0.53	9.272
D_j				
Vehicles	-21.639	0.019	0.61	12.406
Furniture, textiles and shoes	-3.011	0.001	0.64	14.413
O_i				
Electrical equipment	-10.812	0.003	0.89	64.892
D_j				
Electrical equipment	-8.563	0.003	0.91	84.706

Initially it is necessary to associate the orders for which net output data is available to the commodity classification. This association is described in Appendix II.3.

Road and Rail Models

The results are set out in Table 6.5. Improvements over the results obtained with resident population are observed for steel, chemicals, plastics and furniture textiles and shoes. The poor results were obtained, notably, for cement and building materials and for sand, clays and stones which were satisfactorily explained by population.

Road Freight Models

The results for road freight are reported in Table 6.6. Better models were obtained for this mode though the pattern of the relationships found were similar to those reported for road plus rail.

Rail Freight Models

The regressions for rail achieved in general a low level of explanation. For D_j's five equations are significant at the one per cent level. For O_i's however, only two relationship achieved an $R^2 \geq 0.50$. The results are shown in Table 6.7 on Page 65.

TABLE 6.5: COMMODITY TONNAGE BY ROAD PLUS RAIL REGRESSED ON NET OUTPUT (£000)

Standard Regions
F(.001) = 25.4
F(.01) = 11.2

Dependent Variables	Constant b_o	Coefficient b	R^2	F
O_i				
Steel	728.896	37.35	0.88	61.04
Food, drinks & tobacco	-72470.10	720.559	0.59	11.52
Chemicals	1586.625	16.861	0.80	32.78
Plastics	179.609	1.709	0.70	19.23
Fuel oil & petroleum products	4260.703	220.684	0.72	20.67
Vehicles & CKD	-497.517	14.379	0.87	53.70
Furniture, textiles, shoes	1783.075	28.179	0.54	9.40
Electrical equipment	953.264	6.884	0.50	8.29
D_j				
Iron, ores, scrap	2260.648	16.284	0.51	8.46
Food, drinks & tobacco	5277.933	68.700	0.84	41.52
Chemicals	1617.274	16.705	0.84	44.51
Plastics	139.561	1.992	0.80	32.01
Fuel oil & petroleum products	5635.312	166.432	0.52	8.75
Vehicles & CKD	-593.329	15.008	0.81	34.19
Furniture, textiles, shoes	1938.581	26.994	0.52	8.70
Electrical equipment	456.956	8.269	0.83	40.76

TABLE 6.6: COMMODITY TONNAGE BY ROAD REGRESSED ON NET OUTPUT (£000)

Standard Regions
F(.001) = 25.4
F(.01) = 11.2

Dependent Variables	Constant b_o	Coefficient b	R^2	F
O_i				
Steel	699.371	28.994	0.83	39.96
Food, drinks & tobacco	5073.445	69.080	0.82	35.35
Chemicals	1565.203	16.579	0.79	31.72
Plastics	176.239	1.716	0.70	19.41
Fuel oil & petroleum products	3388.719	175.768	0.64	14.65
Vehicles & CKD	-501.209	13.798	0.86	51.64
Furniture, textiles, shoes	1776.504	27.594	0.51	8.32
Electrical equipment	474.809	8.033	0.83	4.35
D_j				
Steel	1291.940	24.183	0.59	11.80
Food, drinks & tobacco	5278.198	67.846	0.83	40.74
Chemicals	1600.035	16.355	0.83	41.58
Plastics	139.439	1.975	0.80	32.02
Fuel oil & petroleum	4011.648	151.079	0.54	9.64
Vehicles & CKD	-569.343	14.242	0.80	33.68
Furniture, textiles, shoes	1933.557	27.000	0.52	8.74
Electrical equipment	459.148	8.173	0.83	40.11

TABLE 6.7: SELECTED RESULTS FOR COMMODITY TONNAGE BY RAIL REGRESSED ON NET OUTPUT

Dependent Variables	Constant b_o	Coefficient b	R^2	F
D_j				
Steel	230.301	6.202	0.68	17.011
O_i				
Food, drinks & tobacco	-4.812	0.911	0.51	8.599
D_j				
Food, drinks & tobacco	-5.676	0.914	0.90	72.422
Chemicals	18.105	0.347	0.69	9.700
Vehicles	-24.036	0.764	0.94	18.291
Electrical equipment	-2.374	0.089	-	130.870
O_i				
Fuel oil & petroleum	695.393	43.341	0.55	97.782

6.1.3 Employment by Industry

Road and Rail Models

The variance explained is, for the majority of the commodity groups, over 60 per cent. The best models are those for terminating traffic of vehicles and originating movements of coal and coke products. The results are shown in Table 6.8.

Road Freight Models

Comparing the results for this mode (Table 6.9) to those for road and rail, it may be seen that the amount of variance explained has markedly decreased for coal and coke products. A slight decrease is observed for vehicles too. For steel movements, a slight improvement is seen for O_i but D_j is now non-significant[1].

Rail Freight Models

As expected, the results (Table 6.10) are best for rail-born commodity groups such as coal and coke products. Another commodity group for which excellent models are observed, is electrical equipment

---

(1) These results are considered in Section 7.3.2. in an attempt to assess the difference in the performance of O_i/D_j, for different commodity groups when different independent variables are used.

TABLE 6.8: COMMODITY TONNAGE BY ROAD LUS RAIL REGRESSED ON EMPLOYMENT BY INDUSTRY

Standard Regions
$F(.01) = 11.26$

Dependent Variables	Constant b_o	Coefficient b	R^2	F (b)
O_i				
Coal & coke products	5449.726	498.413	0.88	60.36
Steel	1051.833	117.762	0.62	13.18
Alloys	455.293	24.705	0.65	14.88
Iron, ores, scrap	1952.600	264.384	0.81	34.11
Food, drinks, tobacco	3343.256	192.523	0.80	32.06
General chemicals	1584.099	65.247	0.82	35.27
Vehicles and machinery	-453.192	22.316	0.72	20.71
Furniture, textiles, shoes	1648.494	39.093	0.63	13.88
Electrical equipment	405.985	13.176	0.79	30.27
D_j				
Coal & coke products	10994.659	331.360	0.79	30.54
Steel	2040.120	90.716	0.47	6.99
Alloys	372.501	31.184	0.76	25.76
Iron, ores, scrap	2070.649	251.361	0.71	20.37
Food, drinks, tobacco	3433.787	191.030	0.83	40.18
General chemicals	1577.824	65.486	0.88	61.16
Plastics	150.894	9.562	0.55	10.17
Cement & building materials	9847.984	674.732	0.48	7.47
Vehicles & machinery	-776.982	25.673	0.82	35.42
Furniture, textiles, shoes	1793.983	37.607	0.61	12.89
Electrical Equipment	394.582	13.334	0.78	28.52

TABLE 6.9: TONNAGE OF COMMODITIES BY ROAD REGRESSED ON EMPLOYMENT BY INDUSTRY

Standard Regions
$F(.01) = 11.26$

Dependent Variables	Constant b_o	Coefficient b	R^2	F (b)
O_i				
Coal & coke products	6237.454	144.342	0.68	17.329
Steel	1208.791	102.542	0.67	16.315
Alloys	350.387	26.881	0.73	21.592
Iron, ores, scraps	1499.006	259.806	0.84	42.216
Food, drinks, tobacco	3357.787	190.187	0.80	32.629
Chemicals	1549.800	64.344	0.81	35.632
Cement & building materials	9910.629	654.039	0.46	7.022
Vehicles & CKD	-425.711	21.067	0.69	18.179
Furniture, textiles, shoes	1645.960	39.078	0.63	13.909
Electrical equipment	410.665	13.009	0.78	29.772
D_j				
Coal & coke products	7619.416	102.798	0.45	6.612
Alloys	327.690	28.582	0.76	25.955
Iron, ores, scrap	1588.371	249.893	0.80	32.590
Food, drinks & tobacco	3444.186	188.837	0.83	39.909
Chemicals	1561.045	64.110	0.87	55.723
Plastics	151.343	9.460	0.55	10.062
Cement & building materials	9492.638	670.361	0.48	7.492
Vehicles & CKD	-732.707	24.249	0.80	33.209
Furniture, textiles, shoes	1790.341	37.601	0.69	12.938
Electrical equipment	397.766	13.186	0.77	28.100

TABLE 6.10: TONNAGE OF COMMODITIES BY RAIL REGRESSED ON EMPLOYMENT BY INDUSTRY

Standard Regions
$F(.01) = 11.26$

Dependent variables	Constant b_o	Coefficient b	R^2	F
O_i				
Coal & coke products	-787.443	354.071	0.87	54.043
Steel	-167.931	41.280	0.75	24.254
Vehicles & CKD	-27.481	1.249	0.58	10.951
Electrical equipment	-4.839	0.168	0.88	62.333
D_j				
Coal & coke products	3375.327	228.569	0.84	45.003
Steel	292.915	25.966	0.64	14.101
Alloys	44.810	2.601	0.57	10.76
Food, drinks & tobacco	-9.229	2.256	0.71	19.162
Chemicals	16.764	1.376	0.68	11.215
Vehicles & CKD	-44.294	1.420	0.83	37.979
Electrical equipment	-3.441	0.149	0.92	103.257

with an $R^2 = 0.88$ for O_i and $R^2 = 0.92$ for D_j. Satisfactory models are also derived for steel traffic ($R^2 = 0.64$ and 0.75 for O_i and D_j respectively).

6.1.4 Other Variables

Production of steel, petroleum consumption, coal consumption and cement deliveries for regions were used as explanatory variables for movements of steel, fuel oil and petroleum, coal and coke products and cement and building materials respectively. Poor results were obtained for regressions of O_i, D_j of coal and coke products on coal consumption. The remaining variables yielded satisfactory results (see Table 6.11).

(i) Cement Deliveries for Regions

As seen from Table 6.11, strong models are obtained for movement of cement and building materials by road plus rail and by road. It is worth noting that this measure of distribution has explained D_j better than O_i traffic.

(ii) Petroleum Consumption

The R^2 and F values obtained for road plus rail and road models are satisfactory. The results for rail are poor as only a small proportion of the commodity group fuel oil and petroleum is transported by rail.

TABLE 6.11: TONNAGE OF SELECTED COMMODITIES REGRESSED ON SELECTED X VARIABLES

Standard Regions
$F(.01) = 11.26$

* Non Significant

Dependent Variables	Independent Variables	Constant b_o	Coefficient b	R^2	F
Cement & building materials ROAD plus RAIL	Cement Deliveries				
O_i		2014.267	16.136	0.97	279.60
D_j		1787.443	16.306	0.98	337.53
ROAD					
O_i		1700.725	16.037	0.97	291.41
D_j		1496.436	16.193	0.98	343.38
Fuel oil & petroleum products ROAD plus RAIL	Petroleum Consumption				
O_i		-634.966	1.647	0.96	245.29
D_j		679.645	1.440	0.94	133.86
ROAD					
O_i		-921.422	1.376	0.95	176.58
D_j		-349.209	1.285	0.95	169.45
RAIL					
O_i		286.476	0.271	0.57	10.81
D_j		1027.398	0.154	0.30	* 3.43
Steel ROAD plus RAIL	Production of Steel Ingot.				
O_i		2479.245	1.129	0.64	12.93
D_j		3937.753	0.555	0.18	* 1.81
ROAD					
O_i		2518.626	0.693	0.55	9.44
D_j		744.553	0.129	0.09	* 0.80
RAIL					
O_i		-39.421	0.436	0.97	386.71
D_j		414.931	0.258	0.73	22.26

(iii) Steel Production

The models are less robust than those derived with the two other variables. This measure of production best explains originating traffic as well as rail movement of steel. For rail, 97 and 73 per cent of the variation is explained in O_i and D_j respectively while only 55 per cent of the variation in O_i by road is explained. The D_j equation by road is not significant.

Summary

The results derived from regressions on resident population and employment agree with findings reported in previous studies. The net output variables, albeit the lack of a good compatibility between the levels of aggregation in the dependent and independent variables, yielded reasonable results for a substantial number of commodity groups. In general, these three types of variables yielded higher R^2 for terminating traffic. However, the converse result (higher R^2 for O_i) was obtained for commodity group whose production is relatively more localized such as steel, iron ores, coal and coke products. This latter comment does not include the variable population as it yielded poor fits for these particular commodity groups. The three other variables viz. cement deliveries for regions, petroleum consumption, steel production proved useful in explaining freight movement for their corresponding industries. It is appropriate to stress on the main findings.

(i) The compatibility between the levels of aggregation in the dependent and independent variables is a key factor of success in the modelling exercise. For example the employment variables for which the best match between the definition of the commodity groups and the industry groups explained satisfactorily most commodity movement.

(ii) Different levels of explanation are achieved for originating and terminating traffic depending on which variable is used.

(iii) As the rail models were generally poor, a mispecification of the mathematical form was suspected. Log and semi-log transforms were unsuccessfully tested. In view of these results as well as the relative success of the rail models with employment and the good fit obtained with the variable steel production, it may be suggested that specific variable like the latter would be more appropriate to calibrate freight movement by rail.

6.2 Correlation Matrices

6.2.1 Net Output and Net Capital Expenditure Variables

For reasons of clarity the original matrix of intercorrelations is subdivided into three sub-matrices. Tables 6.12, 6.13 and 6.14 represent respectively, correlations between net output in different industries, net output and net capital expenditure and net capital expenditure in different industries.

Correlations between Net Output Variables

As seen from Table 6.12, the correlation coefficients are in general greater than 0.30. The pattern of intercorrelations may be summarized as follows: the variables making up the engineering industries (X'5, X'6, X'7), the metal industries (X'4, X'9) and the leather goods and clothing industries (X'11, X'12) correlate highly with one another and similarly with variables in other groups. In addition, the variables X'1, X'2 and X'3 respectively food, coal and petroleum and chemical industries correlate in a similar way with variables from other groups but do not present a significant correlation with one another.

TABLE 6.12: CORRELATION BETWEEN NET OUTPUT MEASURES IN MANUFACTURING INDUSTRIES

Standard Regions

ORDERS		X'1	X'2	X'3	X'4	X'5	X'6	X'7	X'8	X'9	X'10	X'11	X'12	X'13	X'14
III	(X'1)	1	0.02	0.76	-0.35	0.84	0.86	0.81	0.52	0.38	0.15	0.77	0.70	0.52	0.84
IV	(X'2)		1	0.90	-0.03	0.75	0.70	0.71	0.30	0.33	0.34	0.71	0.57	0.56	0.75
V	(X'3)			1	-0.28	0.80	0.76	0.77	0.41	0.32	0.23	0.75	0.65	0.57	0.81
VI	(X'4)				1	-0.09	-0.27	-0.09	0.23	0.47	0.10	-0.27	-0.45	0.30	-0.25
VII	(X'5)					1	0.87	0.91	0.77	0.71	0.18	0.85	0.67	0.75	0.92
VIII	(X'6)						1	0.96	0.57	0.46	-0.12	0.77	0.58	0.42	0.98
IX	(X'7)							1	0.71	0.62	-0.15	0.76	0.53	0.57	0.97
X,XI	(X'8)								1	0.93	-0.11	0.58	0.21	0.82	0.61
XII	(X'9)									1	0.06	0.48	0.13	0.84	0.53
XIII	(X'10)										1	0.28	0.37	0.39	-0.04
XIV	(X'11)											1	0.90	0.63	0.83
XV	(X'12)												1	0.40	0.64
XVI	(X'13)													1	0.52
XVII	(X'14)														1

For convenience, it is referred in the text to the X' variables rather than the order number.

TABLE 6.13: CORRELATION BETWEEN NET OUTPUT AND NET CAPITAL EXPENDITURE IN RESPECTIVE INDUSTRIES

Standard Regions

ORDERS		X'15	X'16	X'17	X'18	X'19	X'20	X'21	X'22	X'23	X'24	X'25	X'26	X'27	X'28
III	(X'1)	0.93	0.15	0.31	-0.39	0.81	0.83	0.79	0.63	0.34	0.03	0.60	0.89	0.62	0.87
IV	(X'2)	0.49	0.71	0.76	0.04	0.71	0.65	0.79	0.56	0.30	0.31	0.61	0.75	0.70	0.82
V	(X'3)	0.67	0.48	0.77	-0.22	0.79	0.76	0.87	0.59	0.29	0.19	0.58	0.85	0.63	0.88
VI	(X'4)	-0.29	0.24	-0.09	0.70	0.00	-0.27	0.15	0.14	0.52	0.13	-0.21	-0.27	0.24	-0.28
VII	(X'5)	0.87	0.14	0.34	-0.20	0.97	0.80	0.91	0.84	0.66	0.15	0.71	0.89	0.91	0.91
VIII	(X'6)	0.84	0.21	0.24	-0.31	0.85	0.94	0.93	0.79	0.39	-0.18	0.52	0.80	0.70	0.97
IX	(X'7)	0.83	0.26	0.25	-0.27	0.89	0.91	0.97	0.89	0.56	-0.18	0.48	0.79	0.78	0.94
X,XI	(X'8)	0.74	-0.06	-0.01	-0.27	0.73	0.53	0.67	0.91	0.93	-0.13	0.32	0.48	0.77	0.53
XII	(X'9)	0.58	0.00	-0.01	0.05	0.70	0.39	0.57	0.84	0.98	0.03	0.33	0.39	0.85	0.45
XIII	(X'10)	0.03	0.21	0.45	0.38	0.18	-0.19	-0.08	-0.18	0.06	0.95	0.62	0.35	0.32	0.06
XIV	(X'11)	0.78	0.22	0.31	-0.34	0.81	0.69	0.79	0.70	0.43	0.28	0.87	0.89	0.76	0.84
XV	(X'12)	0.62	0.16	0.31	-0.40	0.61	0.51	0.58	0.37	0.10	0.41	0.88	0.86	0.49	0.68
XVI	(X'13)	0.64	0.26	0.36	-0.08	0.70	0.34	0.59	0.73	0.87	0.33	0.54	0.58	0.84	0.51
XVII	(X'14)	0.82	0.23	0.29	-0.29	0.89	0.90	0.95	0.80	0.45	-0.06	0.62	0.85	0.78	0.98

TABLE 6.14: CORRELATION BETWEEN NET CAPITAL EXPENDITURE MANUFACTURING INDUSTRIES

Standard Regions

ORDERS		X'15	X'16	X'17	X'18	X'19	X'20	X'21	X'22	X'23	X'24	X'25	X'26	X'27	X'28
III	(X'15)	1	0.00	0.16	-0.51	0.82	0.79	0.79	0.78	0.55	-0.09	0.54	0.79	0.69	0.82
IV	(X'16)		1	0.65	0.18	0.13	0.20	0.32	0.21	0.04	0.16	0.14	0.22	0.23	0.31
V	(X'17)			1	0.10	0.36	0.32	0.43	0.13	0.01	0.40	0.28	0.45	0.26	0.42
VI	(X'18)				1	-0.06	-0.27	-0.27	-0.25	0.03	0.42	-0.08	-0.23	0.08	-0.26
VII	(X'19)					1	0.83	0.90	0.81	0.64	0.16	0.65	0.88	0.88	0.88
VIII	(X'20)						1	0.92	0.75	0.33	-0.25	0.38	0.74	0.58	0.91
IX	(X'21)							1	0.86	0.52	-0.10	0.51	0.81	0.76	0.95
X,XI	(X'22)								1	0.82	-0.22	0.37	0.58	0.32	0.75
XII	(X'23)									1	0.02	0.27	0.34	0.80	0.38
XIII	(X'24)										1	0.65	0.34	0.30	0.01
XIV	(X'25)											1	0.82	0.70	0.65
XV	(X'26)												1	0.73	0.89
XVI	(X'27)													1	0.75
XVII	(X'28)														1

Correlation Between Net Output and Net Capital Expenditure .

The correlation coefficients are in general above 0.70. It is not surprising that this surrogate for investment correlates with net output. The group pattern described above is also true for this matrix.

Correlation Between Net Capital Expenditure

A similar pattern to that of Table 6.12 and 6.13 is observed in Table 6.14. However, the values of the correlation coefficient are relatively lower.

6.2.2 Employment by Industry

The correlation between employment measures in different industries are shown in Table 6.15. Employment in the coal and coke industry (E1) and in the steel industry (E2) correlate with one another but only slightly with the remaining variables. Employment in alloys (E3) and iron-ores (E4) are associated with each other and with the cement and building materials industries (E10). They also correlate, but to a lesser degree, with employment in the vehicle industry (E11). Surprisingly, however, alloys and iron-ores do not show a high correlation with steel, which belong to the same order "metal industries".

TABLE 6.15: CORRELATION BETWEEN EMPLOYMENT VARIABLES

Standard Regions

	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13
E1	1	0.74	0.20	0.19	0.35	0.43	0.16	0.46	0.20	0.14	0.20	0.00	0.39
E2		1	0.30	0.30	0.22	0.17	0.28	0.32	0.36	0.00	0.20	0.00	0.10
E3			1	0.77	0.26	0.03	0.17	0.61	0.35	0.86	0.52	0.00	0.49
E4				1	0.10	0.14	0.02	0.36	0.39	0.73	0.30	0.22	0.26
E5					1	0.20	0.86	0.82	0.28	0.65	0.67	0.90	0.86
E6						1	0.03	0.36	0.54	0.04	0.09	0.10	0.14
E7							1	0.84	0.07	0.59	0.85	0.56	0.84
E8								1	0.05	0.85	0.91	0.35	0.93
E9									1	0.15	0.03	0.41	0.15
E10										1	0.95	0.78	0.90
E11											1	0.80	0.93
E12												1	0.65
E13													1

Key:

E1 MLH 101, 109, 261
E2 MLH 311, 312
E3 MLH 321, 322
E4 MLH 313, 323
E5 Order III
E6 MLH 102, 103
E7 MLH 271, 272, 273, 274, 277, 279
E8 MLH 276, 492, 496
E9 MLH 278
E10 Order XVI
E11 Order XI
E12 Orders XIII, XIV, XV and MLH 472, 473, 474, 475, 476
E13 Order IX

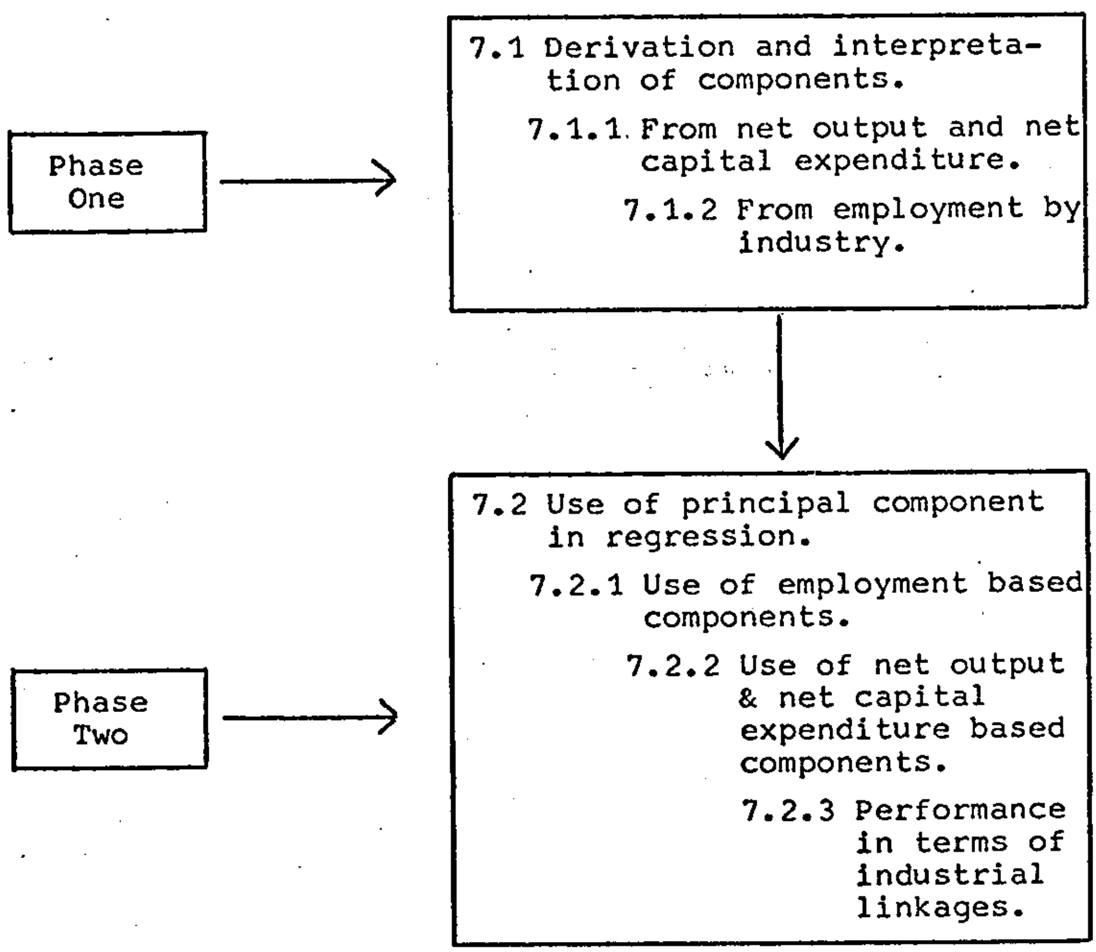
6.2.3 Discussion

The foregoing descriptions have shown that the patterns of intercorrelations are characterized by a high frequency of strong correlations. In particular, the first order correlation coefficients are markedly high between the so called "related industries". This means that attempt to describe industrial linkages by means of stepwise regression on measures in different industrial orders as did Chisholm and O'Sullivan (1973) is likely to fail because of multi-collinearity. Principal component may help isolate the appropriate groups that can be used in stepwise regression.

CHAPTER 7

COMMODITY ANALYSIS (II): USE OF PRINCIPAL COMPONENTS AND DISCUSSION OF RESULTS FOR PART II

The empirical work described in Section 7.1 and 7.2 is summarized in Figure 7.1 below.



7.1 Derivation and Identification of Principal Components

7.1.1 Net Output and Net Capital Expenditure Variables[1]

The results of the principal component analysis for the twenty-eight value variables are set out in Table 7.1. The respective eigenvalues of the four components extracted are 16.30, 3.78 and 2.07. In all, about 92 per cent of the variation is explained by these first few components.

Component EOI

This component accounted for over half the variance in the data set. It is the most complex, since, as might be expected from principal component analysis, a large number of variables correlate significantly with it. The variable most highly associated with this component are: (i) the engineering industry variables viz. net output variables X'_5 to X'_7 and their respective expenditure variables X'_{19} to X'_{21}; (ii) the food industry variables X'_1 and X'_{15}; (iii) the clothing and footwear variables X'_{14} and X'_{26}. None of the negative association is significant at the one per cent level. Nevertheless, it is pertinent to note that the highest negative correlation is with the metal industry variables X'_4 and X'_{18}.

---

(1) Referred to as "value variables" thereafter.

TABLE 7.1: PRINCIPAL COMPONENTS FROM NET CAPITAL EXPENDITURE & NET OUTPUT IN MANUFACTURING INDUSTRY

Standard Regions

Variables	Component EOI (Loadings)	Component EOII (Loadings)	Component EOIII (Loadings)	Component EOIV (Loadings)
X'_{1}	0.842	0.186	0.174	0.016
X'_{2}	0.676	0.143	0.286	0.649
X'_{3}	0.757	0.144	0.227	0.479
X'_{4}	-0.344	0.550	-0.034	0.162
X'_{5}	0.828	0.491	0.235	0.051
X'_{6}	0.959	0.192	-0.087	0.084
X'_{7}	0.892	0.384	-0.120	0.156
X'_{8}	0.415	0.881	-0.075	-0.092
X'_{9}	0.324	0.925	0.034	-0.076
X'_{10}	-0.063	0.047	0.941	0.146
X'_{11}	0.728	0.304	0.411	0.135
X'_{12}	0.606	0.000	0.562	0.093
X'_{13}	0.279	0.843	0.350	0.238
X'_{14}	0.937	0.257	0.023	0.121
X'_{15}	0.753	0.435	0.066	-0.090
X'_{16}	0.136	0.008	0.054	0.947
X'_{17}	0.281	-0.055	0.322	0.732
X'_{18}	-0.199	-0.023	0.236	0.080
X'_{19}	0.846	0.456	0.219	0.021
X'_{20}	0.942	0.133	-0.180	0.103
X'_{21}	0.880	0.346	-0.054	0.268
X'_{22}	0.640	0.709	-0.180	0.141
X'_{23}	0.232	0.966	0.019	-0.004
X'_{24}	-0.089	0.009	0.968	0.122
X'_{25}	0.542	0.155	0.759	0.030
X'_{26}	0.829	0.173	0.442	0.109
X'_{27}	0.621	0.662	0.322	0.106
X'_{28}	0.947	0.178	0.101	0.703
EIGEN VALUES	16.30	3.78	3.57	2.07
PCT OF VARIANCE	58.2	13.5	12.8	7.4
CUM PCT	58.2	71.8	84.5	91.9

Extreme Values of Principal Component Scores

	EO1	EO2	EO3	EO4
HIGHEST :	S. East, Scot.	W.Mid. S.East	Y-H, E.Mid.	N.West, Wales
LOWEST :	W. Mid.,Wales	E. Ang.North	Wales,North	Scot.,E. Ang.

