Monograph:
Clark, S. (1997) National Multi-Modal Travel Forecasts: Aggregate Elasticities from Published Great Britain Transport Statistics. Working Paper. Institute of Transport Studies, University of Leeds , Leeds, UK.

Working Paper 510

Institute for Transport Studies
University of Leeds

EPSRC Research Grant

NATIONAL MULTI-MODAL TRAVEL FORECASTS

AGGREGATE ELASTICITIES FROM PUBLISHED GREAT BRITAIN TRANSPORT STATISTICS

Document Control Information

Title	:	Aggregate elasticities from published Great Britain transport statistics
Authors	:	Stephen Clark.
Reference Number	:	9707/01
Version	:	0.2
Date	:	30 July 1997
Distribution	:	JMP, CAN, MRW
Availability	:	Restricted
File	:	E:\HOME\NMTF\WP.WP6
Authorised by	:	Stephen Clark
Signature	:	

Abstract

This study attempts to form a relationship between the observed volume (or demand) of each mode and various other explanatory measures such as fare, supply and socio-economic measures. The strength of these relationships will be expressed as elasticities which capture the percentage effect of the change in the explanatory measures on the observed volume, based on previous experience.

Reasonable models were obtained for bus and private car travel. The rail model had unusually high elasticities with respect to train kilometres. This was thought to be indicative of simultaneity. Attempts to remedy this situation by the applcation of two stage least squares were unsuccessful

1 INTRODUCTION

The United Kingdom Government collects a wealth of statistics on the operation of the UK economy, industry and society. One component of this exercise is a body of statistics on the operation of the transport industry. Information is collected on both the passenger and freight classes of transport. Within each class there is a further division by mode of transport. For the passenger sectors, which is the only one considered in this study, there are three main modes : private vehicle; bus and rail.

Transport may be regarded as a commodity which individuals consume. They are able to decide in what form this consumption should take, usually by the mode of transport adopted to travel. This may be by the so-called slow modes of walking or cycling (although today, in many cities, cycling has a shorter journey time than other modes during the peak hours) or the motorised modes of car, van, bus or rail travel. This study will examine the relationships in the motorised modes of transport. The primary source for statistics used in this study are those published by the Department of the Environment, Transport and the Regions (DoETR) in Transport Statistics : Great Britain (1996).

This study extends the work conducted by three Masters students at the University of Leeds (Aleem, 1995, Palomo, 1996 and Batley, 1996). The extensions range from the addition of extra years of published data through to a re-structuring of the equations and variables in the model.

The work will be presented in five sections. The first section will discuss the structural form of the relationship between the dependant demand and the independent explanatory variables. The relative merits and de-merits of each form will be discussed. The next three sections will take each mode in turn (Bus, Rail and Private) and present the main findings. The final section will take the rail work and attempt a two-stage least squares re-working of the problem.

2 MODELLING

The formulation of the model relating the demand to the exogenous variables has been split into two components, one of which relates to the structure of the equation and the other to necessary measures to account for statistical features in the observed data.

2.1 STRUCTURAL

For the bus and rail models the data is available at both a national and a sub-national level. In the case of the bus data the sub-levels are the regions of Great Britain whilst with the rail data the sub-levels are the sectors of the national railway. No sub-levels are used in the private travel market. Within the Aleem and Palomo dissertations there are two basic formulations for the demand equations to account for the sub-levels in the bus and rail data. The first is the introduction of sufficient dummy variables to account for all the variation between the sub-levels, this has been termed the **dummy variable** approach. The second is to form a variable which has the correct measure for the appropriate observations and a zero elsewhere, a **sectorised** approach. A third formulation, not used in any of the dissertations, is to have **separate** models for each sub-level.