Given the type of industries with which this component is associated, it may, tentatively be said to represent the "dispersed industries". The pattern of component scores supports this interpretation as the highest scores, are found in the South-East and Scotland which have diversified industries. On the other hand, the lowest scores are found in Wales and the West Midlands, regions that are highly specialised[1].

Component EOII

The highest associations are with X'_{9} and X'_{23}, respectively net output and net capital expenditure in metal goods n.e.s., immediately followed by the ship-building and the vehicle industries measures X'_{8} and X'_{22}. In addition a substantial association is observed with the metal industry variable (X'_{4}). These correlations suggest that this component may be considered as an index for localized industries. The highest scores are observed for the West Midlands and the South-East and the lowest for East Anglia and the North.

Component EOIII

This component is significantly associated with the variables X'_{10}, X'_{24} and X'_{25} respectively net output and net capital expenditure for the textiles industry and net capital expenditure in the leather

---

(1) Evidence may be found in Brown (1972), Keeble (1976). This stands for the rest of the discussion too.

goods and for industry. On these grounds, this component appears in the dimension for the textiles and allied industries. The component scores are gratifying; the highest are found in two well known textiles producing areas namely the East Midlands, Yorkshire and Humberside. The lowest scores appear in Wales and the North.

Component EOIV

The variables that correlate with this component are $X'16$ and $X'17$, net capital expenditure in the coal and petroleum industry and in the chemical industry respectively. These two industries are spatially associated and functionally inter-related, Chisholm and O'Sullivan (1973), Lever (1975). Thus the component may, tentatively be advanced as a measure of these industries. The component scores exposed this suggestion. The highest scores are found in the North West, one of the leading centres of Britain's Chemical industry and in Wales, a coal producing site of major oil import terminals.

7.1.2 Employment Variables

Principal component analysis has been carried out on twenty variables at the S.I.C. order level rather than on the thirteen variables used previously. This procedure may be justified in that firstly the thirteen employment variables do not give a complete idea about regional employment as some industries are not accounted for and that secondly, to consider all the M.L.H. sub-divisions would result in a considerably larger matrix. The results are set out in Table 7.2.

Component CE1

The component is negatively associated with the metal industry variable (X4) and mining and quarrying (X18). It is positively associated with (i) transport and communication (X19), paper, printing and publishing (X20), instrument engineering (X6). The component may be said to represent dispersed industries. The pattern of scores as reported in Table 7.2 gives support to such a description; highest scores appear in the South-East and the North-West whilst the lowest are found in the West Midlands and the North.

Component CE2

Four variables load significantly on this component. These are, by decreasing order of association, X10 (metal goods), X13 (bricks, pottery and glass), X9 (vehicles) and finally X4 (metal industry). It is rather difficult to give an interpretation to this component.

These industries are relatively concentrated in few regions and so it may, tentatively, be suggested that

TABLE 7.2: PRINCIPAL COMPONENTS DERIVED FROM EMPLOYMENT VARIABLES BY INDUSTRY

Standard Regions

Variables	Component CE1 (Loadings)	Component CE2 (Loadings)	Component CE3 (Loadings)
X 1	0.873	0.214	0.137
X 2	0.795	0.207	0.353
X 3	0.813	0.185	0.326
X 4	-0.182	0.777	-0.002
X 5	0.846	0.487	0.092
X 6	0.935	0.153	-0.224
X 7	0.854	0.416	-0.138
X 8	0.468	-0.159	-0.062
X 9	0.530	0.786	-0.012
X 10	0.357	0.920	-0.060
X 11	0.109	-0.015	-0.915
X 12	0.885	0.387	0.190
X 13	0.480	0.853	0.129
X 14	0.876	0.070	0.437
X 15	0.959	0.231	-0.060
X 16	0.824	0.490	-0.043
X 17	0.746	-0.026	-0.555
X 18	-0.169	-0.069	0.309
X 19	0.946	0.234	-0.086
X 20	0.953	0.099	-0.099
EIGEN VALUES	13.05	2.73	2.13
PCT OF VARIANCE	65.3	13.7	10.7
CUM PCT	65.3	78.9	89.6

Extreme Values of Principal Component Scores

	CE1	CE2	CE3
HIGHEST :	S. East, N. East	W. Midlands, S.East	N.West, E.Midlands
LOWEST :	W.Mid., North	S.West, E.Anglia	E. Anglia, S.East

this component represents localized industries.

Component CE3

The third component clearly represents the textile industry as the only variable significantly associated with it is XII (employment in the textile industry). The pattern of scores as seen from the extreme values in Table 7.2 is clear too. Both approaches to principal component analysis have shown that it is a measure of dispersed industry that is first extracted. It has accounted for 58 and 65 per cent of the variability in the two sets of data. Additionally, the components isolated specialisation in metal, textiles and chmeical industries. These major groups coincide with those that have been identified to have spatial or industrial association by other researchers (Lever, (1973, 1975); Smith (1975); Keeble (1976)).

7.2 Regression on Principal Components

7.2.1 Employment Based Components

Road Plus Rail

The stepwise regression results are set out in Table 7.3. The relationships derived for terminating traffic achieved a satisfactory level of explanation with the median $R^2 = 0.82$. All the equations with the exception of that for coal and coke products are significant

TABLE 7.3: TONNAGE OF COMMODITIES BY ROAD PLUS RAIL REGRESSED ON EMPLOYMENT BASED COMPONENTS

Standard Regions
$F(.01)$,2X's= 9.55
$F(.01)$,1X =11.26

Dependent Variables	Constant b_o	b_1	b_2	b_3	R^2	F
O_i						
Coal & coke products	1444.341			131.385	0.47	6.810
Steel	4036.446	-11.446	25.558		0.47	6.286
Alloys	352.582	-0.937*	2.967		0.69	10.549
Iron, ores, scrap	1991.040	-5.603	17.194		0.86	22.253
Food, drinks & tobacco	-48516.712	186.303		-351.175	0.92	44.122
Sand, clays & stones	19640.534	24.721			0.74	21.190
Chemicals	1808.119	3.294			0.53	9.232
Plastics	184.643	0.360			0.54	9.712
Fertilizers	1059.578	1.512			0.61	12.894
Cement & building materials	581.952	1.804			0.56	10.467
Fuel, oil & petroleum	1939.100	12.099			0.93	119.592
Vehicles & CKD	-369.691		5.963		0.86	51.450
Electrical equipment	538.645	5.240		27.224	0.79	13.708
D_j						
Coal & coke products	16459.077			98.308	0.48	7.416
Steel	2105.505	-10.396	26.343		0.76	11.553
Alloys	177.346	-0.819	3.254		0.82	16.406
Iron, ores, scrap	1087.836	-6.185	18.032		0.86	22.731
Food, drinks & tobacco	7782.558	14.578			0.77	26.989
Sand, clays, stones	17077.203	28.729			0.85	46.229
Chemicals	1577.688	3.658			0.71	19.534
Plastics	116.419	0.464			0.75	25.102
Fertilizers	896.487	1.759			0.75	24.255
Cement & building materials	6408.416	35.483			0.97	283.560
Fuel, oil & petroleum	2794.694	10.781			0.94	149.113
Vehicles & CKD	-592.364		6.606		0.90	78.784
Furniture, textiles, shoes	612.388	5.274		25.476	0.81	15.921
Electrical equipment	64.072	2.085			0.87	23.784

* Significant at the .05 level - All the remaining coefficients are significant at the .01 level

b_1 coefficient for CE1 "non-localised industries"
b_2 coefficient for CE2 "localised industries"
b_3 coefficient for CE3 "textiles industries"

TABLE 7.4: TONNAGE OF COMMODITIES BY ROAD REGRESSED ON EMPLOYMENT VARIABLES BASED COMPONENTS

Standard Regions
F(.01),2X's= 9.55
F(.01),1X =11.26

Dependent Variables	Constant b_o	b_1	b_2	b_3	R^2	F
O_i						
Coal & coke products	8224.974			49.069	0.52	9.76
Steel	2951.560	-10.248	24.307		0.69	11.30
Alloys	256.493		1.239		0.50	8.60
Iron, ores, scrap	1517.255	-5.451	16.847		0.90	31.88
Food, drinks & tobacco	7794.279	14.351			0.72	21.22
Sand, clays, stones	18307.066	24.573			0.75	24.33
Chemicals	348.762	3.760	19.497		0.83	17.43
Plastics	181.533	0.361			0.54	9.72
Fertilizers	1015.914	16.052	-9.645		0.73	12.70
Fuel oil & petroleum	1147.081	10.233			0.94	148.67
Vehicles & CKD	-358.574		5.664		0.84	43.13
Furniture, textiles, shoes	513.119	5.238		26.201	0.76	11.66
Electrical equipment	-37.030	2.075		1.759	0.86	52.80
D_j						
Coal & coke products	5290.793	8.505	49.167		0.81	16.46
Steel	1671.573	-10.143	26.761		0.88	27.92
Alloys	137.351	-0.727	2.959		0.84	18.38
Iron, ores, scrap	1518.380	-5.538	17.019		0.94	56.00
Food, drinks, tobacco	7738.584	14.417			0.77	26.95
Sand, clays, stones	11770.313		27.963		0.77	27.12
Chemicals	1456.858	2.790	1.803		0.70	8.22
Plastics	117.113	0.459			0.75	24.69
Fertilizers	804.481	1.757			0.77	27.86
Cement & building materials	6107.016	35.216			0.97	280.90
Fuel oil & petroleum	1558.961	9.595			0.95	167.42
Vehicles & CKD	-549.583	0.420	5.417		0.89	31.18
Furniture, textiles, shoes	612.438	5.265	25.503		0.81	15.87
Electrical equipment	69.992	2.063			0.86	49.77

All coefficient are significant at the .01 level

b_1 coefficients for CE1 "non-localised industries"
b_2 coefficients for CE2 "localised industries"
b_3 coefficients for CE3 "textiles industries"

TABLE 7.5: TONNAGE OF COMMODITIES BY RAIL REGRESSED ON EMPLOYMENT BASED COMPONENTS

Standard Regions
F(.01),2X's= 9.55
F(.01),1X =11.26

Dependent Variables	Constant b_o	b_1	b_2	b_3	R^2	F
O_i						
Fuel oil & petroleum products	385.772	2.153			0.58	11.913
Vehicles & CKD	-6.928	0.151			0.65	14.972
Electrical equipment	-7.923	0.024			0.86	49.33
D_j						
Coal & coke products	7734.594			56.052	0.40	5.581
Alloys	36.208	-0.092	0.293		0.52	5.690
Food, drinks	50.431	0.159			0.55	10.156
Sand, clays, stones	179.985	1.923			0.82	16.401
Chemicals	6.148	0.093			0.68	17.424
Plastics	-0.833	0.049			0.60	12.131
Fuel oil & petroleum	849.926		3.362		0.56	10.237
Vehicles & CKD	-34.956		0.368		0.93	111.179
Furniture, textiles & shoes	-2.088	0.009			0.75	24.297
Electrical equipment	-6.165	0.021			0.91	82.813

All coefficient are significant at the .01 level

b_1 coefficient for CE1 "non-localised industries"
b_2 coefficient for CE2 "localised industries"
b_3 coefficient for CE3 "textiles industries"

at the one per cent level. The originating freight models achieved a more modest level of explanation. In this case, four models viz. coal and coke products, steel, chemicals and plastics are significant only at the five per cent level.

Road

The results are shown in Table 7.4. The level of explanation achieved is higher than that observed for the aggregate mode Road plus Rail. Both the O_i and D_j relationships are satisfactory with median $R^2 = 0.76$ and $R^2 = 0.84$ for O_i and D_j respectively. Only two equations did not achieve the one per cent level of significance that is O_i, alloys and D_j chemicals.

Rail

As seen from Table 7.5, the success for this mode is limited. For O_i traffic for instance, only three models are satisfactory.

7.2.2 Net Output and Net Capital Expenditure Based Components

Road Plus Rail

As seen from Table 7.6, for either O_i or D_j half of the commodity groups considered could be explained satisfactorily. The level of explanation is

TABLE 7.6: TONNAGE OF COMMODITIES BY ROAD PLUS RAIL REGRESSED ON COMPONENTS FROM NET OUTPUT AND NET CAPITAL EXPENDITURE

Standard Regions
F(.01),2X's= 9.55
F(.01),1X =11.26

Dependent Variables	Constant b_0	b_1	b_2	b_3	b_4	R^2	F
O_i							
Coal & coke products	19443.633		-6.915*		9.115	0.42	5.945*
Chemicals	-55931.783				308.788	0.92	94.184
Plastics	-3261.576	5.114	5.875			0.92	41.844
Fertilizers	548.450				5.103	0.89	67.490
Vehicles & CKD	625.903			5.210	-4.481	0.84	18.945
Furniture, textiles, shoes	4091.997	-4.050		20.215		0.78	12.598
Electrical equipment	4190.104	-12.018		30.568		0.79	13.598
D_j							
Chemicals	5365.154	3.212				0.65	15.175
Plastics	-3166.886	7.898				0.94	137.94
Fertilizers	279.891				5.817	0.92	92.545
Fuel, oil & petroleum	5925.607	-4.145		19.629		0.84	19.459
Vehicles & CKD	-4169.051	8.221				0.96	228.754
Furniture, textiles, shoes	2815.797	2.065				0.81	35.303
Electrical equipment	811.554		1.091			0.69	18.44

* Significant at the .05 level - All the remaining coefficients are significant at the .01 level

b_1 coefficient for EOI "dispersed industries"
b_2 coefficient for EOII "localised industries"
b_3 coefficient for EOIII "textiles industries"
b_4 coefficient for EOIV "coal, petroleum & chemicals"

TABLE 7.7: TONNAGE OF COMMODITIES REGRESSED ON COMPONENTS FROM NET OUTPUT AND NET CAPITAL EXPENDITURE

Standard Regions
F(.01),2X's= 9.55
F(.01),1X =11.26

Dependent Variables	Constant b_o	b_1	b_2	b_3	b_4	R^2	F
ROAD							
O_i							
Steel	3168.281	-4.736	8.502			0.66	12.731
Chemicals	2622.543		17.451			0.83	41.938
Plastics	2744.874	19.797	-9.837			0.79	13.369
Fertilizers	389.256			5.081		0.89	67.510
Cement & building materials	1246.982	16.052		-9.645		0.80	14.757
Fuel, oil & petroleum	286.090	2.119				0.92	98.368
Vehicles & CKD	106.307	-1.503	5.324			0.79	13.197
Furnitures, textiles, shoes	2895.155	2.808	2.600			0.71	12.071
Electrical equipment	3409.452	1.926		-1.534		0.78	20.656
D_j							
Coal & coke products	7735.335	-4.589			30.594	0.58	9.976
Alloys	1336.530			5.941		0.79	13.672
Sand, clays, stones	13930.284	-10.702			78.554	0.73	12.908
Chemicals	5297.877	3.189				0.65	15.180
Plastics	-1377.039	10.605				0.95	70.189
Fertilizers	277.025				5.657	0.92	92.64
Fuel, oil & petroleum	4388.075	-4.029		19.779		0.87	24.945
Vehicles & CKD	-4752.950	8.225				0.97	294.610
Furniture, textiles, shoes	2599.474	1.310		2.340		0.80	15.305
Electrical equipment	841.512		1.032			0.70	16.466

All coefficients are significant at the .01 level

b_1 coefficient for EOI "non-localised industries"
b_2 coefficient for EOII "localised industries"
b_3 coefficient for EOIII "textiles industries"
b_4 coefficient for EOIV "coal, petroleum & chemical industries"

TABLE 7.8: TONNAGE OF COMMODITIES REGRESSED ON COMPONENTS FROM NET OUTPUT AND NET CAPITAL EXPENDITURE

Standard Regions
F(.01),2X's= 9.55
F(.01),1X =11.26

Dependent Variables	Constant b_o	b_1	b_2	b_3	b_4	R^2	F
RAIL O_i							
Chemicals	36.042	-0.832			0.578	0.77	12.057
Cement & building materials	405.877	0.462	-0.558			0.53	5.026*
Fuel, oil & petroleum	-110.988			8.241		0.53	9.370*
Vehicles & CKD	-39.163		0.125		0.324	0.66	6.902*
Furniture, leather, shoes	1.050	0.003				0.41	5.567*
Electrical equipment	-5.782	0.016				0.51	8.578*
D_j							
Coal & coke products	3201.675	7.856				0.45	6.750*
Food, drinks & tobacco	30.646		0.359			0.54	9.434*
Sands, clays, stone	404.680	0.80				0.42	5.924*
Chemicals	59.572	0.218			0.218	0.48	7.618*
Plastics	-545.819				3.516	0.88	64.371
Fertilizers	2.914				0.160	0.84	44.430
Cement & building materials	97.862	0.356				0.49	7.812*
Fuel, oil & petroleum	362.906	1.733				0.82	37.062
Vehicles & CKD	325.477	-1.758	3.670			0.75	11.992
Furniture, textiles, shoes	-273.659			1.408		0.71	20.441

* All coefficients are significant at the .01 level

b_1 coefficient for EOI "dispersed industries"
b_2 coefficient for EOII "localised industries"
b_3 coefficient for EOIII "textiles industries"
b_4 coefficient for EOIV "coal, petroleum & chemical industries"

high with the median $R^2 = 0.89$ for O_i and $R^2 = 0.84$ for D_j.

Road

From the results set out in Table 7.7, a relatively higher performance may be noticed for this mode. The variance explained in both originating and terminating traffic is satisfactory ranging from 58 per cent for D_j coal and coke products to 97 per cent for D_j vehicles and CKD.

Rail

As seen from Table 7.8, it is again for rail that the least satisfactory models are obtained. Even the best results reports are, for the majority of commodity groups, significant at the 5 per cent level only. Only six relationships are significant at the one per cent level and have R^2 greater than 0.70.

7.2.3 Performance in terms of Industrial and Spatial Associations

The purpose has been to find out which of the industrial dimensions described by the components, determines primarily the movement of a given commodity group. The employment based component yielded, in general, more plausible models than the net output and net capital expenditure based components. The relationships may be summarised in function of the common components they have included. For employment based component, the groups are:

(i) Metal Industries

For either O_i or D_j, the measure denoted "localised industries" entered the regression first, followed by the measure of "dispersed industries". The latter component has however associated negative coefficient. The inclusion of both components is plausible. The negative sign is probably due to the great concentration of metal whose impact in a small sample is important.

(ii) Coal and Coke Products

A single component identified as representing the textile industry was picked up by the stepwise regression. This is hardly surprising as there is spatial association between the textile and the mining industries. It is pertinent to note that high scores for this component have been derived for mining regions, which do in fact produce textiles too, for example, the North and the East Midlands. Chisholm and O'Sullivan (1973) indicated similar results.

(iii) Electrical Equipment

The O_i relationships was explained primarily by the measure representing textiles and then by the component "dispersed industries" whereas the attraction equation was explained solely by the measure of dispersed industries. The D_j equation is plausible as a number of industries require electrical equipment but the inclusion of the textile measure for O_i does not make an obvious sense.

(v) Vehicles

The movement of this commodity group were reasonably explained by the component localised industries.

(vi) The Remainders in the Commodity Groups

That is petroleum, sand, clay, stone, chemicals, cement and building materials, food and drinks were explained by the measure of dispersed industries.

For the net output and net capital expenditure based component, the results may be summarised as follows.

For the aggregate mode, road plus rail, the component called in the equations for terminating traffic are relatively plausible. The component "dispersed industries" entered in the D_j models for all commodity groups except fertilizers and electrical equipment. These were explained by coal and chemical industries for fertlizers and the component localised industries for electrical equipment. For road, the component representing coal and chemicals did not enter in any of the O_i equations. It entered by itself, however, the D_j model for fertilizers and partly explained sand, clay, stones and plastics. The "dispersed industries" component entered all O_i models but alloys and electrical equipment.

Summary

This exercise has shown that industrial and spatial linkages may be described in freight generation models. However, the investigation has been severely constrained by the sample size. There is a bias due to the domination of most industrial orders by the South-East and to a lesser extent by the North-West.

7.3 Discussion of Results for Part II

7.3.1 Aggregation

(i) Modal Aggregation

The general tendency is, if the transport of a given commodity is largely equalled shared by both modes, then the best results are achieved by the aggregate mode. However, if a commodity group is predominantly moved by one of the two modes, then poor results are obtained for the aggregate mode and the dominant mode yields the best results.

(ii) Commodity Aggregation

The assumption of homogeneity has been made by studies of freight generation at the multi-regional level but the validity of this assumption has not been established (Department of the Environment, 1971; Chisholm and O'Sullivan, 1973). So the objective, here, is to examine whether there are significant differences in the propensity to generate freight tonnage between different industries and group of industries. If no significant differences are found, then aggregation is statistically justified. The test is limited to comparison of regressions of commodity tonnage by road on employment for two reasons. Firstly, these are the most reliable models obtained. Secondly, employment data are available down to the MLM sub-division and so various degrees of aggregation could be experimented with. These were selected on a priority grounds, for example, categories bulk or heavy and light or because of apparent similarities in their simple regression coefficients. The covariance results are set out in Table 7.9.

A variety of commodity groups were first considered so as to attempt a representation of bulk or heavy industries; the groups are defined by their B.R. commodity codes in Table 7.9. The Bartlett χ^2 test for homogeneous variances is significantly different from the $\chi^2_{0.99}$ value for the theoretical distribution thus indicating the presence of heterogeneous

TABLE 7.9: COMPARISON OF REGRESSIONS FOR SELECTED COMMODITY GROUPS: ORIGINATING FREIGHT ON EMPLOYMENT BY INDUSTRY

Standard Regions

F(0.01)	^{2}(0.01)
(4,40)=3.83	(4)=13.28
(2,24)=5.61	(2)= 9.21
(4,44)=3.68	
(2,26)=5.53	

Commodity Groups [1]	Bartlett χ^2 test	Comparison of Slopes F	Comparison of Intercepts F
'Bulk' Commodities (1,2-4,6-8,14,15,20,21)	38.90 (4)	9.10 (4,40)	14.30 (4,44)
Metal Industry (2-4,6,7)	0.00	5.87 (2,24)	14.25 (2,26)
Engineering Industry	5.63 (2)	4.18 (2,24)	7.13 (2,26)
Chemical Industry	0.00	6.84 (2,24)	20.14 (2,26)

(1) In this column, the bracketed numbers refer to B.R. Commodity Groups. In the remainder of the Table, the appropriate degrees of freedom are shown.

variances. This is reinforced by the F tests comparing slopes and intercepts. On the basis of the F tests both hypotheses of identical slopes and intercepts are rejected. Therefore, the commodities groups in this analysis are significantly different from each other in terms of freight generated per employee.

Within this large group, it was decided to examine the metal industries. The resulting χ^2 value is close to zero indicating that the residual variances can be considered as almost perfectly homogeneous. However, the tests of differential slopes and intercepts are significant at the one per cent level suggesting that a common relationship for this grouping is not appropriate. Engineering industries were then examined. Whereas the examination of the first two together viz. vehicles and CKD(24)* and electrical equipment(26) suggested homogeneity with respect to freight generation, the consideration of a third group, engineering products and machinery(28) indicated the presence of significantly different intercepts at the one per cent level. However, the slopes appear to be homogeneous.

Finally, the chemical industry was examined. The χ^2 statistic is again close to zero but the tests for the existence of differential slopes and intercepts proved significant. Indeed, the greater F values,

---

* In brackets B.R. commodity code.

compared to the metal industry example, suggest that heterogeneity within the chemical industry is more marked.

In general, then, there are significant differences in freight generation across industries with differences existing even within the same S.I.C. order. Aggregation, apart from lessening the efficiency of the predictive models, canceals these important differences.

These results may be interpreted in terms of specialisation and differences in the processes of production. In fact, if specialisation is great throughout the country, the same nominal industrial group may generate quite different volumes of freight in different areas, (see Section 2.2 for a theoretical elaboration).

7.3.2 Choice of Variables

At least two important questions related to the choice of variables were raised in the review of previous work. (i) Does the "best" variable vary with the type of traffic examined? (ii) Do O_i and D_j freight necessitate different models? That is, are the determinants of originating and terminating freight different? Only a statement of the main findings is given. The explanation of the results is deferred where a general discussion of both the regional and zonal results is presented.

(i) The Best Variable

With population as the explanatory variable, the poorest results have been obtained for coal and coke products, steel, iron, ores, scrap and alloys. In general final consumption goods have been satisfactorily explained by population. Varying levels of explanation have been derived for commodity groups that comprise both finished products and components destinated for further processing, for example, poor results have been obtained for vehicles and C.K.D. whilst the strongest models have been found for electrical equipment.

Net output appear to best explain commodity groups that have a high ratio of value added to tonnage. Generally, the more the products are processed, the better their movements are explained by net output.

Employment has proved to be the best variable in that it explained satisfactorily the majority of commodity groups.

(ii) O_i, D_j Differences

With population as the explanatory variable D_j's have been systematically better modelled than O_i's. With net output and employment, stronger O_i relationships are observed for commodity groups whose centres of production are relatively localised.

(iii) The Dominant[1] Variables

This problem relates to the technique and to the type of variables used. Although with principal component derivation of significant multiple regression models has been successful, the general tendency confirms previous investigations in that usually one variable accounts for most of the variable (Redding, 1972; G.L.C. and Metra Group Ltd., 1970). The non significance of additional variables has often been related to multi-collinearity. Although multi-collinearity has some bearing on results reported in previous studies, this study has found that:

(1) Even if variables are pairwise uncorrelated and,

(2) Even if when considered, individually, the explanatory variables correlate highly with the dependent variable, the one that has the slightly higher correlation with the dependent variable (r coefficient) will be first to enter the regression and account for most of the variance having very little to be explained by the remaining variables. This point is illustrated in Table 7.10. The extremes examples

---

(1) See for instance Rao and Miller (1971) on this problem in econometric models.

TABLE 7.10: EFFECTS OF DOMINANT VARIABLE ON THE PARTIAL F

EQUATIONS	R SQUARE	RSQ CHANGE	SIMPLE R	F (PARTIAL)
O_i Road Plus Rail Fuel Oil and Petroleum				
Component CE1	0.937	0.937	0.968	57.465
Component CE2	0.964	0.027	0.819	5.368
Component CE3	0.968	0.004	0.131	0.798
D_j Road Plus Rail Fuel Oil and Petroleum				
Component CE1	0.949	0.949	0.974	16.942
Component CE2	0.952	0.003	0.914	0.480
D_j Road Fuel Oil and Petroleum				
Component CE1	0.972	0.972	0.986	43.877
Component CE2	0.974	0.002	0.885	0.417
EXAMPLES WITHOUT DOMINANT VARIABLE				
D_j Road Iron, Ores, Scrap				
Component CE2	0.593	0.593	0.770	85.959
Component CE1	0.941	0.347	0.467	41.405
D_j Road Furniture, Textiles, Shoes				
Component CE1	0.563	0.563	0.750	27.100
Component CE3	0.819	0.256	0.346	10.629

shown are taken from regressions of commodity tonnage on principal components. As seen from the Table, for components, only component CE1 has a significant coefficient and yet the simple correlation of CE1 and CE2 with the dependent variable are respectively 0.96 and 0.81. The column "RSQ Change" shows that the inclusion of CE1 had accounted for 93 per cent of the variance when CE2 was called in the equation. By contrast, regression of D_j road for iron, ores, scrap has yielded a significant coefficient for either of the components. It is worth noting, however, that CE1, which first entered the regression equation has explained only 59 per cent of the variance in the dependent variable and thus leaving a substantial proportion of variance to be explained by CE2.