Algebraically each of these approaches can be expressed as:

Dummy variable

$$DEMAND_{*,t} = f\,(FARE_{*,t},\ SERVICE_{*,t},\ X_{*,t},\ TIME,\ D_A,\ D_B) + \varepsilon_{*,t}$$

Sectorised

$$\begin{bmatrix} DEMAND_{A,t} \\ DEMAND_{B,t} \\ DEMAND_{C,t} \end{bmatrix} = f \begin{bmatrix} FARE_{A,t} & SERVICE_{A,t} & X_{A,t} & 0 & 0 & 0 & 0 & 0 & 0 & TIME \\ 0 & 0 & 0 & FARE_{B,t} & SERVICE_{B,t} & X_{B,t} & 0 & 0 & 0 & TIME \\ 0 & 0 & 0 & 0 & 0 & 0 & FARE_{C} & SERVICE_{C,t} & X_{C,t} & TIME \end{bmatrix} + \begin{bmatrix} \varepsilon_{A,t} \\ \varepsilon_{B,t} \\ \varepsilon_{C,t} \end{bmatrix}$$

Separate

$$DEMAND_{A,t} = f\,(FARE_{A,t},\ SERVICE_{A,t},\ X_{A,t},\ TIME) + \varepsilon_{A,t}$$
$$DEMAND_{B,t} = f\,(FARE_{B,t},\ SERVICE_{B,t},\ X_{B,t},\ TIME) + \varepsilon_{B,t}$$
$$DEMAND_{C,t} = f\,(FARE_{C,t},\ SERVICE_{C,t},\ X_{C,t},\ TIME) + \varepsilon_{C,t}$$

Where $DEMAND_{I,t}$ is the demand in sub-level I for time t (* denotes all sub-levels);
$SUPPLY_{I,t}$ is the supply in sub-level I for time t (* denotes all sub-levels);
$FARE_{I,t}$ is the fare in sub-level I for time t (* denotes all sub-levels);
$SUPPLY_{I,t}$ is the service level in sub-level I for time t (* denotes all sub-levels);
$X_{I,t}$ is a socio-economic measure for sub-level I for time t (* denotes all sub-levels);
TIME is a monotonically increasing trend variable;
D_A is a dummy variable set to 1 when DEMAND is from sub-level A;
D_B is a dummy variable set to 1 when DEMAND is from sub-level B;
f () is a functional form;
$\varepsilon_{I,t}$ is a random error term (* denotes all sub-levels).

The drawback of the dummy variable approach is that only national estimates are available for the parameters of interest, the variations in sub-level having been lost in the dummy variables. Another problem is that the sub-level effects may cancel out or be diluted when aggregated at a national level. Whilst the sectorised approach enables sub-level estimates to be derived, the t-ratios of the parameters will be higher due to the induced multi-colliniarity which exists in the equation. The associated problems are that insignificant parameters may erroneously be included in the relationship and also information on true multi-collinearity will be lost. The separate approach will produce the same estimates as those in the sectorised models and the t-ratios and multi-collinearity statistics will be reliable but the R^2_{adj} figures may be low.

2.2 COPING WITH PROBLEMS

The application of the technique of multiple-linear regression relies on a number of statistical assumptions. Some of these assumptions can be checked before the modelling exercise (such as the distribution of the dependant and independent variables) whilst others are applicable after modelling has taken place. The important post-hoc requirements are:

(A) Homoscedasticity. The magnitude (or variability) of the error terms which emerge from the model should be constant throughout the data set. If this condition fails to hold then the estimates are still linear and unbiased but they are no-longer the best or most efficient estimates. In essence the standard error, and hence the t-ratio, of the parameter estimates are biased.

(B) Autocorrelation. There should be no relationship between successive error terms from the model. If there is a relationship then, as for the presence of heteroscedasticity in the errors, the parameter estimates are linear and unbiased but not best or efficient.

(C) Multi-collinearity. Here some of the independent variables in the model are exact or near exact functions of each other. Parameter estimates in models with collinear variables are still best linear unbiased estimates but they have small variances.

Various remedies are given in texts if any of these features are detected in the diagnostic checks (see, for example, Gujarati, 1995, Stewart and Wallis, 1981 and Wonnacott and Wonnacott, 1979). Amongst the recommendations for heteroscedasticity is either the application of a variance stabilising transformation (usually the square-root or natural logarithm) or a weighted least squares estimation procedure. Two remedies for auto-correlation are the addition of a lagged dependant variable to the list of independent variables or a regression of the variables in differenced form. Of the measures available to remedy multi-collinearity two have already been mentioned (transformations and first differences), additional measures include re-specifying the model (ie dropping variables) or augmenting the data series (with cross-sectional data).