This sheds light on why variables that are thought of as relevant on theoretical grounds may prove non-significant. This argument, along with that of multi-collinearity may base the researchers' decision to opt for other techniques such as analysis of variance or co-variance. These techniques are used in Part III of this thesis.

7.3.3 Mathematical Form of the Relationships

The linear form has been found to be appropriate. Different plots of the residuals have not revealed any systematic violation of the ordinary least squares assumption. Log and semi-log transformation have not improved the goodness of fit for rail models and for sand, clays, stones and fertilizers by road. This unsuccessful transformation, but most importantly, the relative success of rail-born commodity groups models with the variable employment suggest that it is not a mispecification of the mathematical form that has caused the failure of rail model for most commodity groups and of road models for the commodity groups mentioned above. This is further investigated in Part III where a larger sample and the usage of numerical test for the assumptions help assess a firmer conclusion.

P A R T III

Z O N A L A N A L Y S I S

CHAPTER 8

TOTAL FREIGHT MODELS

8.1 Performance of Aggregate Variables

Three explanatory variables were initially considered viz. resident population, population density and zone area. The results are shown in Table 8.1.

8.1.1 Resident Population

Road plus rail: resident population explains a substantial proportion of the variation in terminating freight 84 per cent. A lower $R^2 = 0.75$ is observed for originating freight as expected. Both relationships are significant at the 0.001 level.

Road: 83 and 87 per cent of the variation in O_i and D_j by road is explained by resident population. Both relationships are significant at the 0.001 level.

Rail: the variance explained by resident population was very low .

Tests of Assumptions

The regression equations for road plus rail and for road can be thought of as satisfactory in view of the coefficients R^2 and F ratios. However, as seen from Table 8.1, the values of the S.N.D. of the I statistic range from 5.25 for O_i by road to 7.73 for D_j

TABLE 8.1: TOTAL FREIGHT TONNAGE BY DIFFERENT MODES REGRESSED ON RESIDENT POPULATION AND POPULATION DENSITY

B.R. Zones
$F(0.001)=11.2$
$r_s(0.01)= 0.30$
$S.N.D.(I)(0.01)=\pm 2.57$

Dependent Variables	Constant b_o	Coefficient b	R^2	F	r_s	S.N.D. (I)
(A) Independent Variable: Resident Population						
Road						
O_i	2824.55	23.86	0.83	658.85	0.44	5.25
D_j	2102.39	25.64	0.87	961.73	0.55	5.37
Road plus Rail						
O_i	3948.09	24.67	0.75	396.52	0.39	7.72
D_j	2921.61	27.21	0.84	710.85	0.41	7.73
(B) Independent Variable: Population Density						
Road						
O_i	7094.54	10121.73	0.62	222.01	0.26	1.09
D_j	6702.67	10857.80	0.65	255.55	0.34	0.63
Road plus Rail						
O_i	8400.37	10398.62	0.55	166.22	0.20	2.14
D_j	7848.56	11435.12	0.62	218.20	0.28	1.04

by road plus rail. These values are significant at the 0.01 level and therefore the hypothesis of non-spatial autocorrelation is rejected. Furthermore, as seen from Table 8.1, the rank correlations between the residuals and the independent variable are positive and significant and thus leading to the rejection of the assumption of homoscedastic errors.

8.1.2 Population Density

Road plus rail: as seen from Table 8.1, population density explained only 55 and 62 per cent of the variation in total originating and terminating freight respectively. The relationships are significant at the 0.001 level.

Road: population density explained 62 per cent of the variation in O_i and 65 per cent of the variation in D_j. The relationships achieved the 0.001 level of significance.

Rail Models

Regressions of total tonnage by rail on population density yielded very poor results. Almost zero correlations resulted.

Tests of Assumptions

The equations derived with population density performed better in view of the O.L.S. assumptions. The values of the S.N.D. of I are not significant at the 0.01

level. At this level, the hypothesis of non-spatial autocorrelation cannot be rejected. It may be rejected at the 0.05 level for O_i by road plus rail. Besides, the Spearman rank correlation coefficient are not significant but one, namely the one for D_j by road. Apart from the latter equation, the models do not therefore violate the assumption of constant variances.

8.1.3 Zone Area

Following the reasoning adopted for the regional models, zonal area was introduced into the equations as a further independent variable with resident population. There was no serious risk of collinearity as the first order correlation between the two variables was 0.21. However, the coefficient obtained did not reveal any abvious impact of zone area on total freight generated. The coefficients for zone area were non significant. Moreover, when checking whether such results were due to the dominance of the variable population, the simple correlation between freight and zone area was found to be negligible.

8.1.4 Discussion

The experiments, with one exception, confirmed the results derived for the regions. Firstly, with this type of explanatory variables, the D_j equations are again better explained than O_i's. Secondly, this type of variable are not appropriate for modelling rail movement. Thirdly, population density proved to be useful at the zonal level while it failed to explain regional movement. This might be explained by the effects of aggregation (see Section 9.4). Finally, in view of the presence of heteroscedasticity and of spatial autocorrelation in the resident population models it was thought that some measure of zonal economic structure[(1)] might improve the models. The next two sections assess the effects of zonal hierarchy and the effect of port zones.

Before doing this, experiments were carried out for resident population models, with double and semi-log transforms. For road and for road plus rail no improvement was noticed. For rail, however, the semi-log equation base on the transformation of the X variable increased the R^2 from 0.08 to 0.14 for D_j for example. The level of explanation remained poor.

8.2 Effect of Zonal Hierarchy: Interpretation of the Bias in the Residuals

8.2.1 Centre Periphery Explanation of Freight

It has been suggested that an explanation in terms of central and peripheral locations may account

---

(1) This experiment was undertaken before the Employment data were collected. The analysis is described in its original form so as to highlight the usefulness of the dummy variables approach.

for the unexplained variation in freight generation. Chisholm and O'Sullivan (1973) have examined residuals from regressions of road freight tonnage generation and attraction on population for a 78-zone system. They suggested that central places may be expected to experience high volumes of freight per resident whereas peripheral areas would specialize in activities involving less transport of goods. For this reason and because the ordinary least-squares procedure is an averaging technique (since it takes all observations into account and gives them equal weight) an excess of actual traffic over predicted traffic (positive residuals) may be expected in central places. The converse situation (negative residuals) is likely to occur in peripheral areas. Chisholm and O'Sullivan found no support for this explanation of freight generation in their analysis. However, the spatial distributions of the residuals from equations shown in Table 8.1 contradict Chisholm and O'Sullivan's conclusion. It was thought relevant to improve the economic interpretation of the residuals, as currently the methodology has the shortcoming that there is no systematic way of distinguishing central and peripheral zones, as opposed to the distinction between rural and urban areas. In addition, covariance analysis may be used to investigate the homogeneity of definitions of zonal hierarchy and to enable the direct derivation of regression relationships for different zone types where it is established that such a course of action is appropriate.

8.2.2 The Experiment

Zonal population was first examined, and those urban areas with population totals above the median were identified as in Chisholm and O'Sullivan's work. An examination of regression residuals for this classification of zones shows (Figures 8.1,8.2) that, whereas rural areas have similar residuals, urban areas have disparate results. In particular, for all rural areas, traffic is slightly overestimated. For urban areas the following patterns are observed:

(1) high underestimation for dominant industrial centres such as Central Scotland, the Lea Valley, and South Yorkshire;

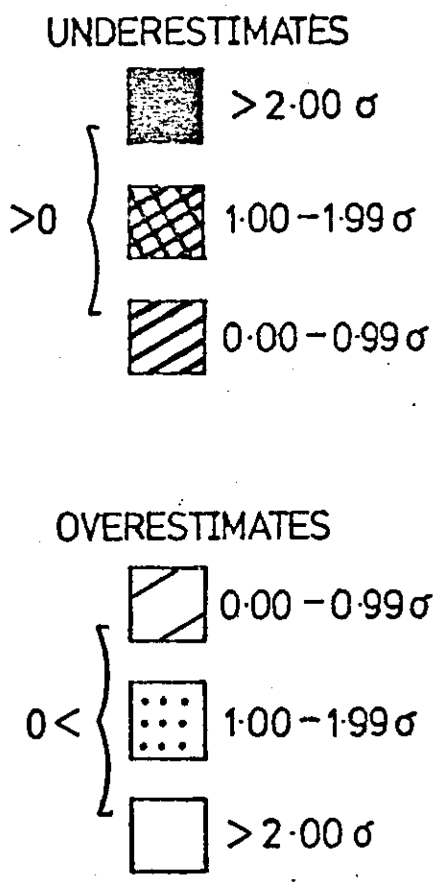
(2) moderate underestimation for declining traditional industrial areas such as Tyneside or South Wales;

(3) high overestimation for urban areas that are predominantly service-oriented, such as Brighton;

(4) moderate underestimation for expanding centres such as Peterborough.

This pattern in the regression residuals suggests that some measure accounting for the basic

LEGEND FOR FIGURES 8.1 & 8.2



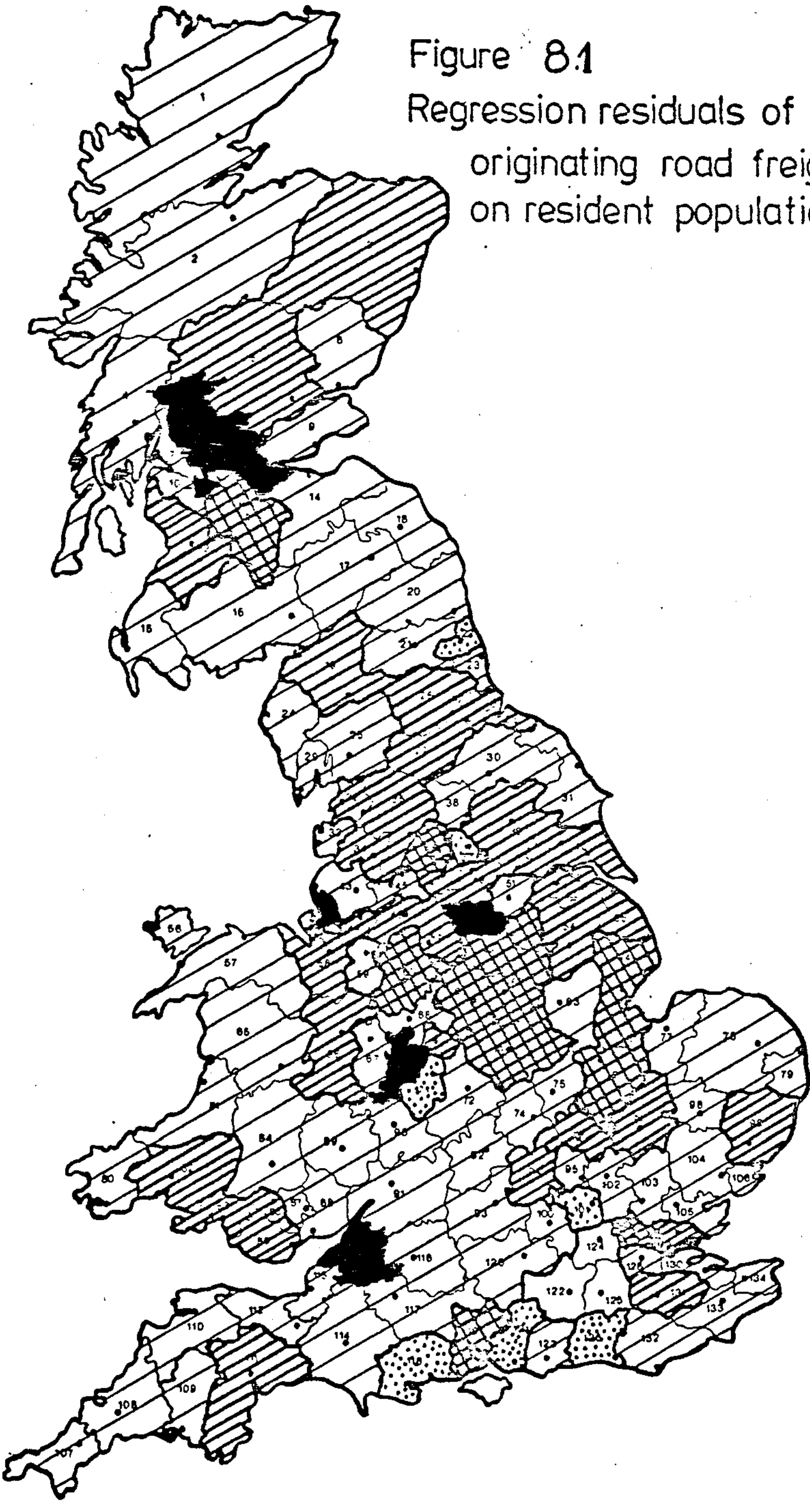


Figure 8.1
Regression residuals of originating road freight on resident population

Figure 8.2
Regression residuals of terminating road freight on resident population

character of different zone types has been omitted. This suggestion is reinforced if some further consideration is given to the way in which regression residuals are computed.

The predicted value of traffic $\hat{Y}_i$ is given by

$$\hat{Y}_i = \bar{Y} + b(X_i - \bar{X}) \qquad (1)$$

where X_i is the independent variable, population, and b the slope coefficient. The residuals are obtained by finding

$$e_i = Y_i - \hat{Y}_i \qquad (2)$$

Equation (1) shows that the variation in the predicted value depends on the size of the zonal population. Yet Equation (2) suggests that the sign and magnitude of the residuals depends also on the actual tonnage Y_i. Chisholm and O'Sullivan (1973) appear to assume that the residual e_i depends solely on population. However, in an analysis in terms of tonnage of freight, if the majority of residents in a zone are engaged in trade and services, the actual tonnage transported is not likely to be in the same proportion as if the residents are not so engaged. Because of this interaction between the size of population and the predominant activity, some urban areas, for instance service-oriented areas, experience high overestimation. This is clearly illustrated in Table 8.2 where zones with similar populations but different orientations are compared. Their different orientations affect both the sign and magnitude of the residuals for these zones.

TABLE 8.2

EFFECT OF ZONAL INDUSTRIAL ORIENTATION ON RESIDUAL MAGNITUDE AND SIGN

Zones	Resident Population (thousands) 1972	Predominant Activity	Residuals in σ	
			O_i	D_j
High Populations				
Wolverhampton (69)	1,338	Heavy Industry	+2	+2
Reigate (125)	1,259	Professional & Scientific Services	-2	-2
Medium Populations				
Swansea (83)	413	Manufacturing Industry	+1	+1
Eastbourne (132)	401	Service Industry	-1	-1

This result suggested that the inclusion of some index of zonal specialization in the regressions would improve the analysis. As employment data at this level of spatial disaggregation were not readily available(2), use was made of the number of commodities

(2) See footnote (1) in this section.

originating in a zone. This approach was successfully used by Loebl and Kenneth (1976) in a study of urban goods vehicle movement. However, it was thought that this crude measure could be improved if it were weighted by the corresponding generation rates for each commodity type.

This gives an index of zonal specialization (Z_i) that may be written formally as

$$Z_i = \sum_{j=1}^{n} \left[c_i^j \left(\sum_{i=1}^{134} o_i^j \Big/ \sum_{i=1}^{134} \sum_{j=1}^{n} o_i^j \right) \right] \quad (3)$$

where c_i^j = number of commodities originating in zone i,

where c_i^j = 1 if the commodity originates in i

= 0 otherwise

o_i^j = originating freight in zone i of commodity j

n = number of commodity groups.

This variable was used to systematically statify zones into four classes:

(1) rural - class D1

(2) urban - class D2

(3) urban manufacturing - class D3

(4) dominant centres - class D4

the index in most cases allowing this distinction to be easily drawn. Further, this stratification broadly corresponds to the variation over zones in the sign and magnitude of the regression residuals from the simple regression of freight tonnage on population. Class D1 corresponds to slight overestimates; D2 to high overestimates and classes D3 and D4 to underestimates. Covariance analysis could now be used to check the appropriateness of the stratification.

These results, set out in Table 8.3, show that the F ratios for tests (I), (II) and (III) are all significant at the one per cent level. That is, the effect of population and the class affiliation is significant as is population with the effects of class affiliation controlled. Further, the differences in intercept between classes is significant. However, test (IV), for the presence of different slope coefficients between the classes, is not significant. This suggests that although the level of explanation can be improved over the simple model based on population, it is not necessary to allow for differences in slope. In fact, whereas, for example, for originating freight by road, R^2 with no stratification is 0.83, R^2 for the equation allowing for differences in intercepts is 0.90 and the predictions for each class may be obtained from the additive model,

$$O_i = 2131.45 + 18.77X_i + 222.68\ D2 + 3143.92\ D3 + 9736.99\ D4$$

$$(17.16) \quad (414.88) \quad (0.06) \quad (16.32) \quad (81.92)$$

$$F(0.01) = 3.44 \qquad (4)$$

TABLE 8.3

COVARIANCE ANALYSIS FOR ORIGINATING FREIGHT AND FOUR CLASSES OF ZONE - RURAL, URBAN SERVICE, URBAN MANUFACTURING, DOMINANT INDUSTRIAL CENTRES

B.R. Zones
F(0.01)
(4,126) = 3.45
(3,126) = 3.94
(1,126) = 6.83

Source of Variation	Mean Square	Degrees of Freedom	F
(I)D and X	3642656137	4	318.5
(II)D, adjusted for X (differential intercepts)	426978009	3	37.3
(III)X, adjusted for D	7525487405	1	658.0
(IV)Interaction of D & X (differential slopes)	0.00	3	0.0
(V)Residuals	11436911	126	-

where the intercepts represents the reference class D1, rural zones. Note that the F ratios in brackets for the coefficients are all significant at the one per cent level, apart from that for D2. This suggests that the urban service classification may not be markedly different in its originating freight characteristics from the reference class. Consequently, the analysis was repeated, pooling the classes D1 and D2 and taking this new class as the reference class.

Table 8.4 gives the results of this exercise. Again

TABLE 8.4

COVARIANCE ANALYSIS FOR ORIGINATING FREIGHT AND THREE CLASSES OF ZONE - RURAL AND URBAN SERVICE, URBAN MANUFACTURING, DOMINANT INDUSTRIAL CENTRES

B.R. Zones
$F(0.01)$
$(3,128) = 3.93$
$(2,128) = 4.16$
$(1,128) = 6.81$

Source of Variation	Mean Square	Degrees of Freedom	F
(I)D and X	4803502600	3	384.0
(II)D, adjusted for X (differential intercepts)	560408637	2	44.8
(III)X, adjusted for D	5283852859	1	422.4
(IV)Interaction of D & X (differential slopes)	8005838	2	0.64
(V)Residuals	12509121	128	-

the zone types are significantly different in terms of their intercepts yet may be considered as having homogeneous slopes. Indeed, the saturated model has insignificant coefficients for the interaction terms. Consequently, the additive model may be used to derive predicting equations for each class of zone. This is,

$$O_i = 2187.23 + 18.85X_i + 3059.95\, D3 + 9609.18\, D4 \qquad (5)$$

$$(23.40) \quad (479.87) \quad (19.16) \quad (102.58)$$

$$F(0.01) = 3.94 \qquad R^2 = 0.90$$

where the F ratios in brackets show each coefficient significantly different from zero at the 1 per cent level.

One further classificatory system was used that owes its origins to central place theory. In a functional sense, the rural and urban service zones may be thought of as peripheral, whilst the remaining urban areas may be thought of as central places from the viewpoint of freight transported. The peripheral zones became the reference class in the analysis and a new class, D5, was designated for the pooled classes D3 and D4.

Table 8.5 shows the covariance table for this stratification of zones and it can be seen that all the F ratios are significant at the one per cent level. Not only should the peripheral and central places have different intercepts, but different slope coefficients are also required for each class. The saturated model, from which predictions for each class may be obtained, has all its coefficients significant at the one per cent level

$$O_i = \underset{(13.81)}{2383.52} + \underset{(101.25)}{18.08X_i} + \underset{(12.40)}{3484.62\ D5} + \underset{(5.20)}{4.48(D5X_i)} \qquad (6)$$

$F(0.01) = 3.94 \qquad R^2 = 0.88$

TABLE 8.5

COVARIANCE ANALYSIS FOR ORIGINATING FREIGHT AND TWO CLASSES OF ZONE - PERIPHERAL (RURAL & URBAN SERVICE) - CENTRAL (URBAN MANUFACTURING & DOMINANT INDUSTRIAL CENTRES)

B.R. Zones
F(0.01)
(2,130) = 4.75
(1,130) = 6.82

Source of Variation	Mean Square	Degrees of Freedom	F
(I)D and X	6965078770	2	471.25
(II)D, adjusted for X (differential intercepts)	640467013	1	43.33
(III)X, adjusted for D	9126654938	1	617.50
(IV)Interaction of D & X (differential slopes)	160116753	1	10.83
(V)Residuals	14780008	130	-

8.2.3 Summary

This analysis has shown, therefore, that however the zones are classified, in terms of the fineness of the classification, the zones differ significantly from one another in terms of the intercept. When grouped as peripheral and central places, the predicting equations appropriate to each class differ in both intercept and slope coefficients.

Such results suggest that the centre-pheriphery explanation of variations in freight generation should not be dismissed as some urban areas' traffic is overestimated. In fact, Chisholm and O'Sullivan (1973), whilst emphasizing the factor of large populations in central areas, overlooked the influence of "central goods" and, as a result, ignored the consequence that areas concentrating on "central goods", such as banking, insurance and other service industries, would give rise to proportionately less freight generation despite their large populations. The notion of centrality is itself relative to the phenomenon investigated, that is, whilst in an analysis of potential contact, based on population, the south-east seaside resorts may be taken as central, they are not necessarily central in terms of freight generated.

In addition, from a predictive rather than an analytical viewpoint, although the simple regression using resident population as explanatory variable yielded a good level of explanation, the distribution of the residuals did not meet the assumption of the O.L.S. technique (see Section 8.1). Introducing dummy variables representing class affiliation, eliminated spatial auto-correlation in the residuals. The values of the standard normal deviate of the I statistic were found to be -1.56, -1.10 and -0.59 for equations 3, 4 and 5 respectively. Besides, the assumption of homoscedastic disturbance terms was not violated. The preferable predicting equation is the last reported (5) since in this case the tests have yielded the lowest values. Finally, covariance tables for terminating traffic have not been derived as the patterns of O_i and D_j are similar (see Figures 8.1 & 8.2). This is to be expected as a regression of O_i on D_j has yielded the following relationship:

$$O_i = 6.97.53 + 0.94\ D_j \quad (6) \quad \text{with} \quad R^2 = 0.97$$

The limited differences between Figures 8.1 and 8.2 may be due to the relative efficiency of population in explaining consumption rather than production. As an example, not that Sheffield's D_j traffic (zone 50) is highly underestimated whilst its O_i movement is only moderately underestimated. This is understandable as Sheffield's industries draw most of their inputs of raw materials from other regions.

8.3 Generation Patterns of Port Zones

As interest lies in the impact of foreign trade on total inland traffic and in the possible comparison of results with previous investigations (Department of the Environment, 1971; Chisholm and O'Sullivan, 1973; Pitfield, 1978), the experiment is carried out for the aggregate mode road plus rail. The former used port variable derived from a conversion of imports and exports in various units (e.g.

square metres of cloth, barrels of beer etc.) into tons, the latter two authors simply described the ratios of imports over O_i and exports over D_j. The experiment discussed below is based on the use of dummy variables derived as follows:

$$P_i = 1 \text{ if zone } i \text{ is a port zone}$$
$$= 0 \text{ otherwise and,}$$
$$MP_i = 1 \text{ if zone } i \text{ is a main port zone,}$$
$$= 0 \text{ zero otherwise.}$$

The cut-off point between port and main port zones as well as the necessity to have different dummy variables for imports and for exports are explained in Appendix 1

.

Port Zones

The covariance analysis results for O_i and D_j are shown in Table 8.6. For originating traffic, the effect of population alone, (Test(III)) is significant as expected. The effect of population and class affiliation together (Test (I)) is also significant. However, Test (II) shows that class affiliation alone, has no significant effect. Moreover the interaction effects (Test (III)) is not significant. This means that the significance of Test (I) is due solely to the variable population.

The results for terminating traffic differ from the O_i results in only one test, that is the test

TABLE 8.6

COVARIANCE ANALYSIS FOR TOTAL FREIGHT BY ROAD PLUS RAIL ON POPULATION AND TWO CLASSES OF ZONE - PORT AND NON-PORT ZONE

B.R. Zones
F(0.01)
(2,130) = 4.75
(1,130) = 6.82

Source of Variation	Mean Square	Degrees of Freedom	F
O_i			
(I)P and X	713035000	2	195
(II)P, adjusted for X (differential intercepts)	11408568	1	0.37
(III)X, adjusted for P	1.3870917E10	1	379.34
(IV)Interaction of X & P (differential slopes)	57042840	(1)	1.56
(V)Residuals	36565923	130	-
D_j			
(I)P and X	8685502600	2	364.49
(II)P, adjusted for X	22626531	1	0.94
(III)X, adjusted for P	1.6760089E10	1	703.34
(IV)Interaction of X & P (differential slopes)	100790910	1	4.22
(V)Residuals	23829060	130	-

for interaction effects (Text (III)) is significant at the five per cent level. The regression for D_j including the interaction terms reduced slightly the bias in the residuals but the assumptions of homoscedastic and independent residuals were still violated.

In view of these results it may be concluded that port zones do not have a pattern of freight generation significantly different from non-port zones. Furthermore, such a zonal stratification adds very little to the predictive power of the models.

<u>Main Port Zones</u>

The results are set out in Table 8.7. For either O_i or D_j the effect of population (test (I)) and the effect of population and classification together (text (III)) are significant. The effect of class affiliation on its own is not significant. This implies that the appropriate predictive equations have homogeneous intercepts but different slope coefficients. The models which accounted for the interaction effect yielded some improvement in the distribution of residuals. However, the decrease in the values of the Spearman test and the S.N.D.(I) was not substantial so as to enable the acceptance of the assumptions of constant variances and spatial independence.

TABLE 8.7

COVARIANCE ANALYSIS FOR TOTAL FREIGHT BY ROAD PLUS RAIL ON POPULATION AND TWO CLASSES OF ZONE - MAIN PORT AND OTHER ZONES

B.R. Zones
F(0.01)
(2,130) = 4.75
(1,130) = 6.82

Source of Variation	Mean Square	Degrees of Freedom	F
O_i			
(I)MP and X	714081290	2	206.60
(II)MP, adjusted for X (differential intercepts)	17112852	1	0.49
(III)X, adjusted for MP	1.3136966E10	1	380.09
(IV)Interaction of X & MP (differential slopes)	230072790	1	6.65
(V)Residuals	34562110	130	-
D_j			
(I)MP and X	8674189400	2	387.70
(II)MP, adjusted for X	0	1	0
(III)X, adjusted for MP	1.6920532E10	1	756.28
(IV)Interaction of X and MP (differential slopes)	312657520	1	13.97
(V)Residuals	22373367	130	-

APPENDIX TO CHAPTER 8

SPATIAL AUTOCORRELATION MEASURES

The spatial autocorrelation test, that is the generalised Moran I Statistic, may be computed under two assumptions (Cliff and Ord, 1973).

(1) Normality: the null hypothesis is that the observed values are the result of taking n values at random from a normally distributed population of values.

(2) Randomisation: the question is, given the particular set of values, "What is the probability that they could have been arranged in the observed way by chance?"

For experimental purposes the tests have been carried out under the two assumptions. The exercise has revealed that, in general, similar results are obtained under either of the assumptions. However, only the results under the second assumption are reported, for it is the safest when dealing with regression residuals (see Ebdon, 1977).

The I Statistic is expressed as a standard normal deviate (S.N.D.(I)) and a two tailed test is used.

Appendix to Chapter 8 (Continued)

The generalised Moran statistic is written as:-

$$I = \frac{n}{W} \left(\sum_i \sum_j w_{ij} \, z_i \, z_j / \sum z_i^2 \right)$$

where w_{ij} is a weight denoting the influence that the i^{th} observation has on the j^{th} and the z_i are deviations from the mean. W is the sum of the weights and n is the number of observations. The distance matrix has been first experimented with. Second, the weights have been expressed as the inverse of the distance. The results obtained with the latter weight matrix were more meaningful in view of the traditional analysis of residuals mapping and D_{ij}^{-1} was therefore considered.