3 BUS MARKET

3.1 DATA SERIES

The most relevant demand (dependant) variable published by the Government is probably the number of passenger journeys undertaken. This statistic is tabulated as table 5.2 in Transport Statistics : Great Britain. A preferable measure would have been passenger kilometres travelled but no direct method of obtaining this measure was apparent. The information in table 5.2 is disaggregated into London; English metropolitan areas; English shire (non-metropolitan) counties; Scotland and Wales sub-levels. The fare measure could have been the fare index (table 5.6) which is indexed within each sub-level to a base year. This means that a 1 unit fare in London may represent a different monetary value to a 1 unit fare in, say, Scotland. This was thought to be undesirable so some monetary form of fare would be more desirable. The passenger receipts (including concessionary fare re-imbursement) are published (table 5.3), so a division of the passenger receipts by passenger trips would yield a passenger fare per trip figure. This has been converted into constant money prices by using the RPI statistic and adopted as the fare variable. Once again a fare per passenger km would have been preferable, but in the absence of passenger km figures, the fare per passenger trip suffices. Level of service provision, as vehicle kilometres, are directly published (table 5.1).

Socio-economic variables of interest are Gross Domestic Product (GDP) per head; population density; private vehicles currently licensed and unemployment. These series are published in the Government reports on Regional Trends (1996) and in the Employment Gazette (1996). Dummy variables thought worthy of inclusion are a deregulation effect outside London to come into practice in 1985 and a time trend variable.

The advances on the work by Aleem are that alternative model formulations are considered, two additional years of data were now available for 1994 and 1995, England (non-London) is disaggregated into metropolitan and non-metropolitan areas (see next paragraph), the fare index was replaced with a monetary valuation of fare and the reciprocal of bus kilometres was used rather than bus kilometres (to discount extra service provision).

The reason for Aleem merging the metropolitan and non-metropolitan area bus statistics is that the socio-economic variables were not tabulated on this bases. These statistics were, however, published at the district level in Regional Trends, and estimates at the metropolitan and non-metropolitan level are obtainable by a using a weighted mean. Thus the estimated Gross Domestic Product value was a mean Gross Domestic Product per head for each of the five metropolitan districts weighted by the population in each district. The non-metropolitan value is a weighted difference between the English value and the derived metropolitan value.

The behaviour of these series over the period of interest (1982 to 1995) is given in figures 1 to 7. The passenger journey series show a decline in passenger journeys outside London, the effect being particularly marked in the metropolitan counties. Conversely bus kilometres have tended to rise throughout this period. Fares have tended to be erratic, with higher fares outside the "regulated" areas of London or the PTA (metropolitan) areas. The Gross Domestic Product graph clearly shows the presence of the early 90's recession. The population density statistic stays level throughout most of the period, with a slight rise in London towards the mid 90's. The vehicles currently licensed per capita is almost a monotonically increasing trend in all the regions under study, all be it with differing slopes. The unemployment statistics accord with a prior expectation of a fall during the boom of the late 80's followed by a rise during the recessionary period.

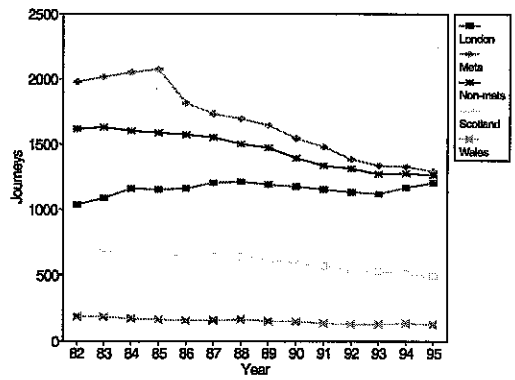

Figure 1 : Passenger Journeys

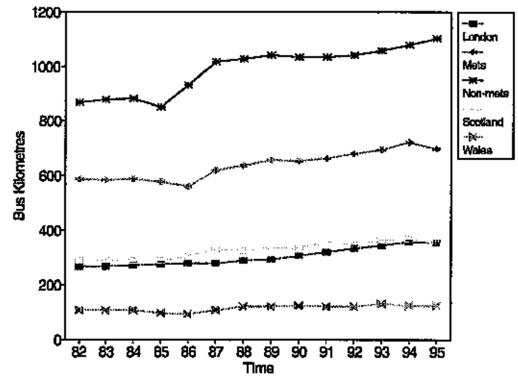

Figure 2 : Bus Kilometres

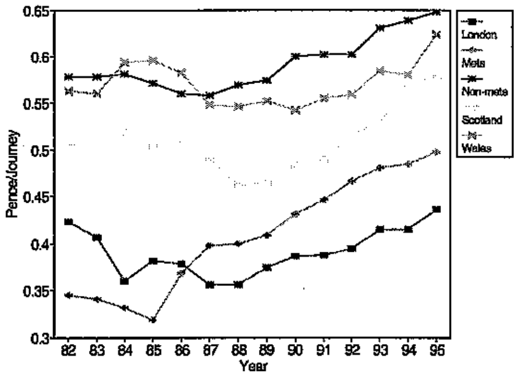

Figure 3 : Passenger Fares

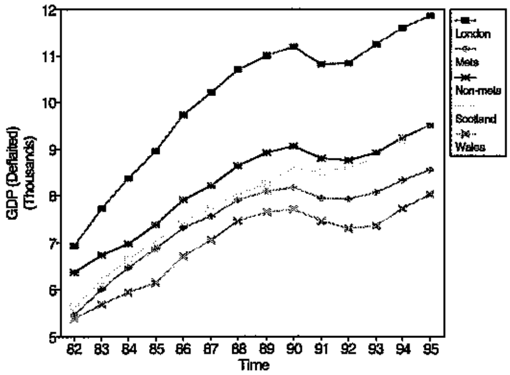

Figure 4 : GDP

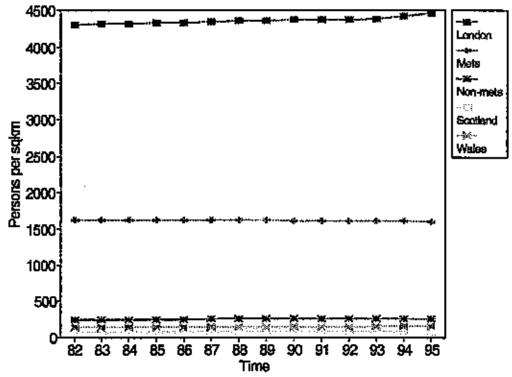

Figure 5 : Population Density

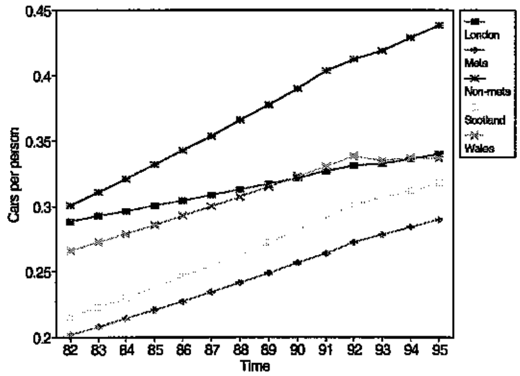

Figure 6 : Vehicles currently licensed

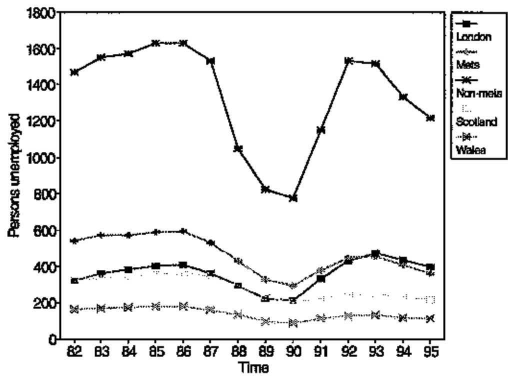

Figure 7 : Unemployment

3.2 ANALYSIS

A fitting of the dummy variable model suggests the presence of heteroscedasticity in the residuals from the model. This suggests that either a single or double log functional form is desirable. The virtue of the double-log formulation is that the parameter estimates have the immediate interpretation as the elasticity measure. Thus the double-log form was adopted for both these reasons.