CHAPTER 9

COMMODITY ANALYSIS

As in Part II(1), simple regression equations that relate O_i and D_j of commodity tonnages to aggregate explanatory variables and to employment in the industries that best match commodity groups are first discussed. Then employment measures by S.I.C. orders are subject to principal component analysis in order to extract independent variables that represent the major dimensions of industrial activity for subsequent use in step-wise regressions. In Section 9.3, some attempts at improving rail models are considered and in Section 9.4, the effects of spatial aggregation are evaluated in a comparison of the results from the regional and zonal analyses.

9.1 Simple Regression Models

Firstly, the explanation of commodity tonnages by zonal resident population and population density are examined. Secondly, the advantages of using employment as the explanatory variables are

---

(1) At the zonal level, the aggregate mode road plus rail is not considered as for most commodity groups only a limited number of zones engage in rail traffic. The singularity of rail freight may be appreciated from Section 9.3.

discussed. Thirdly, the possible gains that might be obtained from using employment density are considered.

Resident Population

Road Models: The results set out in Table 9.1A are as expected. That is the best results are observed for terminating traffic and in most cases this explanatory variable best explains commodity groups that are either (1) consumption goods such as potatoes, fruit and vegetable, other food and drinks, or (ii) products that are used in and around agglomerations, like cement, bricks and building materials, earths and stones and terminating shipments of oil and petroleum products. Further, with resident population as the explanatory variable, none of the minerals such as coal, iron, ores or alloys achieved an R^2 in excess of 0.11.

As seen from Table 9.1A, only three relationships have yielded spatially autocorrelated residuals[(1)]. These are O_i other foods and drinks, D_j earths and stones and D_j chemicals with respective standard normal deviates of 2.91, 3.01 and 7.76. By contrast, in all but three equations the assumption of homoscedastic residuals has been violated.

Rail Models: As expected, resident population proved to be a poor explanatory variable for rail

---

(1) See Appendix to Chapter 8: Spatial Autocorrelation Measures.

TABLE 9.1A: COMMODITY TONNAGE BY ROAD REGRESSED ON RESIDENT POPULATION

B.R. Zones
F(0.001)=11.2
$r_s(0.01)=0.30$
S.N.D.(I)(0.01)=±2.57

Dependent Variables	Constant (b_o)	Coefficient (b)	R^2	F	r_s	S.N.D. (I)
O_i						
Potatoes, fruit & veg.	- 48.11	1.06	0.58	188.52	0.67	1.21
Other food & drinks	- 9.02	1.86	0.59	191.25	0.49	-2.91
Bricks & building mat.	631.68	3.24	0.60	199.45	0.27	-1.22
Electrical Equipment	- 63.05	0.41	0.57	179.10	0.51	1.19
Miscellaneous	-335.47	3.45	0.80	557.46	0.51	-1.63
Engineering products	- 84.67	0.90	0.67	276.81	0.61	2.54
Parcels & newspapers	-133.03	1.09	0.69	302.63	0.61	2.44
D_j						
Potatoes, fruit & veg.	- 98.56	1.19	0.66	260.20	0.64	-1.14
Other food & drinks	32.26	1.76	0.64	239.71	0.46	-2.58
Earths & stones	1126.69	3.32	0.52	147.80	0.22	3.02
Chemicals	15.72	0.68	0.52	146.66	0.56	7.76
Cement	45.68	0.50	0.74	377.20	0.24	1.14
Bricks & building mat.	369.88	3.89	0.81	578.89	0.37	-1.55
Oil & petroleum products	83.67	1.24	0.72	356.09	0.51	1.45
Electrical equipment	- 62.40	0.41	0.59	192.55	0.57	1.27
Miscellaneous	-329.99	3.43	0.79	510.10	0.59	-1.69
Engineering products	- 82.10	0.89	0.70	313.55	0.61	2.18
Parcels & newspapers	-196.84	1.24	0.79	500.05	0.48	1.56

commodity shipments. The coefficients of determination, in most cases, have not exceeded 0.03. The best three relationships obtained are shown in Table 9.1B.

As seen from the Table, the level of explanation achieved is still poor. The relationships are all significant at the one per cent level and their respective residuals meet the hypothesis of spatial independence. However, the assumption of constant residual variances is violated. An explanation of the relative success of these three predictive equations and the general failure of rail models is deferred to Section 9.3.

Population Density

Road Models: The results are set out in Table 9.2. With respect to the commodity groups well explained and the relative performance of originating and terminating traffic, the results are the same as those obtained with the variable resident population. However, although when either of resident population and population density are used as explanatory variables the relationships achieve the 0.001 level of significance, the variation explained by population density is lower. Besides, the results for the tests on the OLS assumptions are different. On the one hand, a high degree of spatial autocorrelation in the

TABLE 9.1B: COMMODITY TONNAGE REGRESSED ON RESIDENT POPULATION

B.R. Zones
F(0.001) = 11.2
r_s(0.01) = 0.30
S.N.D.(I)= ±2.57

Dependent Variable	Constant (b_o)	Coefficient (b)	R^2	F	r_s	S.N.D.(I)
Rail						
O_i Other parcels and newspapers	-6.562	0.034	0.51	142.26	0.53	-2.10
D_j Other foods and drinks	-1.376	0.014	0.44	95.21	0.63	0.72
D_j Milk	-7.705	0.024	0.48	122.83	0.27	1.82

TABLE 9.2: COMMODITY TONNAGE BY ROAD REGRESSED ON POPULATION DENSITY

B.R. Zones
F(0.001)=11.2
r_s(0.01)=0.30
S.N.D.(I)=±2.57

Dependent Variables	Constant (b_o)	Coefficient (b)	R^2	F	r_s	S.N.D. (I)
O_i						
Potatoes, fruit & veg.	128.49	479.38	0.49	130.36	0.43	1.45
Other food & drinks	259.99	913.35	0.59	192.39	0.47	-4.97
Chemicals	108.18	350.71	0.48	126.37	0.35	4.87
Bricks	1223.50	1354.61	0.43	103.31	0.04	0.06
Electrical equipment	-0.58	201.58	0.55	165.50	0.70	4.46
Miscellaneous	271.33	1484.73	0.62	220.76	0.47	0.99
Engineering products	55.86	422.66	0.62	215.79	0.59	4.96
Parcels & newspapers	46.16	493.55	0.59	194.13	0.60	4.91
D_j						
Potatoes, fruit & veg.	108.23	517.42	0.52	145.01	0.37	-0.61
Other food & drinks	304.76	829.27	0.59	194.99	0.37	-4.49
Chemicals	103.87	358.80	0.59	194.67	0.46	3.38
Cement	142.30	201.11	0.49	129.38	0.32	7.01
Bricks	1084.20	1616.27	0.58	188.37	0.36	-0.28
Timber	159.73	265.88	0.46	166.29	0.24	1.59
Oil & petroleum prod.	308.94	526.39	0.54	156.37	0.24	2.78
Electrical equipment	0.12	200.24	0.57	174.98	0.70	4.98
Miscellaneous	277.36	1473.40	0.61	206.72	0.48	0.35
Engineering products	51.36	431.12	0.68	281.14	0.52	4.49
Parcels & newspapers	19.74	543.17	0.62	220.48	0.69	4.36

residuals is observed for most of the relationships as the modal value of the standard normal deviate of the I Statistic is 4. On the other hand, the degree of heteroscedasticity in the residuals is very much lower as seen from the values of the Spearman rank correlation coefficients.

Rail Models: The rail models proved, again, to be very poor as the coefficients of determination for the best relationships have not exceed 0.20.

Summary:

The aggregate variables resident population and population density have reasonably explained O_i and D_j of about half of the commodity groups considered. The commodity groups explained are those known to correlate with the distribution of urban centres. However, only few of the models are robust enough to warrant confident forecasts as some bias is observed in the residuals and the level of explanation remains moderate.

The main motivation of using these two variables is their usual availability and projectability. Employment measures disaggregated by industry are more difficult to collect or forecast but are more meaningful and may be more suitable for an explanation of commodity movements.

Employment by Industry

In view of the difficulties in matching some of the commodity groups to corresponding industrial orders or minimum lists headings (M.L.H.), it was decided to retain only 17 commodity categories for which a reasonable match was achieved in the sense that each of the dependent variable is regressed on employment in the compatible industry (see Appendix II.3).

Road Models: The results are shown in Table 9.3A. As only 17 commodity groups have been considered for this exercise, the rate of success achieved is higher than that attained by either resident population or population density. A reasonable level of explanation is achieved for respectively, 11 and 10, O_i and D_j models. In terms of variation explained however, the variable employment has performed less well than resident population but as well as population density. The coefficients of determination range from 0.48 for O_i fertilizers to 0.78 for O_i steel. The major difference is that employment explains reasonably well commodity groups like coal, steel, alloys, pig and iron castings, vehicles and fertilizers for which resident population and population density have yielded very poor results (R^2's < 0.11). Furthermore, unlike the two aggregate variables,

TABLE 9.3A: COMMODITY TONNAGE BY ROAD REGRESSED ON EMPLOYMENT

B.R. Zones
F(0.001)=11.2
r_s(0.01)=0.30
S.N.D.(I)=±2.57

Dependent Variables	Constant (b_o)	Coefficient (b)	R^2	F	r_s	S.N.D. (I)
O_i						
Coal & coke products	475.99	1.29	0.59	196.92	-0.01	0.84
Steel	51.26	1.11	0.78	489.27	0.28	1.84
Alloys	10.11	0.31	0.62	222.59	0.46	1.74
Pig & iron castings	74.35	1.31	0.64	241.34	0.32	-0.57
Other food & drinks	226.02	1.92	0.61	210.98	0.48	-3.69
Chemicals	98.59	0.68	0.63	225.60	0.50	4.62
Fertilizers	97.67	4.69	0.48	95.99	0.10	1.94
Vehicles	23.50	0.15	0.62	218.01	0.67	2.60
Textiles & furniture	84.93	0.43	0.69	296.75	0.59	1.48
Electrical equipment	8.31	0.15	0.49	129.98	0.63	1.05
Engineering products	18.81	0.56	0.61	212.35	0.39	-0.09
D_j						
Coal & coke products	583.09	0.89	0.49	93.24	-0.02	2.55
Steel	111.63	0.86	0.59	190.23	0.21	4.02
Alloys	21.23	0.30	0.67	273.96	0.38	1.95
Pig & iron castings	92.61	1.05	0.56	172.09	0.28	0.01
Other food & drinks	180.76	1.71	0.60	198.57	0.43	-3.05
Chemicals	114.84	0.62	0.61	213.37	0.41	-0.07
Vehicles	26.14	0.15	0.57	180.56	0.54	2.25
Textiles & furniture	86.68	0.42	0.69	300.71	0.58	1.14
Electrical Equipment	9.99	0.15	0.49	130.84	0.65	0.97
Engineering products	26.00	0.54	0.61	210.74	0.32	1.00

employment best explains originating freight which is reasonable as employment is a better surrogate for production than it is for consumption.

The employment models are also more satisfactory in view of the OLS assumptions. Although the hypothesis of spatially independent residuals is not met by four of the models, the values of the standard normal deviate are in general low. Besides, the degree of heteroscedasticity is relatively moderate.

Rail Models: The results derived for commodity shipments by rail are again very poor with the exception of four regressions which have attained a modest level of explanation as seen from Table 9.3B.

As seen from Table 9.3B, the results for O_i coal and coke products attain a satisfactory level of explanation but for the remaining results R^2's range from 0.49 to 0.56. Further, with the exception of O_i fertilizers, the models have not met the assumption of homoscedastic error terms. For all models however, the null hypothesis of no-spatial autocorrelation in the residual variances is not rejected.

Employment Density:

One of the problems in using âreal data is the effect of zone size bias which may be corrected for, in some cases, by dividing the variable by zone

TABLE 9.3B: COMMODITY TONNAGE BY RAIL REGRESSED ON EMPLOYMENT

B.R. Zones
F(0.001) = 11.2
r_s(0.01) = 0.30
S.N.D.(I)= ±2.57

Dependent Variable	Constant (b_o)	Coefficient (b)	R^2	F	r_s	S.N.D.(I)
O_i Coal & coke products	16.080	2.971	0.77	461.52	0.77	-2.21
O_i Fertilizers	-3.545	1.007	0.56	168.00	0.08	0.75
D_j Steel	33.423	0.192	0.49	92.21	0.46	0.23
D_j Other foods & drinks	-0.082	0.017	0.55	167.81	0.67	1.58
D_j Chemicals	-0.303	0.018	0.51	96.28	0.69	1.52

area, (King, 1972). It was hypothesized that the effect of zone size underlies the persistent occurrence of heteroscedasticity. Indeed, the use of employment density did help correct for heteroscedasticity in the residuals for some road models as seen from Table 9.4.

The values of the Spearman rank correlation coefficients have decreased so as to become non-significant for O_i chemicals, O_i textiles and furniture and D_j chemicals and have slightly been reduced for O_i and D_j electrical equipment. In all cases the values of the S.N.D.(I) have increased but not enough to exceed the critical limits and thus the hypothesis of no-spatial autocorrelation may be retained. Nevertheless, it is worth noting that such results are associated with lower R^2's relative to those for employment.

<u>Summary</u>:

The foregoing results have proved to be consistent with those derived for the Standard regions. The population variable best explains D_j traffic and in general provides a good fit for O_i and D_j movements of consumption goods and building materials. By contrast, the employment variable performs best with O_i models and the commodity groups it best explains can be either consumption or production goods. The variable

TABLE 9.4: COMMODITY TONNAGE BY ROAD REGRESSED ON EMPLOYMENT DENSITY

B.R. Zones
$F(0.001) = 11.2$
$r_s(0.01) = 0.30$
$S.N.D.(I) = \pm 2.57$

Dependent Variable	Constant (b_o)	Coefficient (b)	R^2	F	r_s	S.N.D.(I)
O_i Chemicals	168.874	260.202	0.60	191.93	0.21	1.75
O_i Textiles, furnitures & shoes	202.611	192.362	0.59	130.23	0.28	-1.64
O_i Electrical equipment	24.399	96.882	0.54	156.43	0.61	2.44
O_i Engineering products	82.384	325.019	0.62	220.81	0.48	3.08
D_j Chemicals	186.063	224.711	0.51	139.83	0.16	-2.17
D_j Electrical equipment	25.667	95.505	0.54	159.59	0.51	2.13
D_j Engineering products	82.472	324.874	0.65	255.34	0.50	3.94

population density has proved to be useful for the zonal analysis, whereas it yielded poor results for the regional analysis. This latter result may be explained by aggregation effects (see Section 9.4).

With respect to OLS assumptions, the employment variable performed better than either population density or resident population. The use of population density or employment density helps reduce heteroscedasticity in the residuals but in general has an adverse effect on the spatial independence of the residuals.

9.2 Principal Component Analysis Based Models

Principal Component Analysis

The analysis has retained only S.I.C. orders 1 to 20 as the remaining orders represent service industries. The results are set out in Table 9.5. The six significant components obtained, account for about 80 per cent of the total variance in the data.

Component 1: This component accounted for about half of the variance. It is related to employment in the following S.I.C. orders, paper, printing and publishing (X18), other manufacturing industries (X19) and construction industry (X20). It is also significantly related to the engineering industries

TABLE 9.5: PRINCIPAL COMPONENT ANALYSIS OF EMPLOYMENT IN THE S.I.C. ORDERS

B.R. Zones

Variables (S.I.C. Orders) 1 to 20	Component 1 (Loadings)	Component 2 (Loadings)	Component 3 (Loadings)	Component 4 (Loadings)	Component 5 (Loadings)	Component 6 (Loadings)
X1	-0.092	-0.161	-0.155	-0.031	-0.026	-0.137
X2	-0.037	0.011	0.061	0.793	-0.016	-0.002
X3	0.804	0.116	0.069	0.161	0.227	0.178
X4	0.540	-0.002	-0.376	0.283	-0.023	0.018
X5	0.782	0.150	0.029	0.068	0.026	0.097
X6	0.093	0.945	-0.007	0.106	0.051	-0.031
X7	0.728	0.371	0.307	0.118	0.309	0.048
X8	0.871	0.002	-0.073	-0.080	0.133	-0.139
X9	0.850	0.139	0.027	0.103	0.362	0.066
X10	0.110	-0.056	0.077	0.027	0.037	0.955
X11	0.290	0.284	0.098	-0.019	0.822	0.045
X12	0.348	0.818	0.034	0.028	0.369	-0.043
X13	0.766	0.354	0.336	-0.067	-0.078	0.026
X14	0.766	0.354	0.336	-0.067	-0.078	0.026
X15	0.779	0.121	0.329	0.069	-0.178	0.154
X16	0.213	0.159	0.158	0.724	0.025	0.052
X17	0.884	0.227	0.206	0.015	0.012	0.094
X18	0.936	-0.036	-0.031	0.084	0.193	-0.083
X19	0.886	0.142	0.078	0.075	0.213	-0.023
X20	0.836	0.042	-0.086	0.011	0.015	0.087
Eigenvalues	9.3	1.8	1.4	1.2	0.9	0.9
PCT of variance	46.9	9.3	7.1	6.4	4.9	4.7
cum PCT	46.9	56.2	63.3	69.7	74.6	79.3

namely instrument engineering (X8), mechanical engineering (X7) and electrical engineering (X9). Previous researchers in industrial location have identified these industries to be market oriented (Smith, 1971). The variables X18, employment in the paper industry will be used as a measure of market demands in subsequent regression analysis as it loads most highly on Component 1.

Component 2: This component has a clear interpretation since its sole significant association is with the metal industries that is 0.94 and 0.81 respectively with metal manufacture (X6) and metal goods not elsewhere specified (X12). The employment in S.I.C. order 6 is taken as representative of this component.

Component 3: The only significant loading observed for this component (0.85) is with variable X13, employment in the textile industries and so that variable is used in the step-wise regression.

Component 4: This component has two significant loadings of 0.79 and 0.72 with employment in mining quarrying (X2) and employment in bricks, pottery and cement (X16). These associations suggest that Component 4 is a measure of mining activity. The variable X2 is used as it loads highly on the component.

Component 5: A single significant association is observed for Component 5, (0.82) with employment in the vehicles industry (X11).

Component 6: The last significant component is related to employment in shipbuilding and marine engineering (X10).

Summary

The principal component analysis on employment by S.I.C. order has reduced the set of data to six major dimensions, a market measure and five representations of the main industries in the U.K. The results are consistent with these derived for employment at the regional level though a clearer interpretation is obtained at the zonal level since in most cases a single variable loads significantly on the component. This allows the use of a variable to represent a component, in contrast to the approach used in the regional analysis of Part II.

Principal Components Based Models

Originating and terminating freight by road and by rail for the 24 B.R. commodity groups[1] were regressed on the six variables selected, namely employment in mining and quarrying (X2), in metal manufacture (X6), in shipbuilding and marine engineering

(1) There are no flows by road for commodity group 5. The agricultural products (B.R. codes 9 to 12) were not considered.

(X10), in the vehicle industry (X11), in textiles (X13) and in paper, printing and publishing (X18).

Road Models: The results for O_i and D_j by road are shown in Tables 9.6 and 9.7 respectively.

O_i by Road: As seen from Table 9.6, only 11 out of the 29 regressions achieved a good level of explanation. The coefficients of determination range from 0.58 for alloys to 0.89 for steel, with a median value of 0.71. Both the overall relationships and the coefficients satisfy the 0.01 level of significance with the exception of the constant terms for alloys, pig iron and scrap.

The relationships for coal and coke, alloys and scrap, yielded spatially autocorrelated residuals as their respective standard normal deviates of -3.18, -3.45 and -2.70 exceed the critical limit of -2.57. The remaining relationships meet the assumption of independent error terms. However, the test for the detection of heteroscedasticity is relatively less conclusive as the value of the Spearman correlation coefficient is computed with respect to everyone of the explanatory variables. Nevertheless, it may be said that the equations for scrap, oil and petroleum and coal and coke are virtually free from heteroscedasticity. The regression equations for steel, vehicles and miscellaneous however, show a relatively high

TABLE 9.6: ORIGINATING TONNAGE OF COMMODITIES BY ROAD REGRESSED ON SELECTED EMPLOYMENT VARIABLES

B.R. Zones
F(0.001) = 11.2
$r_s(0.01)$ = 0.30
S.N.D.(I) = ± 2.57

COMMODITY GROUPS	EQUATIONS	F	R^2	r_s (k) • •	S.N.D. (I)
Coal & Coke Products	$y = 181.98 + 1.22 X_2 + 0.27 X_{13} + 0.24 X_6 + 0.22 X_{18}$	89.62	0.73	0.34, 0.40, 0.30, 0.24	3.18
Steel	$y = 11.65 + 0.89 X_6 + 0.06 X_{11}$	567.15	0.89	0.61, 0.54	- 1.84
Alloys	$y = 5.82^{\bullet} + 0.05 X_6 + 0.03 X_{11} + 0.03 X_{18} + 0.04 X_{10}$	39.10	0.58	0.50, 0.47, 0.50, 0.13	- 3.45
Pig & Iron Castings	$y = 6.12^{\bullet} + 0.32 X_6 + 0.08 X_{13} + 0.03 X_{10}$	171.21	0.79	0.34, 0.49, 0.06	0.34
Scrap	$y = 4.89^{\bullet} + 0.15 X_6 + 0.09 X_{18} + 0.02 X_{11} + 0.02 X_{13}$	92.62	0.74	0.30, 0.49, 0.38, 0.46	- 2.70
Oil & Petroleum Prod.	$y = 250.89 + 0.56 X_{18} + 0.26 X_{10} + 0.16 X_{11} + 0.11 X_6$	48.99	0.60	0.21, 0.30, 0.19, - 0.04	1.82
Vehicles	$y = 27.54 + 0.15 X_{11}$	213.39	0.61	0.63	2.90
Textiles	$y = 101.98 + 0.50 X_{13} + 0.45 X_{18} + 0.14 X_6$	110.41	0.71	0.46, 0.50, 0.13	- 1.76
Miscellaneous	$y = 150.67 + 1.83 X_{18} + 0.23 X_{11} + 0.33 X_{13} + 0.24 X_6$	73.42	0.69	0.54, 0.42, 0.44, 0.52	- 1.54
Engineering Products	$y = 11.74 + 0.20 X_6 + 0.32 X_{18} + 0.10 X_{11} + 0.21 X_{10} + 0.06 X_{13}$	83.97	0.77	0.50, 0.51, 0.48, 0.40, 0.20	- 2.41
Parcels & Newspapers	$y = 51.54 + 0.78 X_{18} + 0.05 X_{11}$	195.55	0.74	0.52, 0.47	- 0.31

All the coefficients are significant at the 0.01 level of significance.

The constant terms indicated by • are not significant.

• • r_s for the respective k^{th} explanatory variable.

X2 = mining and quarrying; X6 = metal manufacture; X10 = shipbuilding and marine engineering;
X11 = vehicles; X13 = textiles; X18 = paper and printing.

degree of heteroscedasticity in the residuals. The remaining equations have a moderate degree of heteroscedasticity.

D_j by Road: The results for terminating freight are shown in Table 9.7. Half of the commodity groups considered have moderate to satisfactory coefficients of determination; for example for alloys and for parcels and newspapers the coefficients are 0.58 and 0.88. The median R^2 is 0.68. The relationships are significant at the 0.001 level of significance but the constant terms in the alloys, iron and engineering products models are not significant.

In only two cases is the assumption of independent residuals violated. This is the case for D_j alloys and D_j engineering products where the respective standard normal deviates are -3.11 and -2.88. The equations for coal and coke products, iron, earths and stones, cement, bricks and oil and petroleum meet the assumption constant residual variances. In the remaining equations a moderate degree of heteroscedasticity is observed with the exception of D_j parcels and newspapers which shows a high degree of heteroscedasticity in its residuals.

TABLE 9.7: TERMINATING TONNAGE OF COMMODITIES BY ROAD REGRESSED ON SELECTED EMPLOYMENT VARIABLES

B.R. Zones
$F(0.001) = 11.2$
$r_s(0.01) = 0.30$
$S.N.D.(I) = \pm 2.57$

COMMODITY GROUPS	EQUATIONS	R^2	F	r_s (k) ••	S.N.D. (I)
Coal & Coke Products	$y = 263.83 + 0.77 X_2 + 0.32 X_{13} + 0.31 X_6 + 0.22 X_{18}$	0.68	69.78	0.30, 0.33, 0.24, 0.18	- 2.05
Steel	$y = 22.51 + 0.73 X_6 + 0.06 X_{11}$	0.85	390.57	0.58, 0.37	- 1.18
Alloys	$y = 5.02^{\bullet} + 0.03 X_{11} + 0.04 X_6 + 0.03 X_{18} + 0.04 X_{10}$	0.58	44.95	0.54, 0.53, 0.58, 0.10	- 3.11
Pig Iron	$y = 0.05^{\bullet} + 0.23 X_6 + 0.06 X_{11} + 0.03 X_{13} + 0.09 X_{10}$	0.78	115.28	0.40, 0.35, 0.33, 0.08	0.24
Scrap	$y = 2.81 + 0.18 X_6 + 0.04 X_{18} + 0.02 X_{13} + 0.02 X_{11}$	0.79	100.42	0.45, 0.50, 0.40, 0.20	- 1.08
Earths & Stones	$y = 1335.83 + 1.49 X_{18} + 0.48 X_6 + 0.41 X_{13} + 0.61 X_{13} + 0.25 X_2$	0.54	30.59	0.14, 0.18, 0.18, 0.09, 0.08	2.17
Cement	$y = 123.11 + 0.28 X_{18} + 0.04 X_{11} + 0.04 X_6$	0.65	83.96	0.25, 0.26, 0.26	- 1.13
Bricks	$y = 855.31 + 1.98 X_{18} + 0.38 X_6 + 0.34 X_{11} + 0.30 X_2 + 0.54 X_{10}$	0.67	52.03	0.29, 0.15, 0.10, 0.13, 0.10	- 1.72
Oil & Petroleum Prod.	$y = 250.89 + 0.56 X_{18} + 0.16 X_{11} + 0.11 X_6 + 0.23 X_{10}$	0.60	48.99	0.32, 0.24, 0.14, 0.09	1.93
Vehicles	$y = 29.94 + 0.15 X_{11}$	0.57	178.33	0.50	2.33
Textiles	$y = 59.55 + 0.47 X_{13} + 0.45 X_{18} + 0.08 X_{11} + 0.11 X_2$	0.74	92.85	0.52, 0.53, 0.35, 0.13	- 1.33
Electrical Equipment	$y = 15.17 + 0.27 X_{18} + 0.03 X_6$	0.54	68.82	0.54, 0.35	- 0.92
Miscellaneous	$y = 152.45 + 1.84 X_{18} + 0.24 X_{11} + 0.30 X_{13} + 0.27 X_6$	0.68	70.99	0.54, 0.44, 0.47, 0.55	- 1.98
Engineering Products	$y = 5.22^{\bullet} + 0.33 X_{18} + 0.17 X_6 + 0.10 X_{11} + 0.24 X_{10} + 0.05 X_{13}$	0.67	52.96	0.55, 0.50, 0.49, 0.15, 0.40	- 2.88
Parcels & Newspapers	$y = 1.71 + 0.90 X_{18} + 0.06 X_{11} + 0.02 X_6$	0.88	392.05	0.73, 0.62, 0.47	- 2.09

All the coefficients are significant at the 0.01 level of significance.
The constant terms indicated by • are not significant.
•• r_s for the respective k^{th} explanatory variable.

X2 = mining and quarrying; X6 = metal manufacture; X10 = shipbuilding and marine engineering;
X11 = vehicles; X13 = textiles; X18 = paper and printing.

Performance in Terms of Industrial and Spatial Associations

The models are plausible from the point of view of the variables included in the equations as well as the order in which they entered. The order of inclusion is of interest as in step-wise regression analysis[1], it reflects the relative contribution to the explained variation. The results for O_i and D_j may be summarized as follows:

Coal and Coke Products: Similar relationships are observed for O_i and D_j. The variables included are, in order of inclusion, mining and quarrying (X2), textiles (X13), metal manufacture (X6) and the market measure (X18). The first variable (X2) explains 59 and 40 per cent of the variation in O_i and D_j respectively and the variable (X13) accounts for respectively 8 and 15 per cent of the explained variation in O_i and D_j. This is reasonable as (i) the first variable represents the industry producing the good; (ii) the second variable has obvious spatial association since the major mining areas are also textile producing regions for instance the North and the East-Midlands.