3.2.1 Dummy Variable Regression

The first model presented is a stepwise dummy variable regression. The variables included are listed in table 1 along with the associated elasticity and t-ratio. All the elasticities have the correct, intuitive, interpretation and are of a plausible magnitude. The variables not included are time, population density (which is almost a constant within each region) and the deregulation effects in the non-metropolitan counties and Scotland. The value of the Durbin-Watson statistic is 1.309, which is in the region where neither H_0 or H_1 can safely be rejected. An autocorrelation function (ACF) graph of the residuals from this model calculates a significant first order lag of 0.389 with a standard error of 0.114. Evidence of multi-collinearity exists for the bus kilometres and the region dummy variables.

R^2_{adj}=0.999	η	t-ratio
FARE	-0.247	(-4.289)
BUS KILOMETRES	0.160	(-2.876)
GDP	0.523	(10.782)
VEHICLE OWNERSHIP	-1.473	(-19.828)
DEREG (mets)	-0.075	(-4.052)
DEREG (Wales)	-0.133	(-9.786)
Greater London	0.716	(33.249)
mets	0.796	(15.453)
Non-mets	1.192	(21.639)
Wales	-0.824	(-12.842)
Constant	-1.325	(-2.518)

Table 1 : Simple dummy variable regression

In order to address the problem of auto-correlation in the residuals a lagged dependant variable is introduced into the expression. The short run and long run results are given in table 2.

R^2_{adj}=0.999	SR η	t-ratio	LR η
FARE	-0.170	(-2.798)	-0.235
BUS KILOMETRES	0.146	(-2.567)	0.201
GDP	0.329	(3.813)	0.454
VEHICLE OWNERSHIP	-1.039	(-6.818)	-1.434
POP DENSITY	0.131	(8.938)	0.181
DEREG (mets)	-0.076	(-4.051)	
DEREG (Wales)	-0.088	(-4.179)	
mets	0.163	(2.295)	

Non-mets	0.659	(5.164)	
Wales	-0.674	(-5.597)	
Constant	-0.304	(-0.539)	

Table 2 : Lagged dependant dummy variable regression

The lagged dependant variable has a coefficient of 0.275 and a t-ratio of (2.916). The Durbin-Watson Statistic is unreliable in the presence of a lagged dependant variable. The ACF for first order lags is 0.276 with a standard error of 0.117, still significant but an improvement on the model in table 1. Very strong evidence of multi-collinearity was found between the lagged dependant variable, bus kilometres and the regional dummy variables.

The alternative remedy for autocorrelation is to form the variables as first differences. This modification produced an equation containing fewer terms, and one of the important variables, bus kilometres, was insignificant.

R^2_{adj}=0.479	η	t-ratio
FARE	-0.338	(-6.274)
BUS KILOMETRES	0.026	(-1.583)
DEREG (mets)	-0.085	(-3.258)
mets	-0.030	(-3.325)
Constant (TIME)	-0.114	(-4.006)

Table 3 : First difference dummy variable regression

The R^2_{adj} figure for this relationship is low at only 47.9%. The Durbin-Watson statistic of 1.338 is only a marginal improvement on that reported for the relationship in figure 1. The first order auto correlation in the residuals is 0.276 with a standard error of 0.117. There is no evidence of multi-collinearity in this model. The interpretation on the constant variable is now a monotonic, decreasing time trend. This term is now significant, suggesting that it is picking up the effects of the other, excluded, monotonic variables in the regression (private vehicle ownership or some degree of bus kilometres).

All the dummy variable models still possess problems with auto-correlation in the residuals. There is also evidence for multi-collinearity in the dummy and lagged dependant variable models. The most intuitive results are those from table 2.

3.2.2 Sectorised Regression

The SPSS package will reject any collinear variables with a high (> 10,000) Variance Inflating Factor making it impossible for a full model form to be estimated. The partial results from the first formulation are given in table 4. The main figure in each cell is the elasticity estimate whilst the number in brackets below is the t-ratio.