Metal Industries: Under this heading are the commodity groups steel, alloys, iron and scrap. For these groups most of the explained variation is due to the inclusion of variables (X6) and (X11), that is

(1) The forward stepwise method was used.

employment in metal manufacture and in vehicles. The variable (X6) was the first to be included in the O_i regressions as well as the D_j equations for iron and scrap and it contributed to over 60 per cent of the variance explained in all equations except those for alloys. These results are meaningful as metal manufacture relies largely on intra-industry linkage and the vehicles' industry is its major market. The inclusion of employment in textiles (X13) in the iron and scrap regressions is explained by spatial associations.

Oil and Petroleum Products: Although the same explanatory variables were picked up by the step-wise regressions for O_i and D_j, the order in which they entered the equations is different and so highlights a relatively different pattern for originating and terminating flows. As expected, the market measure (X18) and employment in shipbuilding (X10) first entered the O_i equation and together explained 50 per cent of its variation. Note that X10 represents a spatial association as most oil refineries are located on coastal sites whilst X18 represents large urban centres with major oil depots. The variable (X18) accounted for 43 per cent of the variation in terminating flows and the second major contribution was due to vehicles' industry employment (X11).

<u>Building Materials</u>: Under this heading are the commodity groups, earths and stones, cement and bricks. Good models were only obtained for terminating freight for these commodity groups. Variable (X18) explained 31, 53 and 56 per cent of the variation in respectively earths and stones, bricks and cement. For earths and stones and bricks, the next most important contribution to explained variation was due to employment in metal manufacture (X6). In the case of cement, the variable included after cement was vehicles (X11).

<u>Engineering Products</u>: The same variable entered the O_i and D_j equations for this commodity group. These are metal manufacture (X6), the market measure (X18), vehicles (X11), shipbuilding and marine engineering (X10) and textiles (X13). The first two variables accounted for over 50 per cent of the explained variance in O_i and D_j. The variable X6 first entered the O_i model whereas X18 first entered the D_j explanation.

<u>Vehicles</u>: A single variable, namely employment in the vehicle industry (X11), contributed significantly to the explanation of O_i and D_j. This is probably due to the great concentration of the production and the market around London and the Midlands.

Electrical Equipment: A meaningful relationship was derived for D_j electrical equipment as the variables included are the market measure (X18) and employment in metal manufacture (X6). It is worth noting that substantial industrial linkages relate electrical industries to metal industries (Lever, 1975). Further, X18 represents component I with which employment in instrument and mechanical engineering correlate.

Textiles: The variation in originating traffic was mainly due to employment in textiles (X13) and the market measure (X18). Similar results were obtained for the D_j equation which also included employment in the vehicles' industry (X11) and in mining (X2).

Parcels, Newspapers and Miscellaneous: The variation in O_i and D_j for these commodity groups were primarily explained by X18 and X11.

Summary

The results derived from step-wise regressions on the selected independent employment variables by S.I.C. order may be summarized as follows.

(1) Over half of the 48 regressions achieved a level of explanation in excess of 50 per

cent. Most commodity groups, for which the results are not shown, yielded R^2's in the range 0.35, 0.48; amongst these are chemicals, plastics and other food and drinks. However, by far, the poorest performance was observed for ores, industrial sands and fertilizers for which the coefficients of determination did not exceed 0.12.

(2) The models were plausible in that some causation could be inferred. The major determinant of O_i traffic proved to be the employment in the corresponding industry and for D_j the market dimension. This was assessed on the basis of the relative contribution of the variables as well as the "BETA" coefficients[1]. The regression coefficients are of the correct sign and so compare favourably with those shown in Chisholm and O'Sullivan (1973), where negative sign are numerous.

(3) Compared with the simple regression of commodity tonnages on employment in the respective industry, the improvement in

---

(1) The "BETA" coefficients are standardized regressions coefficients and therefore provide a relative measure of the importance of the explanatory variables in their effect on the dependent variable.

the level of explanation is, in general, very limited as judged from $\bar{R}^2$, that is when the number of explanatory variables is taken into account (Koutsoyannis, 1972). As regards the assumption of independent residuals, the step-wise regressions models are more robust since the negative but non-significant values of the S.N.D.(I) indicate a dispersed pattern whereas the positive values found for the simple models attest that the residuals tend to cluster. For either type of model, however, the assumption of independent residuals is in general met.

(4) The compatibility in the level of aggregation of the dependent and independent variables proved to be important, in particular for some commodity groups like chemicals and fertilizers as multiple regressions based on S.I.C. orders performed less well than univariate models based on M.L.H. groups. It is pertinent to note, that results from covariance analysis in Section 7.3 , show that heterogeneity is great within these industries.

Rail Models: The regressions of commodity tonnage on the selected S.I.C. orders in general yielded poor results. The best models were derived for originating movement of coal and coke, scrap and parcels and newspapers. These attained a level of explanation of respectively 79, 56 and 73 per cent. The next best models were obtained for terminating tonnages of coal and coke, steel, alloys, iron, food and drinks, chemicals and oil and petroleum products. These models yielded R^2's in the range 0.35-0.49 with, in most cases, only one significant explanatory variable.

9.3 Other Attempts at Improving the Models

9.3.1 Rail Models

The generally poor results observed for rail models were expected in view of the insignificant role this mode plays in the transport of the majority of commodity groups considered (see Appendix IV). However, satisfactory models were derived for those commodity groups which constitute most of the rail market. In order to improve rail models for other commodities, or at least to gain a better understanding of rail freight patterns, two lines of inquiry were adopted. These were the transformation of the original variables and weighting the explanatory variables so as to reflect the rail market share.

Transformation of the Original Variables

The rail data are not only biased towards some specific industries but also towards some limited number of areas and so the spatial distributions of the dependent variables tend to deviate from normality. In fact the distributions are positively skewed and leptokurtic. Although the skeweness moderate (the modal value is 5), the kurtosis is high (the modal value is 38). The usual transformation for this bias that is the logarithmic transformation, was undertaken but without success.

Weighted Explanatory Variables

It was hypothesised that if an industry or an area uses rail to transport only a small proportion of its output, then, only a correspondingly small proportion of its employment ought to be used as an explanatory variable. In order to approximate the fraction of employment that would adequately represent the rail share, some index to weight the original variables was needed. To this end, recourse was made to canonical correlation analysis[1]. This method is especially designed to identify correlations between

---

(1) This technique is based on the same principles as those of principal components; an example of its use in freight modelling may be found in Watson (1975).

sets of variables and so if it is applied to the sets of freight and employment data it can provide an objective way to compute an index. As different sub-sets of commodities and employment by industry are likely to load significantly on different canonical vectors, correlations between zonal economic specialisation and mature of rail freight is expected to emerge. Consequently, an index which represents the relative proportion of employment in industries that load significantly on the vectors to the total employment in an area would be appropriate. Furthermore, this technique was selected for, even if no improvement is achieved for the regression models, the canonical correlations may nevertheless aid the understanding of general relationships between the two sets of data.

Regressions on Weighted Employment Variables:

The index used for a given zone i and for each vector takes the form

$$g_i = \frac{\sum_k E_i^k}{\sum_n E_i^n} ,$$

where $\sum_k E_i^k$ is the sum of employment over the k industries that load significantly on the canonical vector and $\sum_n E_i^n$ is the sum over all industries n present at location i. The weighted employment variable was equal to $g_i \times E_i^h$, where, h represents the industry under consideration.

In relative terms, the new variables performed better than the original variables yet in absolute terms the coefficients of determinations were too poor to be considered. However, the canonical correlations between rail freight by commodity and employment by S.I.C. order are of some interest.

Relationships Between Sets of Variables:

First a canonical correlation on the 30 B.R. commodity groups and employment by S.I.C. group was carried out. Seven significant vectors were derived. However, the only variables that yielded significant loadings for the freight set were commodity groups such as coal and coke, ores, steel and alloys. As this exercise was not geared towards these commodities for adequate models had already been obtained, such results were not useful. Consequently, a second canonical correlation analysis, which did not include B.R. commodity groups 1 to 8 (see Appendix II.2) was performed. Five significant canonical vectors were extracted. These are shown in Table 9.8.

The first vector correlates with plastics and engineering products from the freight set and with metal manufacture and other manufacture from the

TABLE 9.8: CANONICAL CORRELATION ANALYSIS: O_i FOR COMMODITY GROUPS 9 TO 30 AGAINST EMPLOYMENT BY S.I.C. ORDERS

B.R. Zones

		Loadings* on Canonical Vectors				
		I	II	III	IV	V
First Set:	Freight					
	Foods and drinks		0.52			
	Chemicals		0.30			
	Plastics	0.36		-0.83		
	Fertilizers					-0.32
	Timber & forest products					0.30
	Oil & petroleum products				0.52	-0.51
	Textiles & furniture					0.39
	Electrical equipment				0.69	0.35
	Miscellaneous				-0.75	-0.45
	Engineering products	0.79		0.72		
	Parcels & newspapers		0.44			0.38
Second Set:	Employment					
	Agriculture & forestry				-0.53	
	Food, drinks, tobacco		0.94	-0.33	-0.38	
	Coal and petroleum				-0.43	
	Chemicals					-0.54
	Metal manufacture	0.30				-0.69
	Mechanical engineering					0.71
	Instrument engineering			0.53		
	Textiles				0.34	
	Timber and furniture					0.45
	Other manufacture	1.21		1.15		
	Construction		0.30	-0.63	-0.69	-0.71
Canonical Correlations		0.90	0.77	0.68	0.53	0.50

* Only loadings greater than 0.30 are shown.

employment set. On this basis it may be suggested that areas with high employment in "other manufacture industries"[(1)] and metal manufacture are likely to use rail to transport their output of engineering products and to a lesser extent (weak correlation) their plastic products. The highest loadings on the second vector are from O_i foods and drinks, general chemicals and parcels and newspapers and from employment in food industries and construction. This result indicates that large urban centres generate a substantial tonnage by rail of chemicals and food products. The third vector represent high associations (i) between employment in instrument engineering and other manufacture industries and originating movement of engineering products; (ii) between employment in construction and originating shipments of plastics. The next canonical vector also shows two types of associations - the first between originating movement of oil and petroleum products and electrical equipment and employment in textiles. The second type of association is between O_i miscellaneous and employment in chemicals and construction. The last canonical vector suggests that areas with high employment in metal manufacture, chemicals and

---

(1) Note that S.I.C. order 19, that is "other manufacturing industries" is essentially an aggregate of plastics' industries as it includes rubber, linoleum, plastic floor etc. (see Appendix II.1).

construction generate a substantial volume of freight in fertilizers, oil and petroleum and miscellaneous. It also shows a plausible correlation for O_i timber and forest products, and textiles and furniture with employment in timber and furniture.

Summary

It has been difficult to adequately model rail freight with the explanatory variables considered by this research. The attempts at improving the models summarized in this section have been unsuccessful. It has been, however, possible to derive cononical correlations between sets of originating freight by commodity and sets of employment by S.I.C. orders. These results suggest that zone with substantial employment in construction, other manufacture, mechanical engineering, instrument engineering and chemicals engage significantly in rail freight. Furthermore, when modelling rail freight for all zones, it is worth noting that for a few commodity groups, no more than 15 out of the B.R. 134 areas have non-zero flows. In a distribution modelling exercise, this leads to highly degenerate problems; in the context of econometric models this feature causes "flat regressions" if the O.L.S. technique is used (Goldberger, 1964; Epperson et al, 1980). That is, the high frequency of zero O_i and D_j explains the persistent failure of rail models and to improve these it may be necessary to alter the original data base and use other statistical methods. These possible solutions are outlined in the concluding chapter.

9.3.2 Impact of Port Activity

The relevance of the distinction between port and major port zones from non-port zones when modelling total freight was assessed in Section 8.3. This section investigates the usefulness of such a stratification of areas when modelling freight by commodity group.

Commodity volumes by road were regressed on resident population and a dummy variable[1] representing port zones. Then a dummy variable representing major port zones was included in the equations. The results are set out in Table 9.9.

Port Zones: The port variable contributed significantly to the explanation of O_i and D_j for oil and petroleum products, chemicals, food and drinks, timber and potatoes, fruits and vegetables. Although statistically significant, the contribution of the port variable to the explained variation was substantial

(1) The dummy variables are defined in Section 8.3.

TABLE 9.9: TONNAGE OF COMMODITIES REGRESSED ON RESIDENT POPULATION AND DUMMY VARIABLES: PORTS AND MAJOR PORTS

B.R. Zones
F(0.001) =
$r_s(0.01) = \pm 0.30$
S.N.D.(I)= ± 2.57

Dependent Variables	Constant (b_o)	Coefficient (b_1)	Coefficient (b_2)	R^2	F	$r_s(K)$ *	S.N.D. (I)
A. Resident Population and Port Zones							
O_i							
Potatoes, fruit & veg.	-100.35	1.02	288.66	0.61	106.44	0.52, 0.06	1.68
Foods & drinks	- 19.68	1.83	255.25	0.63	109.65	0.40, 0.10	-2.14
Chemicals	- 10.04	0.62	129.85	0.41	46.81	0.54, 0.16	7.83
Timber	- 16.94	0.59	189.53	0.40	44.91	0.35, 0.13	-1.12
Oil & petroleum products	- 58.19	1.10	825.44	0.40	42.86	0.33, 0.29	0.24
D_j							
Potatoes, fruit & veg.	-155.34	1.17	162.97	0.68	140.36	0.41, 0.02	0.01
Foods & drinks	- 44.34	1.74	210.63	0.66	127.14	0.45, 0.09	-1.75
Chemicals	- 20.67	0.69	94.45	0.55	67.33	0.24, 0.03	6.92
Timber	40.59	0.53	110.35	0.50	67.33	0.24, 0.11	1.73
Oil & petroleum products	85.63	1.54	68.19	0.73	179.92	0.45,-0.33	0.95
B. Resident Population and Major Port Zones							
O_i							
Chemicals	6.14	0.57	327.58	0.56	69.65	0.60,-0.37	6.37
Plastics	5.71	0.65	29.57	0.27	24.60	0.31,-0.46	1.12
Timber	20.90	0.55	323.83	0.44	51.49	0.33,-0.32	-1.94
Oil & petroleum products	118.25	0.94	1284.58	0.42	39.45	0.34,-0.36	-1.80
D_j							
Electrical equipment	- 75.66	0.40	0.63	0.63	111.51	0.45, 0.31	-0.35

b_1: coefficient for population

b_2: coefficient for ports in A. and for major ports in B.

* r_s for the k^{th} explanatory variable; All coefficient satisfy the 0.01 level of significance.

only for oil and petroleum products and timber (7 and 4 per cent respectively). For the remaining commodity groups, the contribution of the port variable to the explained variation clustered around 2 per cent. The inclusion of this variable, however, reduced the bias in the residuals in all models with the exception of O_i and D_j for chemicals.

Major Port Zones: The inclusion of the major importing zones, dummy variable contributed significantly to the explanation of originating movements of plastics, chemicals, timber and oil and petroleum products. However, the major exporting zones dummy variable, proved to be significant only for the explanation of D_j electrical equipment. In terms of the relative contribution to the explained variation, this stratification of zones performed better than the simple distinction between port and non-port zones as the contribution ranged between 4 and 9 per cent. Again, with the exception of chemicals, a reduction of the bias in the residuals is observed. The error terms are spatially independent and only moderately heteroscedastic.

Summary

The results support the hypothesis put forward in Appendix 1 which states that foreign trade activity may have a significant impact on internal trade for a few commodity groups (see Table A.2, Appendix 1). However, the results suggest that the impact of port activity is more relevant for the main port and whereas imports play a significant role in inland originating freight, exports have a weaker effect on terminating traffic to major exporting zones. Further, this type of stratification of zones ensured the derivation of more robust models for commodity groups as the bias in the residuals was significantly reduced.

9.4 Comparison with the Regional Models (Effects of Spatial Aggregation)

This section evaluates the effects of spatial aggregation on the results and draws comparisons between the analyses at regional and zonal levels.

Areal aggregation results in smaller between zone variances compared to the variance between behavioural units (households, plants). As in the case of aggregation over industries, spatial aggregation is based on the assumption of homogeneity. Therefore, the efficiency of aggregation will be a function of the degree of homogeneity achieved in drawing the boundaries of the areal units.

A simple procedure that may be used to demonstrate the effects of aggregation consists of an examination of the distribution of variables at different levels of aggregation. For example the mean standard deviation for each variable shows the variability of the data with respect to the mean, and usually as the degree of aggregation increases so the range of variability decreases (Kassof and Deutschman, 1969). Consequently, the loss of variation in the data as aggregation take place may be demonstrated. However, more information may be gathered if analysis of variance is used to decompose the variation of the data in terms of the total sum of squares (TSS), the between-groups sum of square (BSS), and the within-groups sum of squares (WSS). The expression of WSS as a percentage of TSS provides an estimate of the degree of homogeneity achieved, with a lower percentage indicating greater homogeneity. As regression analysis works on the variation between groups, expressing BSS as a percentage of TSS enables the success of aggregation to be judged in terms of variation left to be explained. When there is a lack of homogeneity within the groups, it is possible for more variation to be left unexplained than is explained by the models (Kassof and Deutschman, 1969; Dalvi and Martin, 1977). Table 9.10 provides some examples.

These results indicate that by far the greatest proportion of the total variation is within the zones. In the best case, that of total freight by

TABLE 9.10: VARIATION WITH DATA AGGREGATION

Areal Units	Total: O_i by Road		Steel: O_i by Road		Steel: O_i by Rail	
	WSS as % of TSS	BSS as % of TSS	WSS as % of TSS	BSS as % of TSS	WSS as % of TSS	BSS as % of TSS
Zones (134)	0	100	0	100	0	100
Regions(10)	46	54	87	13	95	5

road, the $R^2 = 0.96$ obtained for example with resident population (at the regional level) means that 96 per cent of a relatively small fraction of the data is explained, whilst the seemingly lower $R^2 = 0.83$ obtained at the zonal level represents the explained proportion of a much greater total variance.

It may be concluded that, in terms of homogeneity, the regions are not an efficient areal framework, since, in the best case the variability within the groups amounts to 46 per cent and in the worst case to 95 per cent of the total variation.

As some zones are part of more than one region[1], another approach is used to illustrate the consequences of spatial aggregation for selected dependent and independent variables. The results are shown in Table 9.11. The means, standard deviations and the mean standard deviation as well as the loss of variation (Column Δ) are shown in that table. The standard deviations give a first hint about the reduction in dispersion of values but the last two columns are more informative. The mean standard deviations of chemicals, for instance, show that at the zonal level the values are scattered within a range defined by about $\pm$ 1.45 per cent of

(1) See Appendix III.2.

TABLE 9.10: MEANS AND STANDARD DEVIATIONS FOR VARIABLES AT VARYING LEVELS OF AGGREGATION

Areal Units	Mean	Standard Deviation	Mean Standard Deviation	Δ
	Total O_i by Road			
Zones	12471.79	10978.86	0.88	0.29
Regions	166458.79	97396.00	0.59	
	Total D_j by Road			
Zones	12470.73	11482.30	0.92	0.32
Regions	166578.51	100400.08	0.60	
	Total O_i by Rail			
Zones	1453.25	2738.74	1.88	1.14
Regions	19466.36	14448.36	0.74	
	Total D_j by Rail			
Zones	1453.25	2204.70	1.51	
Regions	19415.5	9236.39	0.48	1.03
	Chemicals O_i Road			
Zones	296.65	429.86	1.45	
Regions	3580.83	2179.86	0.61	0.84
	Furniture, textiles, etc. O_i Road			
Zones	408.15	691.36	1.69	
Regions	5506.121	3832.11	0.70	0.99
	Population			
Zones	401.46	420.77	1.04	0.24
Regions	5393.75	4404.45	0.80	
	Population density			
Zones	532.27	858.29	1.61	0.82
Regions	315.90	249.83	0.79	

the mean whereas the regional values cluster within only ± 0.61 per cent of the regional mean value. The biggest losses of variability is observed for rail freight and furniture and textiles by road; immediately followed by chemicals. In terms of earlier chapters, these commodity groups are characterized by relatively localized industries and therefore it may be suggested that the more the industries are localized the worse are the effects of aggregation.

However, the results also suggest that if the loss of variation for some variable is too great its values tend to be almost constant across the observations. This, in turn, lead to a lower R^2 at the aggregated level and often to a flat correlation with R^2 close to zero. For example, population density does not discrimate significantly between standard regions whilst for a between-zones comparison it is a good discrimatory variable. This has two implications. First, prior to including a variable in a regression, its coefficient of variation ought to be examined. Second the traditional correlation measures are not sensitive to such variable distributions as in the case of rail, with over 90 per cent of the values being zero. Therefore if the variables are of interest for some purpose, then other techniques or other statistics of goodness of fit ought to be considered.

As a conclusion, it may be said that if a great disparity exists within regions with respect to a given variable, the aggregation procedure lead to a great loss of variation, which is not accounted for by the usually "inflated R^2's". Consequently, the zonal models appear as more robust and certainly more reliable for forecasting purposes. Nevertheless, from the variety of models discussed, a consistency in the results does emerge for analyses at the regional and zonal levels and so gives support to the conclusions drawn in Part II. Regional models are therefore worthwhile from an explanatory point of view, they are not from a forecasting point of view as, in time more changes may result in sub-regions than expected on the basis of regional figures.

CHAPTER 10

ACCESSIBILITY AND FREIGHT GENERATION

10.1 Introduction

In Section 2.2, the relevance of accessibility in freight generation models has been outlined and it has been underlined that accessibility has not been accounted for in previous models; with the exception of Redding's unsuccessful attempt that was based on the crude variable, distance from Charing-Cross, (1972). Not until recently have attempts to incorporate accessibility into urban passenger trip-end models been relatively successful, Leake and Huzzayin (1979), Morris et al (1979). The main drawback of early accessibility measures is their coarseness in that they have not discriminated between categories of individuals and trip purposes, Koenig (1975). Another problem, though specific to the Hansen type of accessibility index, is one of multicollinearity between the indices and the remaining explanatory variables. However, these rather frustrating results have not weakened the belief that accessibility may constitute the policy variable par excellence for it is accepted, in theory, to be a determinant of transport demand and it may also be related to the adequacy of the transport system. Such a belief is illustrated in recent urban studies in two ways; on the one hand researchers have experimented with mathematical programming models in which accessibility is the objective function to be maximised (for example, Leonardi (1978)). On the other hand, accessibility based user benefit measures have been derived, Neuberger (1971), Williams and Senior (1977). All these works make use of the duality theory of mathematical programming on which the work reported in this chapter is also based.

Although these studies have inspired the derivation of the indices of accessibility used here, two points are worth noting. Firstly, this study is concerned with incorporating an accessibility measure in generation models, whereas the studies referred to above have sought primarily to establish benefit measures for evaluation purposes. Secondly, such studies have dealt with passenger trips while this research addresses multi-regional commodity shipments.

The ultimate question is whether the variation in originating and terminating is partly explained by the relative access of suppliers to markets and relative access of customers to sources of supply. The intermediate questions are how may accessibility be understood and how can it be measured in this context? The next section will underline the limits of traditional measures for the problem in hand and set out a basis for the indices to be used.

Section 10.3 shows the derivation of the measures and in Section 10.4 the results are discussed.

10.2 Limits of Previous Accessibility Measures

The Route Opportunity Approach

In a strict sense, accessibility has been defined as the relative ease with which distance is overcome. In this view, the relative transport cost, travel time or distance would be reasonable surrogates for accessibility. Operationally, the indices have been derived either (i) by means of graph theory methods of which Shimbel's index is most powerful, see Leake and Huzzayin (1979) or (ii) by other indices which vary in their degree of sophistication. The difference between various indices usually reflects the concern to produce useful measures which would preferably be different for private and public transport (see Leake and Huzzayin for a review). The advantage of such an approach from a transport planner's point of view is its concern with the performance of the transport system. However, such indices cannot indicate to the urban planner why two areas with identical "access" attract different demands, nor can they help assess traffic diversion resulting from the creation of a new facility as they overlook the supply of opportunities across areas. This particular point is documented in a recent study by Dolin and

Gibson (1979) who look at the effect of changes in the supply of recreational facilities on leisure trips.

The Location Opportunity Approach

The alternative approach to accessibility considers an element of distance friction as well as opportunities at the destination. In this view it is understood that improved accessibility may result from an upgrading of the transport system or from a modification of the locational pattern of opportunities.

Most authors who espouse this approach have derived their indices by calibrating a singly constrained gravity model. One of the differences between them is in the type of attraction measures selected which, in particular, are expected to be different for different trip purposes. The most commonly used attraction variables are the number of jobs for home-based work trips and shopping floor area for shopping trips. Another difference is the form of deterrence function, for example, exponential or power, Nakkash and Grecco (1972), Vickerman (1974), Koenig (1978).

Of the two alternative approaches to accessibility the location opportunity approach has the advantage of being more comprehensive in that distance friction is but a component in the measure of accessibility. The pitfall of considering a measure based solely on distance friction in a multi-regional freight context may best be appreciated by the following points. Firstly, planners have recognised that the idea of stimulating development in depressed areas by providing them with better access (i.e. investing in transport) would fail[1] unless complemented by other measures outside the transport sector (for the E.E.C. case, see Blonk (1977)). Secondly, recent research in location theory has pointed to the decline of the relative importance of transport costs as a factor of locational choice, Smith (1971), Norcliffe (1975). Such a decline is explained by the decreasing weight of transport costs in unit production costs. This is understandable in view of the progress of transport technology and changes in production processes. However, the incidence of such progress varies across industries, across countries and across regions. For example, in Sweden, Czamanski and Czamanski (1979), stress that "with the exception of building materials, heavy construction and iron and steel, there are few industries in which transport costs form more than an insignificant part of the unit production cost". In

---

(1) In fact, Blonk claims that "Cases are known in the Community of, for instance, a motorway built to an underdeveloped area without accompanying measures; instead of boosting the area in question, this has resulted in a vacuum".

the U.K., on the other hand, Edwards' works (1970, 1975) imply that spatial variation in transport cost is significant for the manufacturing sector. In addition, many processes are sensitive to delays in obtaining inputs, (Fahle and Hertzfield, 1972) and delays in shipments of output may put a strain on storage capacities. Therefore factors such as co-location remain important. Consequently, it would seem to be appropriate to take into account not only the distance element but also the opportunities. The apposite measure also needs to be industry specific.

Researchers in freight modelling have used measures of total employment and population potential to test some hypotheses on accessibility, Chisholm and O'Sullivan (1973), Gordon (1976). However, such aggregate measures are not to be preferred to indices that are industry specific. Furthermore, both of the above two measures (and their disaggregate counterparts which may be based on employment by industry) would be inappropriate in generation modelling as they are likely to be correlated with the other explanatory variables.

Another procedure which has proved useful in the urban context (Nakkash and Grecco, 1972) and in this research (see Chapter 8), is the stratification of zones into central and peripheral zones. Despite its usefulness such an approach is limited in that areas are allocated to groups in the hierarchy and consequently the zone specific position is overlooked. A continuous measure is thus preferable.

Accessibility may best be gauged in terms of benefits or advantages of particular areas. The approach suggested in this thesis is based on the premise that different regions possess an uneven distribution of opportunities which represent economic surpluses to those who have access to them. As these economic surpluses vary with relative accessibility the idea is related to the concern in urban studies about the degree to which accessibility affects "real income" of different community groups, Pahl (1971), Wachs and Kumagai (1973), Burns (1979). Indeed, rather than approaching accessibility in physical terms or solely in cost terms it is proposed to consider accessibility as relative benefit over space.

10.3 Derivation of the Measures

Interaction and Accessibility in the Multi-Regional Freight Context

In principle, accessibility measures are rooted in the empirical analysis of the type of spatial interaction under consideration. Interregional freight flows have usually been represented by the linear programming or by the gravity formulations. While it has been found that the linear programming formulation yields a good prediction for flows of homogeneous commodity groups, Chisholm and O'Sullivan (1973), Benabi (1978), overall, the performance of the gravity model has proved to be more successful, Pitfield (1978b). The strictly deductive nature of the linear programming model and its assumption of homogeneity make it unrealistic. In particular, it fails to represent the sub-optimality observed in real life and so it has been suggested that it would best describe the case of perfect competition or that of a centrally planned economy, Hay (1973). In existing centralised or decentralised economies, however, the observed patterns of freight movement are characterised by a sub-optimal dispersion of flows. Although the gravity model captures part of such a dispersion (see below), it has been criticised for its lack of a behavioural basis. Therefore, some effort has been devoted to reflecting this dispersion of flows and in providing a behavioural explanation for it. Two approaches, individual based methods and group based methods, have been adopted to these ends. However, with the data in hand, it is not possible to use models based on the behaviour of the individual shipper; the group locational surplus method is used. Nevertheless, so as to help in the interpretation of the group surplus model parameters, a method based on the individual decision maker, that is, the random utility method, is adapted to the freight case (see Appendix to this Chapter).

The Group Locational Surplus Model

A model that accounts for the dispersion of flows and conforms with the random utility approach but which is designed for aggregate data has been suggested by Neuberger (1971) and Champernowne et al (1976). It takes the form:-

$$\text{Max}\left\{LS\right\} = -\frac{1}{\beta}\sum_i\sum_j\left(T_{ij}\text{Ln}\,\frac{T_{ij}}{O_i D_j} - T_{ij}\right) - \sum_{i,j} T_{ij}C_{ij} \quad (1)$$

subject to

$$\sum_i T_{ij} = O_i \quad (2)$$

$$\sum_j T_{ij} = D_j \quad (3)$$

It may be seen from (1) that the first term of the objective function is the entropy which accounts for the diversity and that the second term is the objective function of the transportation problem. Clearly, if β tends to infinity, the first term vanishes and we are left with the transportation problem.

The solution to the programming problem (1) to (3) is given by:-

$$T_{ij} = O_i D_j\, e^{-\beta(\alpha_i + \gamma_j + C_{ij})} \quad (4)$$

where α_i and γ_j are the Lagrangian multipliers associated with the constraints (2) and (3). The Lagrangian or dual variables associated with the origins are given by:-

$$\alpha_i = - \beta^{-1} \operatorname{Ln} A_i \tag{5}$$

where, A_i is the usual balancing factor. The destination dual variables are calculated from:-

$$\gamma_j = - \beta^{-1} \operatorname{Ln} B_j \tag{6}$$

where, B_j is the balancing factor.

From (5) and (6) it may be seen that α_i and γ_j are equivalent to p_i^k and p_j^k in the random utility model[1] and so a similar interpretation may be attached to them. That is α_i are equilibrium prices that represent surplus for suppliers with respect to their access to markets and that the prices are an increasing function of the demand potential. Higher dual variable α_i will thus be associated with higher access. While γ_j is a mean utility for buyers and is a decreasing function of the supply potential, that is higher prices γ_j imply lower access to supply.

The dual variables α_i and γ_j are used as explanatory variables for originating and terminating

---

(1) See Appendix to this Chapter.

freight respectively. In addition, experiments with the following indices are carried out:-

$$I_i = \frac{\alpha_i}{U_i} \quad \text{and} \quad I_j = \frac{\gamma_j}{V_j} \ , \quad \text{where}$$

U_i and V_j[(1)] are origin and destination prices for the transportation programming model. The use of the indices I_i and I_j is justified as follows. Accessibility measures based on observed behaviour have been criticised as they "have built into them the inefficiencies of the existing system" Brehney (1979). In order to correct for this, researchers have recently used "Indicators which would enable an ideal (optimum) network to be compared with the existing network" Leake and Huzzayin (1979). By contrasting the locational surplus variables to the linear programming variables a similar result may be achieved as the former model best fits the observed data whereas the linear programming solution may be considered as the ideal. Consequently, zones with higher values of I_i and I_j would show closeness to the optimum.

10.4 Results

Calibration of Distribution Models

The group locational surplus maximizing model with a negative power deterrence function was calibrated using Williams (1976) acceleration of Hyman's (1969) method.

---

(1) U_i & V_j were derived in an earlier study at this department (Benabi, 1978).

The advantages in using this technique when calibrating gravity like distribution models are discussed in Pitfield (1978b).

The choice of the negative power deterrence function maybe justified on a priori grounds, as Wilson (1971) has suggested that such a function is appropriate to interregional long distance flows, whereas the negative exponential deterrence function is more appropriate to intra-urban short distance flows. Further previous work has found that such a specification of the deterrence function produces the best fit of the observed freight flow matrices,Pitfield (1978b).

Choice of Mode and Commodity Groups

Only commodity flows by road were considered. Six commodity groups are used in the regression analysis; steel, iron, vehicles, earths and stones, bricks and oil and petroleum products. These were selected so as to represent different ranges of values of the (β) parameter as illustrated in Table 10.1.

TABLE 10.1

RANGES OF β VALUES FOR THE SELECTED COMMODITIES

Range of Values for the β Parameter		
LOW	MEDIUM	HIGH
Steel	Oil and Petroleum	Earths & Stones
Iron		Bricks
Vehicles		

These values are inversely related to the index of industrial concentration as seen from Figure 10.1. In fact the rank correlation between the values of β and the index of concentration[1] yielded a value of the Spearman rank correlation coefficient of -0.64. This value satisfies the 0.001 level of significance.

It is argued in the appendix to this chapter that the more widespread is the availability of a commodity, the smaller will be the utility created by transport services and vice-versa. Consequently, it is hypothesised that the impact of the dual variables on freight generation will be greater for steel, iron and vehicles than for the remaining commodity groups selected for this analysis.

Regression Results

As the main purposes of this section is the study of the relevance of accessibility, the dual variable are incorporated into the simple regression models established in Section 9.1 so as to facilitate comparisons. However, on the resulting equations multicollinearity was suspected. In fact, the paired correlations between employment and the dual variables were

---

(1) This was the commonly used ratio of the four main shippers to the total. This simple index is thought to be appropriate as for all industries there is the same number of observations (134).

FIGURE 10.1: PLOT OF INDICES OF INDUSTRIAL CONCENTRATION AGAINST

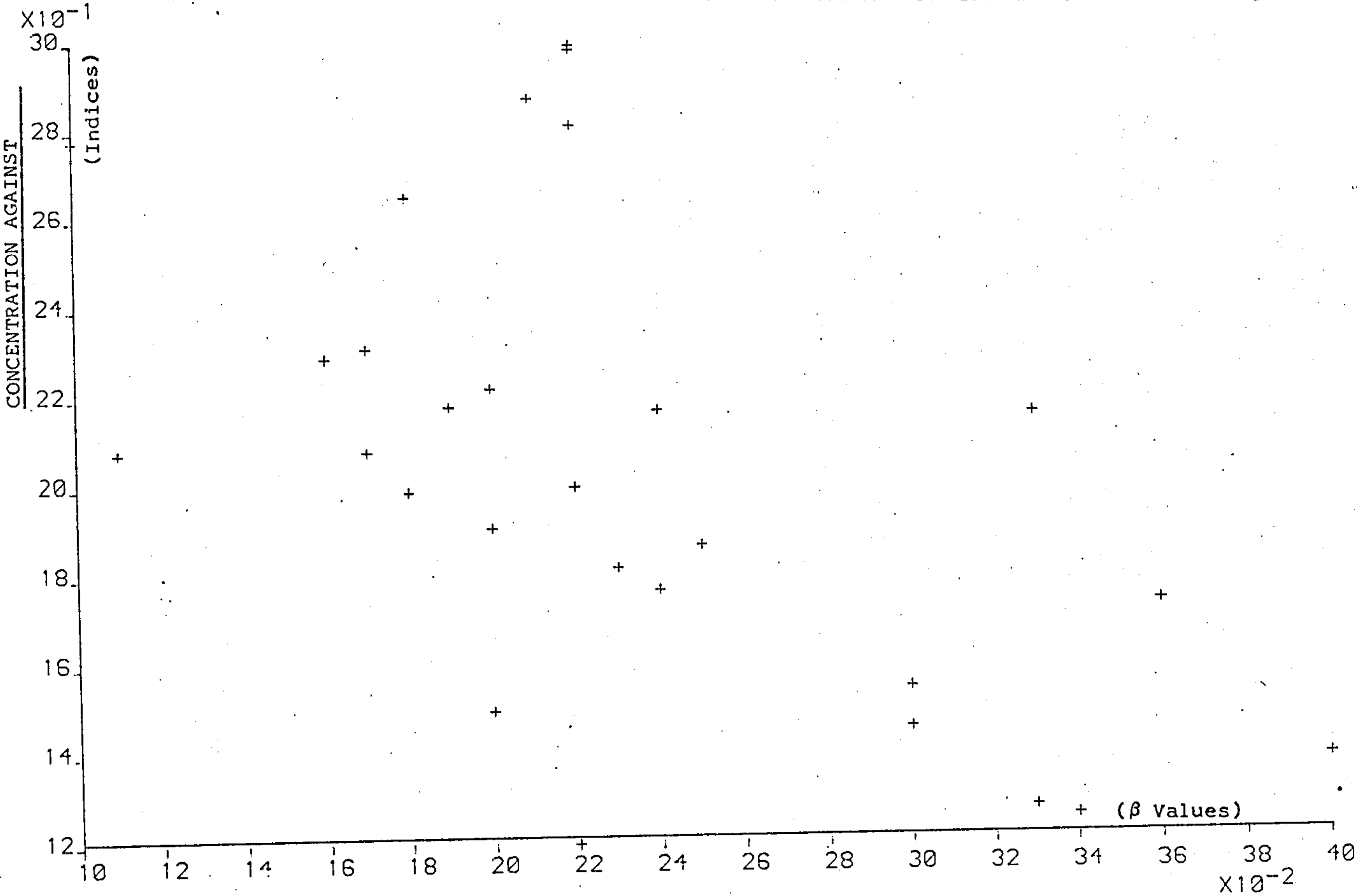


TABLE 10.2: TONNAGE OF COMMODITIES REGRESSED ON EMPLOYMENT AND THE DUAL VARIABLES $_i$ AND $_j$

B.R. Zones
$F(0.001) = 11.2$
$r_s(0.01) = \pm 2.57$

Dependent Variables	Constant b_o	Coefficient b_1	Coefficient b_2	R^2	F	r_s (k)	S.N.D. (I)
O_i							
Steel	-778.01	1.07 (0.85)*	251.39 (0.13)	0.80	265.68	0.19, 0.30	0.24
Iron	-318.09	1.23 (0.76)	159.40 (0.17)	0.67	134.20	0.42, 0.42	-1.06
Vehicles	- 54.95	0.15 (0.75)	44.15* (0.11)	0.63	113.65	0.70, 0.64	2.99
D_j							
Steel	269.03	0.79 (0.70)	-269.95 (-0.14)	0.60	100.76	0.33,-0.47	2.45
Iron	189.86	0.93 (0.67)	-199.73 (-0.26)	0.62	108.50	0.35,-0.61	-1.15
Oil & Petroleum	521.161	0.68 (0.55)	-432.53 (-0.29)	0.55	81.35	0.26,-0.33	5.48
Vehicles	72.61	0.13 (0.65)	- 74.30+ (-0.16)	0.59	95.05	0.52,-0.52	-1.80

* standardized regression coefficients are shown in brackets;

\+ significant at the 0.05 level of significance. The remaining coefficients are significant at the 0.01 level.

b_1 employment;

b_2 α_i and γ_j for O_i and D_j respectively.

high in the case of vehicles (0.64) and in the remaining cases varied from 0.21 to 0.41. The regression results are set out in Table 10.2. The equations for O_i and D_j earths and stones, bricks and O_i oil and petroleum products are not reported. Although the dual variables contributed significantly to the explained variation for these commodity groups, the general level of explanation was too low for the models to be usefully used in analysis of the standardized regression coefficients and for comparisons across commodities. The models that are shown in Table 10.2 are significant at the 0.001 level. The coefficients satisfy the 0.01 level of significance with the exception of the coefficients for O_i and D_j vehicles which only achieve the 0.05 level. These results for vehicle movements were expected in view of the high collinearity between the accessibility variables and employment in the vehicles industry.

As regards the O.L.S. assumptions, the inclusion of the dual variables had a beneficial influence only on the D_j equations for steel and vehicles as the respective values of the S.N.D.(I) decreased from 4.02 to 2.45 and from 2.25 to -1.80. However, the values for the test of heteroscedasticity have increased in all cases.

The regressions based on employment and the indices I_i and I_j, which contrast the locational surplus variables to the linear programming variables performed less well than the above equations with the exception of O_i oil and petroleum products.

Relative Effects of the Variables

As judged from the standardized regression coefficients (shown in brackets in Table 10.2), the impact of accessibility on freight generation is much lower relative to that of employment. However, a comparison across commodity groups reveals a higher influence on iron and steel than on the remaining commodity groups and thus gives support to the hypothesis made earlier. Furthermore the impact on D_j is higher than on O_i.

Finally, as collinearity usually leads to high joint effects of the explanatory variables on the dependent variable, the direct and joint effects of employment and accessibility were computed. The joint effects were not substantial and so enhanced the higher net effect of employment relative to accessibility.

Discussion

The analysis has indicated that the dual variables representing accessibility have a significant effect on the variation of originating and terminating freight by commodity. The signs of the coefficients are as expected, that is for terminating freight the higher the price γ_j, the lower is the attraction of freight and this may imply that inaccessible areas tend to specialize in industries which involve little transport. The sign for α_i is positive, that is the higher the rent on accessibility the higher the freight originating.

In addition, the fact that the effects of the dual prices on terminating traffic are stronger than on originating traffic suggests that buyers are more constrained by lack of accessibility; D_j reacts more sensitively to price than does O_i. The weaker sensitivity of O_i may be explained: (i) either by the fact that supplies have a monopoly position; (ii) or because they are constrained by productive capacities.

As hypothesized, the effect of accessibility varies across industries. It is worth noting that as suggested by the behavioural model, as $\beta \rightarrow 0$, the factor distance becomes important. In fact, the overall limited impact may be due to the use of distance as a proxy for transport costs. As freight patterns are likely to reflect overall distribution costs, similar price variables derived from a model with a properly specified generalized cost function may be more useful.

Finally, the analysis has revealed that, like the activity-accessibility measures, these variables are likely to be correlated with the principal explanatory variables.

APPENDIX TO CHAPTER 10

RANDOM UTILITY THEORY IN A FREIGHT CONTEXT

Random utility theory employs utility maximizing decision rules so as to explain the choice behaviour of individuals and in the transportation field was first used in binary mode choice models. It has recently been extended to multiple choice situations (mode, route, destination) by Williams (1977). The theory assumes that each individual picks that alternative which provides him the greatest net benefit (or utility). In the multi-regional freight context, the situation is as follows.

The demand for a good k from amongst the set of products K, may be satisfied at any location L_i. The individual is assumed to select a single alternative from the product set $L(X)K$. Let t^k_{ij} be the probabilities denoting how often individual h chooses product k from region i and define O^k_i and D^k_j as the volume of product k supplied in i and received in j respectively as a result of decisions of all individuals. In a micro-model, the objective is to find the probability P^k_{ij} that sum of individuals in region j choose product k from region i, in order to derive the expected trade flow.

$$T^k_{ij} = D^k_j \; P^k_{ij} \qquad (1)$$

where, T^k_{ij} is the sum of all individuals decisions t^k_{ijh}.

Assume the decisions that underly P^k_{ij} are basedon the net surplus $S^k_i = p^k_i - \mathcal{E}^k_i$ (2)

where, p^k_i is the exfactory price of product k in region i and $\mathcal{E}^k_i$ is a preference or utility indicator for product k supplied by region i.

The assumption that $\mathcal{E}^k_i$ is a random variable and that all $\mathcal{E}^k_i$ belong to the same probability distribution (which is assumed to be of a Weibull form) enables the derivation of a gravity type formula as demonstrated by Sheppard (1980). This is:-

$$P^k_{ij} = \frac{O^k_i \exp\left[-\beta(p^k_i + c_{ij})\right]}{\sum_i O^k_i \exp\left[-\beta(p_i + c_{ij})\right]} = O^k_i \exp\left[\beta(p^k_j - p^k_i - c_{ij})\right] \quad (3)$$

where, $p^k_j = -\frac{1}{\beta} \text{Ln} \sum_i O_i \left[-\beta(p^k_i + c_{ij})\right]$ (4)

From (1) and (3), the interregional flows can be derived by:-

$$T^k_{ij} = D^k_j P^k_{ij} = O^k_i D^k_j \exp\left[\beta(p^k_j - p^k_i - c_{ij})\right] \quad (5)$$

Letting $p^k_i = -\beta^{-1} \text{Ln}\, A_i$ (6)

$p^k_j = -\beta^{-1} \text{Ln}\, B_j$ (7)

Substituting in (5) leads to:-

$$T_{ij}^{k} = O_i^k D_j^k A_i B_j^{-\beta c_{ij}} \qquad (8)$$

which, with the equilibrium conditions:-

$$\sum_j T_{ij}^k = O_i^k \qquad (9)$$

$$\sum_i T_{ij}^k = D_j^k \qquad (10)$$

gives the well known doubly constrained gravity model.

C_{ij} is the usual measure of separation, that is the cost of transport and β is the dispersion parameter which is equivalent to $\Pi/6^{\frac{1}{2}}$; σ is the standard deviation of the Weibull distribution. In random utility theory, σ^2 represents the variance of the stochastic preference terms ε_i^k. When β tends to ∞, the preference term σ^2 tends to 0. This occurs when the nearest location is patronized as assumed for example by linear programming. Random utility theory explains the dispersion of preferences as utility may be gained by engaging in business with more distant markets. Cost minimization is not the only basis for the dispersion of preferences.

In a multi-regional freight case at least four factors are likely to influence β. These are:-

(i) The relative contribution of transport services to increased consumer surplus. It may be argued that if a commodity enjoys a widespread availability, then the additional utility created by transport is small. By contrast, for commodity groups whose production is concentrated in a few locations over the country, the utility created by transport is high.

(ii) The second factor is the relative price of the commodity to the cost of transport. A highly valued good can absorb a substantial increase in transport cost, but a low valued good cannot.

(iii) The third factor may be called behavioural sources of dispersion of flows. These sources can be, for example, misinformation on the location of opportunities, different perception of prices across individual firms, constraints on behaviour due to organisation; for instance, corporations, nationalised enterprises or simply psychological constraints due to habits.

(iv) Finally, the measure of cost used. In this research, for instance, the distance is used as a surrogate for cost and this has implications for the form of the deterrence function used as this depends on the hypothesized relationship between transport costs and distance. Further, the actual distribution of commodities will reflect the total distribution costs, not just those of transport.

With the data in hand, only the first factor was tested for. The relationship between β and the index of industrial concentration is discussed in Chapter 10.

The prices p_i^k and p_j^k can be derived as dual variables associated with the constraints (9) and (10) respectively. The prices p_j^k have been shown to be utility indicators associated with access to opportunities (neuberger, 1971; Williams, 1976). This fact may also be seen from Equation (4). The prices p_i^k appear as equilibrium prices in the gravity formula (3) and may be viewed as a rent on the comparative advantage of the supply regions with respect to their access to markets. This interpretation also coincides with the analysis in terms of potential as p_i^k is an increasing function of the demand potential defined by:-

$$M_j = \sum_i T_{ij}^k / e^{\beta c_{ij}}$$

while p_j^k is a decreasing function of the supply potential defined by:-

$$M_i = \sum_j T_{ij}^k / e^{\beta c_{ij}}$$

PART IV

CONCLUSIONS AND LIST OF REFERENCES

++++++++++++++++++++++++++++++++

"One is entitled to be right but, with facts, one possess"

Thoughts of a biologist

Author's translation, from:

"On n'a le droit d'avoir raison qu' avec les faits dont on dispose"

Jean Rostand,
Pensées d'un biologist

+++++++++++++++++++++++++++++++

CHAPTER 11

CONCLUSIONS

This final chapter is concerned with three issues. The first is to summarize the main conclusions that have emerge from the empirical work and draw general conclusions. The second is to outline the potential for future research and finally, to draw planning implications from the major points raised in the thesis.

++++++++++++++++++++

The starting point of the study was the fact that the potential uses of multi-regional freight generation models are many and yet the area appears to have received little attention. The objectives of the thesis were to review previous work, to examine what could be done to further knowledge on freight generation modelling on the basis of available data and to see what may be learnt from the models.

In Chapter 2, the findings of both urban and multi-regional freight studies were examined. Some deficiencies have been revealed in both types of study, in particular, the neglect of a number of relevant determinants of freight movement and some statistical shortcomings in the calibration of models. Chapter 3 outlines the difficulties in collecting and assembling socio-economic data for the 134 British Rail zones and justifies the use of the standard regions as a spatial framework for an exploratory study. In Chapter 4, the main analytical techniques used are described. Although, regression analysis has been found to be appropriate, the type of data in hand and the problems investigated have suggested the potential usefulness of principal component analysis and analysis of covariance. In addition, an explanation of the reasons for the violation of the O.L.S. assumptions in cross-section studies of freight are provided along with simple corrective procedures.

These first four chapters represent the framework of the study. The most salient conclusions to be drawn from the subsequent analysis reported in Parts II and III may be categorized under four headings; these are choice of variables, aggregation effects, analytical techniques and mathematical forms of the relationships.

9.1 CHOICE OF VARIABLES

The determinants of freight generation that have been examined may be grouped into three categories; the socio-economic variables, the related industries variables and the location specific variables.

Socio-Economic Variables

The major conclusions to be drawn from the relative performance of the different socio-economic variables examined at the regional and zonal levels are as follows.

First, this research has confirmed previous findings on the usefulness of resident population and total employment when modelling total freight generation. However, for analysis of commodity movement, resident population has a limited usefulness compared to employment by industry in that it well explains only a few commodity groups and in general yields poor results for total and commodity movements by rail.

Second, data availability at the regional level has enabled the performance of regional G.D.P., net output by industry and measures of production and consumption to be examined. The first variable has proved to be a good explanatory variable for aggregated freight whereas the second has well explained movements of goods that are characterised by high value/weight ratios. The third type of variable has been found to be useful in modelling total freight and freight movements of the respective industries.

Third, the determinants of originating and terminating freight are different especially in the case of disaggregated freight. It is important, however, to bear in mind that the level of spatial aggregation directly affects the anticipated performance of a variable as a surrogate for production or consumption (see below). Nevertheless, it may be said that resident population as well as the measures of consumption give better fit for terminating traffic whereas employment and the measures of production best explains originating freight.

The Related Industries Variables

The exercise with principal component analysis have proved rewarding as (i) the major dimension of regional and zonal industrial structure have been isolated; (ii) a consistency has emerged between different sets of variables at the regional level and at different spatial levels for employment by S.I.C. orders. In addition, the models so derived, in particular at the zonal level have proved to be more plausible than those described in previous work (Department of the Environment, 1971; Chisholm and O'Sullivan, 1973). The principal components based models have shown that the main determinants of freight by commodity group is employment in the industry producing the good and in client industries. Some models have included variables representing spatial associations rather than industrial linkages.

The Location Specific Variables

Two questions have been investigated; these are whether accessibility affects freight generation and whether imports and exports have an impact on the volume of inland freight to and from port zones?

Accessibility

This question has been examined for total freight and for commodity movements.

For total freight, the centre-periphery model has been found to be a relevant explanation. This analysis has shown that when zones are grouped into central and peripheral places, the predicting equations appropriate to each group differ both in intercept and in slope coefficients. Furthermore, the inclusion of dummy variables representing zonal group affiliation improve the predictive ability of the models. For commodity freight, dual variables derived from the calibration of the group locational surplus model have been experimented with. For the commodity groups examined, it has been found that accessibility, as represented by the dual variables, has a significant effect on freight generation. As hypothesized, the impact of accessibility varies with the commodity group considered and is weaker than the impact of employment variables. Also terminating traffic is more sensitive to accessibility than originating freight.

Port Zones

Previous research has suggested that foreign trade may have some impact on inland flows for some commodity groups but may not affect total freight (Chisholm and O'Sullivan, 1973; Pitfield, 1978a). This research has shown that major ports are significantly different from other zones when total freight is considered but when looking at freight movement for given types of commodities, both ports and major ports have generation patterns significantly different from non-port zones. Foreign trade at major ports has a substantial effect and its consideration improves the reliability of the models. In addition, as may be expected in view of the composition of British foreign trade, the impact of imports on originating movement is stronger than that of exports on terminating freight to ports.

9.2 AGGREGATION

Aggregation Over Modes

Although the results obtained for total freight by road and rail are statistically significant, it is advisable to avoid such an aggregation since modal characteristics are greatly different, for example in mean haul or in terms of products carried. However, it was thought that if rail freight generation was difficult to model successfully, a prediction for rail could be derived from knowledge of forecasts for road and for the aggregate road plus rail (Chisholm & O'Sullivan, 1973)

Aggregation Over Industries

The assumption of homogeneity across industries has been investigated and great dissimilarities in freight generation per employee have been found even within the same industrial order. Consequently, a fine level of aggregation approaching the M.L.H. sub-divisions is needed if accuracy is sought in the models.

A further point which related to the question of aggregation over industries is one of compatibility in the level of aggregation of the dependent and independent variables. In this respect, unlike previous multi-regional studies, some effort has been devoted to assembling variables representing employment in industries that best match the definition of the commodity groups. This exercise has been rewarding, since the simple regression models derived perform as well as and sometimes, better than the multiple regression models based on employment by S.I.C.

order. In particular, it has been found that the more aggregated the independent variable is relative to the dependent variable, the higher is the degree of heteroscedasticity in the residuals.

Spatial Aggregation

The empirical consequence of spatial aggregation has been found to be a loss of variation and this varies amongst variables depending on whether, initially, they represent dispersed or highly localized industries. Such a loss of variation usually leads to inflated correlation, but if too great, it may affect the explanatory power of variables and lead to flat correlations with R^2's close to zero.

9.3 ANALYTICAL TECHNIQUES

Regression analysis has proved to be a useful analytical technique and when supplemented by other multivariate techniques, it constitute a particularly powerful tool. However, it has often been used in a mechanistic fashion as the main apparent concern in previous work has been the achievement of a good fit. This research has shown that such an approach may lead to, at least, three major shortcomings in the analysis of freight generation.

First and foremost, Starkie's (1974) criticism of previous work for not considering the assumptions that underlie least squares estimation, has been found to be relevant. In fact, results from the tests on O.L.S. assumptions, cast doubt on the conclusions that have been put forward in previous studies. This is especially so, because, the relative performance of alternative variables and alternative mathematical forms has been solely based on the level of explanation achieved. However, an interesting outcome of the examination of the O.L.S. assumptions has been that an increase in the coefficient of determination (R^2) may be achieved at the cost of violating the assumptions. Hence the criterion of model reliability may overule that of apparently explained variance. Furthermore, as the coefficients of the variable which causes heteroscedasticity are overestimated, conclusions from any comparison of the b or β coefficients ought to be considered with caution.

Secondly, most variables included in previous work have been selected on an ad-hoc basis, for example, on the basis of results achieved rather than on an a priori theory of freight generation. Consequently, the models appear to be associative rather than causal. An example of this kind of analysis is the study of freight generation by Chisholm and O'Sullivan (1973) where step-wise regressions of 15 aggregated commodity groups on 24 employment variables are likely to pick-up significant variables. By contrast, the theory presented in Chapter 2 of this thesis has led to a successful search for the major dimensions in the spatial distribution of industries and so causal relations have been distinguished from mere spatial associations. In this respect principal component analysis has been valuable for its explanatory role and for the statistically desirable property of independence in the variables it helps select. Furthermore, due attention has been devoted to the examination of residuals over the zones and such an exercise has aided the search for meaningful variables to be included in the regressions.

Thirdly, the reason why previous investigations (Redding, 1972; G.L.C. and Metra Group Ltd., 1970) favour simple regressions has been considered. This study has revealed that one reason for this preference is multi-collinearity. Another reason is the effects of variables referred to in econometric literature as "dominant variables".

9.4 MATHEMATICAL FORMS OF THE RELATIONSHIPS

As judged from the coefficients of determination and the tests and plots of the residuals, the appropriate functional form has been linear. When R^2's have been too low to enable sound conclusions to be drawn from the residual plots, the skewness and kurtosis of the variables distributions have been inspected but the different functional forms suggested have not yielded satisfactory results.

As the assumption of homoscedastic residuals has been violated relatively frequently, a weakness of this analysis has been the choice of the Spearman test rather than the Glejser test. An advantage of the latter test over the Spearman test is that it also gives information on the form of heteroscedasticity and so may be useful in deciding what transformation is required. Nevertheless, the Glejser test is less powerful according to Johnston (1972, p.221) and combined with the spatial autocorrelation test would have put considerable strain on the computing time.

9.5 SUGGESTIONS FOR FURTHER STUDY

Data Related Suggestions

The difficulties encountered in gathering data for the zonal analysis were mostly due to the uniqueness of the B.R. zonal system. A convergence towards a more uniform zonal framework for which a variety of information was available, would be most helpful for this kind of study. Further, as there is a strong interdependence between the zoning system and the results, a number of algorithms to derive optimal zoning systems have been suggested (Hannan and Burnstein, 1974; Openshaw, 1978). If these procedures are found to be time consuming, an alternative approach would be to group the zones according to the corrected Geary's index as shown in Lebart (1966).

Improvement of Rail Models

The attempts to model rail freight have rested on the use of the original data. Two methods that alter the original data may aid explanation. The first method would lead to a reduction in the number of commodities and zones by means of a Q-Mode factor analysis. The factor analysis used in this thesis is known as R-Mode with the objective of indentifying relationships between the variables. In Q-Mode factor analysis, it is the individuals or the zones that are factor analysed. Only a limited number of zones would load significantly on the main factors representing commodity groups and so regression models may be run for the significant zones and commodities. The second approach stems from the fact that as the value of variables are zero for a large proportion of the observations, O.L.S. estimation is not appropriate (Goldberger, 1964; Epperson et al, 1980). However, as the alternative method, these authors have suggested, that is the Tobit analysis, deals only with binary data (0-1), its use implies changing the original tonnages recorded for the zones that generate freight to the value 1. Nevertheless, the Tobit model has several desirable characteristics. First, it enables the derivation of the rate of participation of zones in rail freight. Secondly, it takes into account the high frequencies of zero values in the dependent variable.

Other General Suggestions

If it is found in any forecasting application of generation models that the variable resident population is important both for its availability and projectibility, some prior corrections may strengthen the models. First, a correction that takes into account cross-zonal commuting would improve the O_i equations as around conurbations there is substantial cross-zonal commuting. Secondly, a correction for spatial component, as in Chapter 8, so as to remove the contiguity effects or having recourse to the approach suggested by Moellering and Tobler (1972) that is related to the nested design in analysis of variance.

9.6 SOME PLANNING IMPLICATIONS

An initial point to be made is that rail movements are largely dominated by a few industrial sectors especially coal, iron, ores and steel. Therefore, changes in rail freight patterns may be dependent on a limited number of decisions. These, in turn, are largely related to government action for the sectors in question are dominated by nationalized industries. However, if major changes in population patterns resulting from for example the creation of new growth centres, occurred, although rail freight may be little affected, a substantial impact on road freight patterns may be expected.

As regards the evaluation of the economic feasibility of road investments, it has been shown that the output of generation models may be useful (Gw illiam 1972; Dodgson, 1974). More important, perhaps, is whether improved accessibility has a real impact on regional development, in particular, in less-developed areas. It is true that improved accessibility decreases transport costs, yet there is a conflict in regional economics on the importance of transport costs in influencing patterns of regional prosperity in Britain (Chisholm and O'Sullivan, 1973). This study has shown that in terms of total freight, peripheral areas do seek to avoid some of the penalties of their inaccessible locations by engaging in activities that require smaller tonnages of freight than central areas. The analysis at commodity level by contrast, suggests that when transport costs are represented by distance, accessibility has a relatively limited impact. However, evidence show that transport costs become more important when supply is relatively more concentrated. This suggests that accessibility may be achieved by reorganizing opportunities over space which, in turn, supports the suggestion that to be more effective in regional policy, transport investments must be complemented with measures such as the relocation of industries. In this respect, the principal component analysis models may be useful in monitoring the likely effects of relocation on freight movements between related industries.

+++++++++++++++++++

Various models have been examined in this research with the objectives of improving the models' reliability and potential usefulness in explanation so as to aid the assessment of policies. The potential usefulness must, however, be related to the requirements of the user and to the time perspective.

Any use of the models in aiding decision making is constrained by the availability and projectability of the variables included. Whilst projections are basically dependent upon the time span considered, the availability of data depends upon traditions in governmental data collection and the confidentiality attached to some information by enterprises. Consequently, although the models based on employment by industry appear to be more reliable than the resident population models, the latter ought to be considered for long term nationwide forecasts or for forecasts in connection with major investments, such as construction of new towns, for at least two reasons.

(i) Demographic forecasts are relatively advanced and data are usually available at a variety of administrative levels.

(ii) The employment variables, by contrast, are relatively more difficult to forecast in view of technological and economic changes that may occur.

However, analysis of spatial and industrial linkages has shown that employment data in relatively few industries are needed to explain freight generation and preliminary tests show that freight in one industry may be estimated from that of a related industry. Hence it may be worthwhile, at least for medium term forecasts, to concentrate on gathering information for these few industries and then to make use of sets of simultaneous equations.

Finally, any success in gaining further knowledge of the multi-regional freight generation process is likely to benefit general transport planning exercises. The issues of accessibility and foreign trade dealt with in this research may also be related to the distribution phase. For example, when the impact of foreign trade on generation patterns is found to be relevant, the hypothesis of a closed economy needs to be relaxed. Furthermore, the relevance of location specific variables in freight generation imply their potential usefulness in modal split models. At the very least, the multivariate methods used in the generation phase will be valuable in analyzing, for example, consignments characteristics or shippers perceptions of modal characteristics. Most simply, output from generation models that are capable of being used in a forecasting sense, are a necessary requirement for the pursuit of a generation-distribution-modal split model framework for some design year.

REFERENCES

BAXTER, R.S. (1976): Computer and statistical techniques for planners. (Methuem & Co. Ltd., London.)

BENABI, B. (1978): An analysis of British interzonal freight flows: an application of Linear programming. (Unpublished, M.Sc. thesis, University of Technology, Loughborough, Leics.)

BREHNEY, M.J. (1978): "The measurement of spatial opportunity in strategic planning", Regional Studies, 12, 463-479.

BRITISH RAIL, (1975): Freight market background study: Report, (British Rail, strategic planning studies London, mimeographed).

BROWN, A.J. (1972): The framework of regional economics in the U.K., (Cambridge University Press, London).

BURNS, L.D. (1979): Transport, temporal, and spatial components of accessibility. (D.C. Heath and Company, Massachusetts).

CHAMPERNOWNE, A.F. et al (1976): "Some comments on urban travel demand analysis, model calibration and the economic evaluation of transport plans", Journal of transport economics & policy, 10, 267-285.

CHILD, D. (1971): The essentials of factor analysis, (Rinhart and Winston Ltd., London).

CHISHOLM, M. & O'Sullivan, P. (1973): Freight flows and spatial aspects of the British economy. (Cambridge University Press, London.)

CLIFF, A.D. & ORD, J.K. (1973): Spatial autocorrelation, (Pion, London).

Commissariat general du plan (1972): Transport, (Librairie Armand Colin, Paris).

Commission for European Communities (1975): The United Kingdom economy. Economic and financial series, 9 (N.I.E.S.R., London).

CZAMANSKI, S. & CZAMANSKI, D.Z. (1979): "A contribution to the study of locational preferences of industries", Paper presented to the European Congress of the Regional Science Association, London.

DALVI, M.Q. & MARTIN, K.M. (1977): "The effects of spatial aggregation on predicting trip rates: some empitical results", W.P. 89, (Institute of transport studies, University of Leeds, Leeds).

Department of the Environment (1971): "Initial attempts at modelling road freight flows", Note 211 Mathematical Advisory Unit, D.O.E. (Mimeographed).

DODGSON, J.S. (1974): "Motorway investment, industrial transport costs and sub-regional growth: a case study of the M62", Regional Studies 8, 75-91.

DOLIN, J. & GIBSON, J.G. (1979): "The demand for new recreational facilities: a Coventry case study", Regional Studies, 13, 181-190.

DRAPER, N.R. & SMITH, H. (1966): Applied regression analysis, (John Wiley & Sons, Inc. New York).

EBDON, D.E. (1977): Statistics in geography: a practical approach (Basil Blackwell, Oxford).

EDWARDS, S.L. (1970): "Transport cost in British industry", Journal of transport economics & policy, 4, 265-285.

EDWARDS, S.L. (1975): "Regional variation in freight cost", Journal of transport economics & policy, 9, 115-126.

EPPERSON, J.E. et al (1980): "Program planning with tobit analysis - The U.S. ford stamp case." Socio-Economic Planning Science, 14, 91-95.

FAHLE, V.I. & HERTZFIELD, H.R. (1972): "The role of transport costs and market size in treshold models of industrial location", Paper R.S.A. 28, 189-202.

FLEET, C.R. & ROBERTSON, R.S. (1968): "Trip generation in the transportation planning process", Highway Research Record, 240, 11-25.

G.L.C. & Metra Group Ltd. (1970): "Study of generation of goods vehicle movements in selected town centres. Vols. I to IV (mimeographed).

GOLDBERGER, A.S. (1964): Econometric Theory, (Wiley, New York).

GORDON, I.R. (1976): "Gravity demand functions, accessibility and regional trade", Regional Studies, 10, 25-37.

GWILLIAM, K.M. (1972): "Economic evaluation of urban transport projects: the state of the art", Transportation Planning and Technology, 1, 123-142.

HANNAN, M. & BURNSTEIN, L. (1974): "Estimation from grouped observations", American Sociological Review, 39, 374-392.

HAY, A. (1973): "Transport for the space economy, (MacMillan Press Ltd., London).

HEYMAN (see Department of the Environment, 1971).

H.M.S.O., (1975): The economic activity, County leaflets: general explanatory notes. (H.M.S.O., London).

HYMAN, G.M. (1969): "Calibration of trip distribution models". Environment and Planning 1, 105-112.

JOHNSTON, J. (1972): Econometric methods, second edition. (McGraw-Hill, New York).

KASSOF, H. & DEUTSCHMAN, H.D. (1969): "Trip-generation: a critical appraisal", Highway Research record 297, 15-30.

KEEBLE, D. (1976): Industrial location and planning in the U.K. (Methuen Edition, London).

KING, L.J. (1970): Statistical Analysis in Geography, (Prentice-hall, New York).

KOENIG, J.G. (1975): "A theory of urban accessibility: a new working tool for the urban planner" in urban traffic models PTRC Summer Meeting, University of Warwick, Coventry, England.

KOUTSOYIANNIS, A. (1973): Theory of econometrics. (MacMillan, London).

LAWLEY, A.E. & MAXWELL, D.N. (1971): Factor analysis as a statistical method, Second Edition, (Butterworths & Co. Ltd., London).

LEAKE, G.R. & GAN, R.C.H. (1973): "Traffic generation of selected industrial groups", Traffic engineering and control, 15, 349-353.

LEAKE, G.R. & GRAY, J.E. (1979): "Trip generation of selected industrial groups", WP 113 (Institute of transport studies, University of Leeds, Leeds).

LEAKE, G.R. & HUZZAYIN, A.S. (1979): "Accessibility measures and their suitability for use in generation models", Traffic engineering & control, 20, 566-572.

LEBART, L. (1966): _Les variables socio-economiques departementales ou regionales_, (I.E.D.E.S., Paris).

LOELB, S.A. & KENNETH, W.C. (1976): "Aspects of demand for urban goods movements in city centres", _Transport Research record_, 591, 38-40.

LEONARDI, G. (1978): "Optimum facility location by accessibility maximizing" _Environment and Planning A_, 10, 1287-1306.

LEVER, W.F. (1972): "Industrial movements, spatial association and functional linkages", _Regional Studies_, 6, 371-384.

LEVER, W.F. (1975): "Mobile industry and levels of integration in sub-regional economic structure". _Regional Studies_, 9, 265-278.

MALTBY, D. (1973): "Traffic models for goods, services and business movement to manufacturing establishment" _Transport Planning and Technology_, 2, 21-39.

MORRIS, J.M. et al (1979): "Accessibility indicators for transport planning", _Transportation Research_, 13A, 91-109.

MOELLERING, H. & TOBLER, W. (1972): "Geographical variances", _Geographical Analysis_, 4, 34-50.

NAKKASH, T.Z. & GRECCO, W.L. (1972): "Activity-accessibility models of trip generation", _Highway Research record_, 392.

NEUBERGER, H.L.I. (1971): "User benefit in the evaluation of transport and land use plans", Journal of transport economics and policy, 5, 52-75.

NIE, N.H. et al : Statistical package for the social sciences (Second Edition), (MacGraw-Hill, New York).

NORCLIFFE, G.B. (1975): "A theory of manufacturing places", (in Lyndhurst and David F. Walker eds), Locational Dynamics of Manufacturing Activity, (John Wiley & Sons, Inc., New York).

OPENSHAW, S. (1978): "A empipical study of some zone-design criteria", Environment and Planning A, 10, 781-794.

PIDOT, G.B.J.R. (1969): "A principal component analysis of the determinant of local government fiscal patterns", the Review of Economics and Statistics, 51, 176-185.

PITFIELD, D.E. (1977): "The freight distribution data", Department of Transport Technology, University of Technology, Loughborough, Leics.,(mimeographed).

PITFIELD, D.E. (1978a): "The volume of internal and external trade: is Britain a closed economy?" Regional Studies, 12.

PITFIELD, D.E. (1978b): "Freight distribution model predictions compared: a test of hypotheses". Environment and Planning A, 10, 813-836.

RAO & MILLER. R. LeRoy (1971): Applied econometrics (Wadsworth Publishing Company, Inc., Belmont, California.)

REDDING, B.G. (1972): "Traffic generated by industrial premises in London", Traffic Engineering and Control, 14, 170-173.

ROBERTS, P.O. (1977): "Forecasting freight demand", Proceedings of the World Conference on Transport Research, (Martinuso Nijhoff, The Hague).

ROBSON, B.T. (1971): Urban analysis: a study of city structure with special reference to Sunderland. (Cambridge University Press, Cambridge.)

SLAVIN, H.L. (1976): "Demand for urban goods vehicle trip", Transportation Research Record, 591, 31-37.

SHEPPARD, E.S. (1978): "Theoretical underpinnings of the gravity hypothesis", Geographical Analysis, 10, 387-402.

SMITH, D.M. : Industrial location: an economic geographic analysis, (John Wiley & Sons Ltd., New York).

SNECEDOR, G.W. & COCHRANE, W.G. (1969): Statistical methods, (Iowa University Press, Ames, Iowa).

STARKIE, D.N.M. (1967): Traffic and industry: A study of traffic generation and spatial interaction. (London School of Economics and Political Science, Geographical Papers.)

STARKIE, D.N.M. (1971): "The treatment of curvilinearities in the calibration of trip-end models", Universities Transport Study Group, Sheffield Conference.

STARKIE, D.N.M. (1974): "Forecasting urban truck trips in the U.K.", Transportation Research Record 496, 28-35.

VICKERMAN, R.W. (1974): "Accessibility, attraction and potential: a review of some concepts and their use in determining mobility", Environment and Planning A, 6, 675-691.

WACHS, M. & KUMAGAI, T.G. (1973): "Physical accessibility as a social indicator". Socio-Economic Planning Science, 7, 437-456.

WATSON, P.L. (1975): Urban goods movement, (D.C. Heath and Company, Massachusetts).

WILSON, A.G. (1971): "A family of spatial interaction models, and associated developments", Environment and Planning, 3, 1-32.

WILLIAMS, H.C.W.L. (1976): "Travel demand models, duality relations and user benefit analysis", Journal of Regional Science, 16, 147-177.

WILLIAMS, H.C.W.L. (1977): "On the formation of travel demand models and economic evaluation of user benefit", Environment and Planning A, 9, 285-344.

WILLIAMS, H.C.W.L. & SENIOR, M.L. (1977): "Model based transport policy assessment 2: removing fundamental inconsistencies from the models", Traffic Engineering and Control, 18, 464-469.

WILLIAMS, I. (1976): "A comparison of some calibration techniques for doubly constrained models with an exponential cost function", Transportation Research, 10, 91-104.

APPENDIX I

FOREIGN TRADE VARIABLES

1. The Problem

Some proportion of originating traffic in port zones will consist of imported goods and terminating traffic in these zones will be partly goods for export. The question is whether for these zones the proportion of external traffic in internal flows is substantial enough to warrant inclusion of variables to account for foreign trade. Chisholm and O'Sullivan (1973) concluded that in contrast to an analysis in value terms, within an analysis in volume terms, the British economy can be regarded as closed since the share of foreign trade in inland flows was small. In this study, the question is more rigourously dealt with. This Appendix provides the background to the statistical analysis described in Section 8.3.

2. Total Traffic and Difference between Volume and Value

Some 351.9 million tons of traffic passed through the ports of Britain in 1972. Imports valued at £9,136 million accounted for 80 per cent of foreign trade in volume. Table A.1 shows the proportion of external traffic by principal ports for each region.

TABLE A.1: O_i and D_j, INTERNAL TRAFFIC FOR ROAD AND RAIL, THOUSAND TONS

	O_i	D_j	IMPORTS(%) O_i	EXPORTS(%) D_j
South East				
London and Medway (105,124,127,128,129, 130,133)	213870	215449	21.75	5.20
Dover (134)	7706	7702	19.93	9.34
Shoreham (126)	9099	9590	9.98	0.38
Southampton (118,119)	28401	27389	66.31	7.94
South West				
Teignmouth (111)	16508	16592	0.12	2.08
Plymouth (109)	8692	7848	2.60	3.84
Fowey and Par (108)	6933	6846	0.04	24.38
Bristol (113,115)	39598	37960	7.92	0.61
Wales				
Newport (88)	17392	19504	23.37	4.71
Cardiff, Port Talbot, Swansea (83,86)	42686	40212	15.71	6.25
Milford Haven (80)	4813	4641	558.36	154.66
Holyhead (56)	2126	1882	15.52	8.66
North West				
Liverpool and Manchester (41,42,44,46,59)	124507	127390	21.16	4.47
Preston (33,34,43)	41585	46683	0.89	0.62

The figures between brackets are the zone numbers.

Cont.../

TABLE A.1 (continued)

	O_i	D_j	IMPORTS(%) O_i	EXPORTS(%) D_j
Scotland				
Ardrossan and Clyde (4,7,10,11,12,15)	71470	78195	15.26	1.61
Aberdeen (3)	13162	13796	2.84	0.37
Dundee and Forth (6,8,9,14)	72386	67519	5.60	1.71
Northern				
Tyne (22,26)	34010	37486	6.02	1.08
Tees and Hartlepool (28)	25182	24014	60.56	11.34
Whitehaven, Workington and Barrow (24,29)	7829	919	13.85	0.73
Yorkshire and Humberside				
Hull, Grimsby, Immingham and Goole, (31,39,40,53,54,55)	70733	81851	22.50	5.57
East Midlands				
Boston (64)	11984	7415	4.88	1.77
East Anglia				
Kings Lynn and Great Yarmouth (77,78,79)	22824	24427	3.45	1.25
Felixstowe, Ipswich and Harwich (99,106)	16104	16268	22.17	12.77
	909.600	928.578	20.94	4.97

Source: D.E. Pitfield (1978)

For comparison with previous research and because data on pipeline and coastal shipping are not reliable, the aggregate mode road plus road is consider (see Chapter III, data).

Column 3 and 4 in Table A.1 show ratios of imports over originating traffic and of exports over terminating traffic respectively. The import ratios vary from 0.04 per cent in the case of Fowey and Par to 558.36 per cent for Milford Haven and exports vary from 0.37 in Aberdeen to 154.66 for Milford Haven. It seems that the leading ports activity represents a substantial share of inland traffic. Accordingly, special treatment of main port zones may be warranted.

3. Commodity Composition of Foreign Trade

Britain's exports are predominantly manufactured and semi-manufactured products whereas imports consist chiefly of raw materials. Table A.2 gives the commodity break-down of foreign trade for 1972.

This table shows the relative share of different commodity groups in foreign trade. The dominant commodity is obviously petroleum (38.33% for export and 62.07% for imports). This commodity can be identified to commodity 23 (fuel oil) of the British Rail commodity classification. The ratios of imports to O_i's and of exports to D_j's for this commodity are 129 per cent and 7.53 per cent respectively.

TABLE A.2

	Exports (000)tons	% of Total	Imports (000)tons	% of Total
Total	50,713	100.00	56,175	100.00
Foodstuffs	2,428	4.79	3,257	9.63
Basic materials	9,555	18.84	11,185	16.86
Manufactured goods	17,043	33.61	19,483	8.80
Coal etc.	2,251	4.44	3,444	2.64
Petroleum	19,436	38.33	18,806	62.07

Source: Abstract of Statistics 1976.

4. Derivation of the Port Variables(1)

To obtain measures of imports and exports, Heyman (Department of the Environment (1971)) carried out conversion of the original measures available in various units such as square yards of cloth, barrels of bear, etc. into tons. In this study, rather than get involved in such operations, it was thought more appropriate to experiment with dummy variables since port zones may be considered as a special category of zones. The investigation considers first all port zones and then only the main port zones.

For the port zones, the classification is straightforward but for the main port zones some cut-off point is to be selected. This is based on the following conditions:

---

(1) This is for the zonal analysis since for the regions data were readily available in tons.

Imports / O_i $>$ National average,

Exports / D_j $>$ National average, where

National average is total imports/exports for port zones as a percentage of total O_i/D_j by road plus rail for port zones.

This results in two different sets of dummy variables, one for imports and one for exports.

APPENDIX II.1

STANDARD INDUSTRIAL CLASSIFICATION, REVISED 1968 - ORDERS 1 to 24

Minimum List Heading	
	ORDER I -AGRICULTURAL, FORESTRY, FISHING
001	Agriculture and horticulture
002	Forestry
003	Fishing
	ORDER II -MINING AND QUARRYING
101	Coal mining
102	Stone and slate quarrying and mining
103	Chalk, clay, sand and gravel extraction
104	Petroleum and natural gas
109	Other mining and quarrying
	ORDER III-FOOD, DRINK AND TOBACCO
211	Grain milling
212	Bread and flour confectionery
213	Biscuits
214	Bacon curing, meat and fish products
215	Milk and milk products
216	Sugar
217	Cocoa, chocolate and sugar confectionery
218	Fruit and vegetable products
219	Animal and poultry fooods
221	Vegetable and animal oils and fats
229	Food industries not elsewhere specified
231	Brewing and malting
232	Soft drinks
239	Other drink industries
240	Tobacco

Contd..../

Minimum List Heading	
	ORDER IV -COAL AND PETROLEUM PRODUCTS
261	Coke ovens and manufactured fuel
262	Mineral oil refining
263	Lubricating oils and greases
	ORDER V -CHEMICALS AND ALLIED INDUSTRIES
271	General chemicals
272	Pharmaceutical chemicals and preparations
273	Toilet preparations
274	Paint
275	Soap and detergents
276	Synthetic resins and plastics materials and synthetic rubber
277	Dyestuffs and pigments
278	Fertilizers
279	Other chemical industries
	ORDER VI -METAL MANUFACTURE
311	Iron and steel (general)
312	Steel tubes
313	Iron castings, etc.
321	Aluminium and aluminium alloys
322	Copper, brass and other copper alloys
323	Other base metals
	ORDER VII -MECHANICAL ENGINEERING
331	Agricultural machinery (except tractors)
332	Metal-working machine tools
333	Pumps, valves and compressors
334	Industrial engines
335	Textile machinery and accessories
336	Construction and earth-moving equipment
337	Mechanical handling equipment
338	Office machinery

Contd.../

Minimum List Heading	
339	Other machinery
341	Industrial (including process) plant and steelwork
342	Ordnance and small arms
349	Other mechanical engineering not elsewhere specified
	ORDER VIII-INSTRUMENT ENGINEERING
351	Photographic and document copying equipment
352	Watches and clocks
353	Surgical instruments and appliances
354	Scientific and industrial instruments and systems
	ORDER IX -ELECTRICAL ENGINEERING
361	Electrical machinery
362	Insulated wires and cables
363	Telegraph and telephone apparatus and equipment
364	Radio and electronic components
365	Broadcast receiving and sound reproducing equipment
366	Electronic computers
367	Radio, radar and electronic capital goods
368	Electric appliances primarily for domestic use
369	Other electrical goods
	ORDER X -SHIPBUILDING AND MARINE ENGINEERING
370	Shipbuilding and marine engineering
	ORDER XI -VEHICLES
380	Wheeled tractor manufacturing
381	Motor vehicle manufacturing
382	Motor cycle, tricycle and pedal cycle manufacturing
383	Aerospace equipment manufacturing and repairing
384	Locomotives and railway track equipment
385	Railway carriages and wagons and trams

Contd.../

Minimum List Heading	
	ORDER XII -METAL GOODS NOT ELSEWHERE SPECIFIED
390	Engineers' small tools and gauges
391	Hand tools and implements
392	Cutlery, spoons, forks and plated tableware, etc.
393	Bolts, nuts, screws, rivets, etc.
394	Wire and wire manufactures
395	Cans and metal boxes
396	Jewellery and precious metals
399	Metal industries not elsewhere specified
	ORDER XIII-TEXTILES
411	Production of man-made fibres
412	Spinning and doubling on the cotton and flax systems
413	Weaving of cotton, linen and man-made fibres
414	Woollen and worsted
415	Jute
416	Rope, twine and net
417	Hosiery and other knitted goods
418	Lace
419	Carpets
421	Narrow fabrics (not more than 30 cm. wide)
422	Made-up textiles
423	Textile finishing
429	Other textile industries
	ORDER XIV -LEATHER, LEATHER GOODS AND FUR
431	Leather (tanning and dressing) and fellmongery
432	Leather goods
433	Fur
	ORDER XV -CLOTHING AND FOOTWEAR
441	Weatherproof outerwear
442	Men's and Boys' tailored outerwear

Contd.../

Minimum List Heading	
443	Women's and girls' tailored outerwear
444	Overalls and men's shirts, underwear, etc.
445	Dresses, lingerie, infants' wear, etc.
446	Hats, caps and millinery
449	Dress industries not elsewhere specified
450	Footwear
	ORDER XVI -BRICKS, POTTERY, GLASS, CEMENT, ETC.
461	Bricks, fireclay and refractory goods
462	Pottery
463	Glass
464	Cement
469	Abrasives and building materials, etc. not elsewhere specified
	ORDER XVII-TIMBER, FURNITURE, ETC.
471	Timber
472	Furniture and upholstery
473	Bedding, etc.
474	Shop and office fitting
475	Wooden containers and baskets
479	Miscellaneous wood and cork manufactures
	ORDER XVIII-PAPER, PRINTING AND PUBLISHING
481	Paper and board
482	Packaging products of paper, board and associated materials
483	Manufactured stationery
484	Manufactures of paper and board not elsewhere specified
485	Printing, publishing of newspaper
486	Printing, publishing of periodicals
489	Other printing, publishing, bookbinding, engraving, etc.

Contd.../

Minimum
List
Heading

ORDER XIX -OTHER MANUFACTURING INDUSTRIES

491 Rubber

492 Linoleum, plastics floor-covering, leathercloth, etc.

493 Brushes and brooms

494 Toys, games, children's carriages, and sports equipment

495 Miscellaneous stationers' goods

496 Plastics products not elsewhere specified

499 Miscellaneous manufacturing industries

ORDER XX -CONSTRUCTION

500 Construction

ORDER XXI -GAS, ELECTRICITY AND WATER

601 Gas

602 Electricity

603 Water supply

ORDER XXII -TRANSPORT AND COMMUNICATION

701 Railways

702 Road passenger transport

703 Road haulage contracting for general hire or reward

704 Other road haulage

705 Sea transport

706 Port and inland water transport

707 Air transport

708 Postal services and telecommunications

709 Miscellaneous transport services and storage

ORDER XXIII-DISTRIBUTIVE TRADES

810 Wholesale distribution of food and drink

811 Wholesale distribution of petroleum products

Contd.../

Minimum List Heading	
812	Other wholesale distribution
820	Retail distribution of food and drink
821	Other retail distribution
831	Dealing in coal, oil, builders' materials, grain and agricultural supplies
832	Dealing in other industrial materials and machinery

ORDER XXIV -INSURANCE, BANKING, FINANCE AND BUSINESS SERVICES

860	Insurance
861	Banking and bill discounting
862	Other financial institutions
863	Property owning and managing, etc.
864	Advertising and market research
865	Other business services
866	Central offices not allocable elsewhere

APPENDIX II.2

British Rail Commodity Classification

Data-base classification	CSTE classification	Item number	Division number	Category number
1 Coal and coke products	coal (anthracite, bituminous)	177	27	11
	briquettes of coal	179	27	11
	lignite	181	28	11
	lignite briquettes	183	28	11
	peat	185	28	11
	coke and semicoke of coal	187	29	11
	coke and semicoke of lignite or peat	189	29	11
2 Steel	crude steel	275	45	16
	blooms, billets, slabs, sheet bars, and coils for rerolling	277	45	16
	other semifinished iron and steel products	279	45	16
	wire rod of iron or steel	281	46	16
	rounds and squares, bars, sheet pilings and angles, shapes and sections of iron or steel	283	46	16
	universals, plates, and sheets of iron or steel	285	46	16
	hoop and strip of iron or steel	287	46	16
	rails and railway-track construction material of iron or steel	289	46	16
	iron and steel wire (excluding wire rod)	291	46	16
3 Alloys, nonferrous metals	copper and copper alloys, unwrought	297	47	16
	aluminium and aluminium alloys, unwrought	299	47	16
	lead and lead alloys, unwrought	301	47	16
	zinc and zinc alloys, unwrought	303	47	16
	other nonferrous metals and their alloys (including silver and platinum), unwrought	305	47	16
	bars, rods, plates, sheets, wire, pipes, tubes, castings, and forgings of nonferrous metals	307	47	16
4 Ores	iron ore and concentrates (except roasted iron pyrites)	157	23	8
	roasted iron pyrites	159	23	8
	slag for resmelting, flue dust	161	23	8
	ore and concentrates of copper	165	25	9
	bauxite and concentrates of aluminium	167	25	9
	ore and concentrates of manganese	169	25	9
	ore and concentrates of other nonferrous base metals	171	25	9
5 Lime and limestone for steelmaking	none			
6 Pig and iron castings	pig iron, spiegeleisen, sponge iron, iron and steel powders and shot, and ferroalloys	273	45	16
	tubes, pipes, and fittings of iron and steel	293	46	16
	iron and steel castings and forgings, unworked, NES	295	46	16

APPENDIX II.2 (Continued)

Data-base classification	CSTE classification	Item number	Division number	Category number
7 Scrap	iron and steel scrap	163	24	8
	nonferrous-metal scrap	173	25	9
8 Fly ash, waste, flasks	slag (not for resmelting), ash, dross	151	22	7
9 Grain and cereals	wheat (including spelt) and meslin, unmilled	17	5	1
	rice	19	5	1
	barley, unmilled	21	5	1
	maize, unmilled	23	5	1
	rye, unmilled	25	5	1
	oats, unmilled	27	5	1
	cereals, unmilled, NES	29	5	1
	meal, flour, and groats of cereals	31	6	3
	malt	33	6	3
	other cereal preparations and preparations of flour; starch of fruit and vegetables	35	6	3
10 Potatoes, fruit, and vegetables	citrus fruit	37	7	2
	other fresh fruit and nuts	39	7	2
	potatoes	41	7	2
	other vegetables, fresh or frozen	43	7	2
	dried or dehydrated fruit, preserved fruit, frozen fruit	47	8	3
	dried leguminous vegetables	49	8	3
	hops	51	8	3
	preserved vegetables and vegetable preparations	53	8	3
11 Animal feed, sugar beet	sugar beet	45	8	3
	cereal hay, straw, and husks	63	10	3
	oilseed cake and other residues resulting from the extraction of vegetable oils	65	10	3
	bran, sharps, food waste, and prepared animal feed	67	10	3
12 Meat and fish	live animals	1	1	3
	fresh, chilled, and frozen meat	3	2	3
	dried, salted, smoked, and preserved meat	5	2	3
	fish, crustacea, and molluscs (fresh, chilled, frozen, salted, smoked, or dried)	13	4	3
	preserved fish, crustacea, and molluscs, canned or not canned	15	4	3
	crude animal and vegetable materials, NES	175	26	10
	animal or vegetable oils and fats, and derivatives	209	33	4

APPENDIX II.2 (Continued)

Data-base classification	CSTE classification	Item number	Division number	Category number
13 Other foods and drinks and milk products	milk and cream (evaporated, condensed, or powdered); butter and cheese	9	3	3
	eggs	11	3	3
	raw sugar	55	9	3
	refined sugar	57	9	3
	molasses	59	9	3
	sugars and syrups, NES; natural and artificial honey; sugar confectionary and other sugar preparations (except chocolate confectionary)	61	9	3
	coffee	69	11	3
	cocoa and chocolate	71	11	3
	tea and maté, spices	73	11	3
	margarine and cooking fats	75	11	3
	food preparations, NES	77	11	3
	nonalcoholic beverages, NES	79	12	3
	wine of fresh grapes (including grape must)	81	12	3
	beer	83	12	3
	other alcoholic beverages	85	12	3
	tobacco, unmanufactured	87	13	3
	tobacco, manufactured	89	13	3
14 Industrial sand/clays	sand for industrial use	129	21	7
	clay and clay earth	137	21	7
15 Other earths and stones	other sand or gravel	131	21	7
	crushed stones, macadam, tarred macadam, pebbles	133	21	7
	pumice stone (including pumiceous sand and gravel)	135	21	7
	building and monumental (dimension) stone	139	22	7
	plasters	141	22	7
	limestone for industry	143	22	7
	sulphur, other than sublimed, precipitated, or colloidal sulphur	145	22	7
	iron pyrites, unroasted	147	22	7
	salt, crude or refined	149	22	7
	chalk	153	22	7
	other crude minerals	155	22	7
16 Chemicals	calcium carbide	217	34	14
	aluminium oxide and hydroxide	219	34	14
	other basic chemicals	221	34	14
	benzole	223	35	13
	pitch, mineral tar, and other crude chemicals from coal or natural gas	225	35	13
	dyeing, tanning, and colouring materials	239	38	14
	medicinal and pharmaceutical products; perfumes and polishing preparations	241	38	14
	explosives and pyrotechnic products; hunting ammunition	243	38	14
	starch, starchy substances, gluten	245	38	14
	chemical materials and products, NES	247	38	14
17 Plastics	plastic materials, regenerated cellulose, and artificial resins	237	37	14

APPENDIX II.2 (Continued)

Data-base classification	CSTE classification	Item number	Division number	Category number
18 Fertilizers	natural sodium nitrate	121	20	6
	natural phosphates, whether or not ground	123	20	6
	natural potassic salts, crude	125	20	6
	natural fertilizers, of animal or vegetable origin, not chemically treated	127	20	6
	basic slag (Thomas slag)	227	36	6
	other phosphatic fertilizers and fertilizer materials	229	36	6
	potassic fertilizers and potassic fertilizer materials	231	36	6
	nitrogenous fertilizers and nitrogenous fertilizer materials	233	36	6
	fertilizers, NES	235	36	6
19 Gasses, acids bulk chemicals	gas, natural and manufactured	207	32	12
	sulphuric acid, oleum	211	34	14
	caustic soda (sodium hydroxide)	213	34	14
	natural sodium carbonate (soda ash)	215	34	14
20 Cement	cement	263	43	15
21 Bricks, other building materials	lime, including lime for fertilizing	261	43	15
	pumiceous agglomerates, pieces of concrete or cement, asbestos cement and fibro-cement, and other fabricated building materials (except manufactures of clay or glass)	265	44	15
	clay construction materials and refractory construction materials	267	44	15
	glass	269	44	15
	glassware, pottery, and other mineral manufactures	271	44	15
22 Timbers and forest products	oilseeds, oil nuts, and oil kernels	93	15	4
	crude rubber, including synthetic and reclaimed	95	16	10
	pulpwood (including broad-leaved)	97	17	5
	pit props (mine timber)	99	17	5
	other wood in the log	101	17	5
	railway sleepers and other shaped or simply worked wood	103	17	5
	fuel wood and charcoal; wood waste; cork, raw and waste	105	17	5
	wood pulp	107	18	10
	veneer sheets, plywood, wood panels, artificial or reconstituted wood, and wood and cork manufactures (excluding furniture)	253	40	19
23 Fuel oil and petroleum products		191	30	12
	petroleum, crude or partly refined	193	30	12
	motor spirit, other light oils			
	lamp oil and white spirit (kerosene,	195	30	12
	illuminating oil, jet fuel)	197	30	12
	distillate fuels	199	30	12
	residual fuel oils	201	31	12
	lubricating oils and greases	203	31	12
	bitumen and bituminous mixtures	205	31	12
	other nonenergy petroleum by-products			

APPENDIX II.2 (Continued)

Data-base classification	CSTE classification	Item number	Division number	Category number
24 Vehicles and CKD	tractors and agricultural machinery and equipment, whether or not assembled; parts thereof	313	49	18
	transport equipment, whether or not assembled; parts thereof	319	50	18
25 Furniture, textiles, and shoes	wool and other animal hair	111	19	10
	cotton	113	19	10
	silk, jute, flax, sisal, hemp, and other vegetable fibres, NES	115	19	10
	synthetic and regenerated (artificial) fibres	117	19	10
	waste materials from textile fabrics, including rags	119	19	10
	leather, leather manufactures, NES, and dressed fur skins	249	39	19
	rubber manufactures, NES	251	39	19
	textile yarn, fabrics, made-up articles, and related products	259	42	19
	furniture (new)	321	51	19
	travel goods, handbags and similar articles, clothing, knitted and crocheted goods, footwear	323	51	19
26 Electrical equipment	electrical machinery, apparatus, appliances, and parts thereof	317	49	18
27 Miscellaneous, NES	hides, skins, and fur skins, undressed	91	14	10
	manufactured articles, NES	327	51	19
	used packaging	329	52	20
	used building equipment; circus vehicles and equipment	331	52	20
	furniture-removal equipment	333	52	20
	gold, coins, medals	335	52	20
	firearms ammunition for military purposes	337	52	20
	goods which cannot be classified by their nature	339	52	20
28 Engineering products and machinery	finished structural parts and structures	309	48	17
	other manufactures of metal	311	48	17
	other nonelectrical machinery and parts thereof	315	49	18
29 Other parcels and newspapers	paper waste and old paper	109	18	10
	paper and paperboard	255	41	19
	articles made of paper pulp, of paper, or of paperboard	257	41	19
	printed matter	325	51	19
30 Milk (fresh)	Milk and cream (fresh)	7	3	3

Key: NES—not elsewhere specified or included.

APPENDIX II.3

The actual difficulty in approximately matching the freight commodity classification with the industrial breakdown of the employment data varied from case to case. The matching was based on a detailed list of M.L.H. subdivisions on the one hand and the British Rail commodity groups as C.S.T.E. and S.I.C.T. categories (Standard International Trade Classification - Commission of H.M. Customs and Excise, 1972) on the other hand.

At the regional level, sometimes there was a need to aggregate British Rail commodity groups too, as to correspond to employment by S.I.C. order when suitable data at M.L.H. level were not available for all regions. Another constraint for the regional nalysis was the availability of net output measure at the order level and for the manufacturing industries only. (See Table A.3).

T A B L E A.3

REGIONS

COMMODITY GROUPS		EMPLOYMENT	NET-OUTPUT
(1)	Coal & Coke Products	M.L.H. 101, 261	Order IV
(2)	Steel	M.L.H. 311, 312	
(3)	Alloys	M.L.H. 321, 322	Order VI
(4,6,7)	Iron, Ores, Scrap	M.L.H. 313, 323	
(9-13)	Food, Drinks, Tobacco	S.I.C. Order III	Order III
(14,15)	Sand, Clays, Earths	M.L.H. 102, 103	Order XVI
(16)	Chemicals	M.L.H. 271-274, 277, 279	
(17)	Plastics	M.L.H. 276, 492, 496	Order V
(18)	Fertilizers	M.L.H. 278	
(20,21)	Cement & Building Mat.	S.I.C. Order XVI	Order XVI
(24)	Vehicles & C.K.D.	S.I.C. Order XI	Order XI
(25)	Furniture, Textiles, Shoes	S.I.C. Order XIII, XIV, XV & M.L.H. 472, 474	Order XIII, XIV, XV, XVII
(26) *	Electrical Equipment	S.I.C. Order IX	Order IX

ZONES

Commodity Groups 1,2,3,16,17,18,24,25,26 as for Regions.
Remaining Commodity Groups:

(6) Iron	M.L.H. 313	(13) Other Foods & Drinks	M.L.H. 215-217, 229,231,232, 239,240
(9) Grain & Cereals	M.L.H. 001,211,212,213	(20) Cement	M.L.H. 464
(10) Potatoe, Fruit & Veg.	M.L.H. 218	(21) Bricks & Other Building Mat.	M.L.H. 461,462,
(12) Meat & Fish	M.L.H. 003,221,214	(28) Engineering Products	M.L.H. 332-339.

* Commodity 23 (oil & petroleum) was not considered for regressions on employment by industry, as data were missing for some regions

APPENDIX II.4

SOURCE OF REGIONAL ECONOMIC DATA

Total Employment and Employment by Industry[5].

Agriculture Area[1].

Gross Regional Domestic Product[1].

Coal Consumption All Classes[1].

Region Area[2].

Cement Deliveries/Regions[3].

Number of Goods Vehicles[1].

Population[2].

Exports[1].

Imports[1].

Petrol Consumption[1].

Retail Turnover[1].

Net Output Manufacturing Industry[4].

Net Capital Expenditure[4].

Gross Domestic Profits[1].

Income from Employment[1].

Public Investment[1].

Agriculture Investment[1].

Production Iron[1].

Production Steel[1].

1. C.S.O. (1974) Abstract of Regional Statistics, (H.M.S.O., London)

2. Aggregated from Zonal Data.

3. Institute of Geological Studies (1974), U.K. Mineral Statistics, (H.M.S.O., London).

4. Business Statistical Office (1977), 'Report on the Census of Production of 1972", Business Monitor, P.A. 1002, (H.M.S.O., London).

5. Department of Employment (1973), "Annual Census of Employment", Department of Employment Gazette, August 1973, (H.M.S.O., London).

APPENDIX III.1

Zone centroids for the 134 zones

Zone number	Centroid	Zone number	Centroid	Zone number	Centroid
1	Bonar Bridge [a]	46	Stockport	91	Gloucester
2	Daviot	47	Huddersfield	92	Banbury
3	Aberdeen	48	Buxton	93	Oxford
4	Inveraray	49	Wakefield	94	Milton Keynes
5	Perth	50	Sheffield	95	Luton
6	Dundee	51	Doncaster	96	Bedford
7	Dumbarton	52	Worksop	97	Cambridge
8	Falkirk	53	Scunthorpe	98	Bury St Edmunds
9	Kirkcaldy	54	Lincoln	99	Ipswich
10	Paisley	55	Grimsby	100	High Wycombe
11	Glasgow	56	Holyhead [b]	101	St Albans
12	Mauchline	57	Colwyn Bay	102	Stevenage
13	Wishaw	58	Chester	103	Harlow
14	Edinburgh	59	Winsford	104	Colchester
15	Stranraer	60	Stoke	105	Chelmsford
16	Lockerbie	61	Derby	106	Harwich
17	Jedburgh	62	Nottingham	107	Truro
18	Coldstream	63	Grantham	108	Bodmin
19	Penrith	64	Boston	109	Plymouth
20	Morpeth	65	Newtown	110	Barnstaple
21	Hexham	66	Shrewsbury	111	Exeter
22	Newcastle	67	Telford	112	Taunton
23	Sunderland	68	Stafford	113	Burnham
24	Whitehaven	69	Wolverhampton	114	Yeovil
25	Kendal	70	Lichfield	115	Bristol
26	Bishop Auckland	71	Birmingham	116	Calne
27	Darlington	72	Coventry	117	Warminster
28	Teesside	73	Leicester	118	Bournemouth
29	Barrow	74	Northampton	119	Eastleigh
30	Thirsk	75	Kettering	120	Reading
31	Scarborough	76	Peterborough	121	Portsmouth
32	Lancaster	77	King's Lynn	122	Guildford
33	Blackpool	78	Norwich	123	Chichester
34	Blackburn	79	Yarmouth	124	NW London
35	Colne	80	Pembroke	125	Reigate
36	Halifax	81	Aberystwyth	126	Brighton
37	Leeds	82	Carmarthen	127	NE London
38	Harrogate	83	Swansea	128	SE London
39	York	84	Brecon	129	Basildon
40	Hull	85	Merthyr Tydfil	130	Sittingbourne
41	Liverpool	86	Cardiff	131	Tonbridge
42	Birkenhead	87	Pontypool	132	Eastbourne
43	Wigan	88	Newport	133	Ashford
44	Warrington	89	Hereford	134	Canterbury
45	Bury	90	Worcester		

[a] The true centroid for zone 1 is about six miles due north of Bonar Bridge, and mileages refer to this point.

[b] The true centroid for zone 56 is about ten miles east of Holyhead, and mileages refer to this point.

APPENDIX III.2

TABLE A.4

CONVERSION OF B.R. TRAFFIC ZONES TO THE STANDARD REGIONS

REGIONS	B.R. TRAFFIC ZONES
SCOTLAND	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17 and part of 18 (about 50% by visual inspection of area). N.B. Traffic Zones exclude or kney, Shetland, West Isles and other Islands.
NORTH	19,20,21,22,23,24,25,26,28,29 and part of 18 (c. 50%) " " " 27 (c. 50%) " " " 30 (c. 70%) " " " 31 (c. 85%)
YORKSHIRE AND HUMBERSIDE	37,38,39,40,47,49,50,51,53,55, and part of 27 (c. 50%) " " " 30 (c. 30%) " " " 35 (c. 45%) " " " 36 (c. 98%) " " " 54 (c. 70%) " " " 64 (c. 35%) " " " 31 (c. 15%)
NORTH WEST	32,33,34,43,41,42,45,46,59 and part of 36 (c. 2%) " " " 35 (c. 55%) " " " 60 (c. 30%) " " " 58 (c. 30%)
WALES	56,57,65,80,81,82,83,84,85,86,87,88 and part of 58 (c. 70%)

Cont..../

TABLE A.4 (Continued)

REGIONS	B.R. TRAFFIC ZONES
WEST MIDLANDS	66,67,68,69,70,71,72,89,90 and part of 60 (c. 70%) " " " 92 (c. 30%)
EAST MIDLANDS	48,52,75,73,74,62,61,63 and part of 54 (c. 30%) " " " 64 (c. 65%) " " " 92 (c. 10%)
EAST ANGLIA	76,77,78,79,97,98,99 and part of 104 (c. 40%)
SOUTH EAST	93,94,95,96,100,101,102,103,105,106,119, 120,121,122,123,124,125,126,127,128,129, 130,131,132,133,134 and part of 92 (c. 60%) " " " 104 (c. 60%) " " " 117 (c. 20%) " " " 118 (c. 15%)
SOUTH WEST	91,107,108,109,110,111,112,113,114,115, 116 and part of 117 (c. 80%) " " " 118 (c. 85%)

Regional Definitions from:
C.S.O. (1974) Regional Statistics

A P P E N D I X III.3

CONVERSION CHART OF B.R. ZONES TO ADMINISTRATIVE AREAS

B.R.TRAFFIC ZONES Number	ADMINISTRATIVE AREAS
01	X03, X05, X06, X07
02	X00, X04, X08, X14, X16
03	X12, X13, X17, X18, X19, X1X, X20
04	X02, X09
05	X0X, X0Y, X01
06	X10, X11, X15, X1Y, X30, X31
07	XXX, XXY, X5Y
08	XX0, XX1, XX2, XX3, XX4, XX5, XX6 XX7, XX8, XX9, XY2, XY3, X66, X67, X68, X69
09	X32, X33, XY0, XY1, XY4, XY5, XY6, XY7, XY8, XY9, XYX, XYY
10	X51, X56, X57, X58, X59, X5X, X9Y
11	X50
12	X90, X91, X92, X93, X94, X95, X96, X97, X98, X99, X9X
13	X52, X53, X54, X55, X60, X61, X62, X63, X64, X65
14	X40, X41, X42, X43, X44, X45, X46, X47, X48, X49, X82, X83, X84, X85, X89, X8X, X8Y
15	X71, X72
16	X70, X73, X74, X75, X76, X77, X79, X7X
17	X78, X80, X81, X86, X87
18	12Y, X88, 121
19	100, 103, 107, 108, 109
20	120, 124, 125, 126, 127, 129
21	111, 116, 119, 122, 123, 12X, 140
22	110, 112, 113, 117, 118, 11Y, 128, 115

Cont../

B.R.TRAFFIC ZONES Number	ADMINISTRATIVE AREAS
23	114, 11X, 146, 147, 148
24	101, 102, 104, 105, 106, 10X
25	160, 161, 162, 163, 164, 165
26	141, 143, 149, 14X, 14Y
27	142, 144, 145, 154, 15Y
28	130, 132
29	051, 056, 05Y, 10Y
30	151, 152, 155, 157, 158
31	150, 153, 159, 15X, 200, 201, 203, 207, 208
32	052, 053, 057
33	050, 054, 055, 058, 059, 05X
34	030, 033, 034, 035, 036, 037, 038, 040, 042, 045, 046, 047, 049, 04X
35	041, 043, 044, 048, 04Y, 224, 228, 22X
36	229, 238, 251, 252, 255, 25Y
37	250, 258, 25X
38	221, 227, 22Y
39	156, 204, 220, 223, 226
40	202, 205, 206, 209, 20X, 20Y, 210, 211, 222, 225
41	022, 070, 073, 075, 078, 079
42	071, 072, 07X, 07Y
43	020, 023, 204, 025, 026, 027, 028, 029, 02Y, 031, 032, 074
44	005, 021, 02X, 076, 077
45	061, 063, 064, 065, 066, 067, 069
46	060, 062, 068, 06Y, 006, 06X
47	237, 239, 253, 256, 257
48	010, 011, 309
49	230, 235, 236, 23X, 254, 259
50	232, 240, 241, 242, 243, 244, 245
51	231, 233, 234
52	301, 307, 308, 350, 351, 353, 354, 355, 356, 357, 359

Cont../

B.R.TRAFFIC ZONES Number	ADMINISTRATIVE AREAS
53	263, 267
54	268, 269, 330, 33X
55	260, 261, 262, 26X
56	800 to 807
57	840 to 849, 84X, 84Y, 852, 853, 854, 856, 857, 85X, 85Y, 897
58	000, 00Y, 850, 851, 855, 858, 852
59	001, 004, 009, 00X
60	925, 926, 92Y, 002, 003, 007, 008, 930 to 933
61	300, 302, 304, 305, 306, 30X
62	352, 358, 35X, 360, 361, 362, 363, 364, 365, 366, 367
63	332, 333, 335, 336, 339, 33Y
64	264, 265, 266, 331, 334, 337, 338
65	890, 891, 892, 893, 894, 895, 896, 898, 8X0, 8X1, 8X2, 8X3, 8X4, 8X5, 8X9
66	910, 912, 914, 916, 918, 91Y
67	911, 913, 919, 91X
68	921, 927, 928, 92X
69	929, 951, 954, 95Y, 961, 962, 963, 964, 967, 968, 969, 96X
70	920, 922, 923, 924, 30Y
71	955, 958, 95X, 960, 965, 966
72	940 to 945, 947, 948, 949, 94X, 94Y, 970, 971, 972, 973
73	310 to 319, 31X, 31Y, 320 to 324, 370 to 373
74	340, 342, 343, 34X, 34Y
75	344 to 349
76	424, 426, 427, 429, 42X, 42Y, 470 to 479, 47X, 47Y
77	481, 487, 489
78	482 to 486, 488, 48X, 4Y0

Cont../

B.R.TRAFFIC ZONES Number	ADMINISTRATIVE AREAS
79	480, 491, 492, 494, 498
80	8Y0 to 8Y9, 8YX
81	824 to 827
82	820 to 823, 874
83	865, 869, 86Y, 870 to 873
84	810 to 819, 81X, 8X6, 8X7, 8X8, 8XX, 8XY
85	860, 861, 862, 867, 868
86	830, 832, 835, 836, 863, 864, 866, 86X
87	838, 882, 884, 885, 886, 887, 888, 88Y
88	831, 833, 834, 837, 880, 881, 883, 889, 88X
89	900 to 909, 90X, 90Y
90	950, 952, 953, 956, 957, 959
91	740 to 749, 74X
92	341, 681, 682, 686, 687, 946
93	611, 614, 615, 161, 617, 618, 619, 680, 683, 684, 685, 688, 689
94	5X2, 5X4, 5X5, 5X6, 630 to 635
95	601, 605, 58Y
96	600, 602, 603, 604, 606, 607
97	420 to 423, 425, 428, 4X3
98	48Y, 4X0, 4X4, 4X8, 4X9, 4XX
99	490, 493, 495, 496, 497, 49 49X, 49Y
100	5X0, 5X1, 5X3, 5X7, 5X8, 5X9, 5XX
101	586, 587, 588, 58X
102	580 to 584, 589
103	577, 579, 5YX, 5YY, 585, 628, 62X, 62Y
104	4X1, 4X2, 4X5, 4X6, 4X7, 620, 624, 626, 627
105	571, 57X, 57Y, 625, 629

Cont../

B.R.TRAFFIC ZONES Number	ADMINISTRATIVE AREAS
106	621, 622, 623
107	700, 701, 702, 705, 706, 707, 708
108	703, 704, 709, 70Y
109	718, 724, 730, 731, 732, 733
110	721, 722, 728, 729, 72X, 72Y
111	710 to 717, 719, 71X, 71Y, 720, 723, 725, 726, 727
112	761, 769, 76X
113	762, 764, 768
114	763, 765, 767, 770, 771, 772, 773, 776, 778, 77X, 77Y
115	750 to 759, 760, 766, 787, 78
116	781 to 785, 789
117	76Y, 780, 786, 788, 78X, 669
118	640, 660, 662, 664, 774, 775, 777, 779
119	661, 665, 666, 6X0, 6X5, 6X6
120	5XY, 5Y0 to 5Y9, 610, 612, 613, 663, 66X
121	667, 668, 66Y, 690 to 695, 6X1 to 6X4
122	550, 552, 553, 554, 557, 558, 559
123	6Y1 to 6Y4, 6Y9, 6YX, 6YY
124	501, 502, 503, 505, 520 to 525, 530, 556
125	531 to 534, 551, 555, 55X, 55Y, 590, 592
126	6Y0, 6Y5, 6Y6, 6Y7, 6Y8, 591, 594, 595, 596, 597, 650, 654, 656
127	500, 508, 509, 50Y, 510, 511, 512, 513, 514, 526, 52X, 573, 578
128	506, 507, 50X, 540, 541, 562, 567, 568
129	570, 572, 574, 575, 576
130	560, 563, 569, 679, 67X, 67Y
131	561, 564, 565, 566, 56X, 56Y

Cont../

B.R.TRAFFIC ZONES Number	ADMINISTRATIVE AREAS
132	593, 651, 652, 653, 655, 657, 658, 659, 65X, 65Y
133	672, 673, 674, 678
134	670, 671, 675, 676, 677

+++++++++++++++++

The codes correspond to local authorities as defined in:

Ministry of Transport, Direction of Statistics (August 1968) "Geographical System for Great Britain." Note: S.3/A9/012 (mimeographed).

APPENDIX IV

ANNUAL TONS AND AVERAGE LENGTH OF HAUL 1972

BY COMMODITY GROUP AND BY MODES

Commodity Group	ROAD Tons Millions	ROAD Average Haul, Miles	RAIL Tons Millions	RAIL Average Haul, Miles
Coal	110.4	16.8	108.9	48.4
Steel	42.8	52.2	10.7	88.7
Alloys	6.9	54.3	0.8	142.5
Ores	3.3	13.9	15.2	39.9
Lime	0.0	-	2.2	98.2
Pig, iron	22.5	55.9	1.1	100.0
Scrap	15.5	32.9	3.8	71.3
Waste	22.9	12.1	2.7	78.5
Grain	36.0	37.5	0.6	275.0
Fertilizer &veg.	51.3	32.1	0.6	210.0
Animal feed	40.2	35.2	0.3	220.0
Meat	44.9	36.4	0.1	180.0
Food & drink	100.0	34.7	0.6	233.0
Sand	15.0	18.9	2.2	144.0
Earth/stones	331.4	14.7	12.3	79.2
Chemicals	39.5	54.5	0.7	171.0
Plastics	4.2	70.9	0.03	166.0
Fertilizer	19.6	34.4	1.0	207.0
Bulk chemicals	4.3	49.7	2.2	154.0
Cement	33.4	28.6	3.5	101.0
Building Materials	260.5	21.7	1.3	98.5
Timber	40.3	35.4	0.4	95.0
Oil	78.9	28.1	20.2	94.2
Vehicles	16.1	50.8	0.9	245.0
Furniture	55.2	40.0	0.05	240.0
Electrical Equip.	14.3	46.0	0.1	220.0
Miscellaneous	142.3	29.8	1.0	148.0
Engineering/Mach.	37.7	44.0	0.05	180.0
Parcels/Newspapers	41.4	46.9	1.1	214.0
Milk	41.3	21.6	0.3	210.0