R^2_{adj}=0.999	Greater London	Mets	Non-mets	Wales	Scotland
FARE	-0.512 (-4.340)	-1.170 (-6.551)	-0.606 (-1.539)	-0.473 (-1.804)	-0.491 (-2.853)
BUS KILOMETRES	0.152 (-2.350)	0.025 (-0.527)		-0.382 (6.287)	0.449 (-1.696)

VEHICLE OWNERSHIP	0.411 (1.071)		-0.580 (-1.971)	-0.975 (-3.152)	
UNEMPLOYMENT	-0.020 (-0.612)	-0.021 (-0.500)	0.011 (0.352)	-0.174 (-3.529)	0.083 (1.180)
POPULATION DENSITY					0.483 (3.866)
DEREGULATION		0.024 (0.910)	0.0146 (0.334)	-0.058 (-1.852)	0.003 (0.093)

Table 4 : Simple sectorised regression

The results are confused. The signs on the parameters across variables are not consistent and some parameters are insignificant. All in all, not a very credible body of information. Gross Domestic Product is never able to enter into the equation. The Durbin-Watson statistic at 1.396 is once again in the in-determinant region and there is evidence of strong multi-collinearity. The next stage is to correct for possible auto-correlation in the residuals and see what impact this has on the plausibility of the results. These results are presented in table 5.

R^2_{adj}=0.999	Greater London		Mets		Non-mets		Wales		Scotland	
	SR η	LR η	SR η	LR η	SR η	LR η	SR η	LR η	SR η	LR η
FARE	-0.18	-0.41	-0.98	-2.2	-0.19	-0.43	-0.71	-1.6	-0.38	-0.86
BUS KILOMETRES			0.17	0.38			-0.15	-0.34		
GDP					0.08	0.18				
VEHICLE OWNERSHIP	-1.21	-2.72			-0.93	-2.09	-1.00	-2.25	-0.63	-1.42
UNEMPLOYMENT	0.03	0.06	-0.02	0	0.01	0.023	-0.01	-0.02	0.10	0.225
POPULATION DENSITY									0.08	0.18
DEREGULATION			-0.02	0	0.00	0	-0.04	-0.09	-0.02	-0.05
TIME	0.015									
Constant	1.296									

Table 5 : Lagged dependant sectorised regression

The parameter associated with the lagged dependant variable is 0.556 with a t-ratio of (3.382). The first order ACF is 0.092 with a standard error of 0.117 but there is still evidence of strong multi-collinearity. The other remedy for auto-correlation and also a remedy for multi-collinearity is to take first difference for the variables. The elasticity and (t-ratio) results are presented in table 6.

R^2_{adj}=0.877	Greater London	Mets	Non-mets	Wales	Scotland
FARE	-0.392 (-4.983)	-0.510 (-2.699)	-0.453 (-2.105)	-0.638 (-5.239)	-0.467 (-4.556)
BUS KILOMETRES	-0.258 (1.255)	0.150 (-1.107)	0.103 (-0.800)	0.260 (-3.940)	0.255 (-2.288)
GDP	0.189 (1.673)	0.471 (2.685)	0.395 (2.667)	0.535 (3.654)	0.416 (2.886)

VEHICLE OWNERSHIP	-0.845 (-1.106)	-0.963 (-1.932)	-1.197 (-3.718)	-2.069 (-6.505)	-1.247 (-3.479)
UNEMPLOYMENT	-0.001 (-0.058)	0.013 (0.504)	0.002 (0.074)	0.065 (1.914)	0.060 (1.661)
POPULATION DENSITY	6.547 (4.470)	9.346 (0.820)	0.833 (0.237)	0.539 (0.490)	0.471 (0.870)
DEREGULATION	N/A	-0.057 (-1.573)	-0.021 (-1.200)	-0.086 (-4.841)	-0.029 (-2.082)
Constant	0.004 (0.562)				

Table 6 : First difference sectorised regression

These results are a much more consistent body of evidence. Only in two situations, Greater London bus kilometres and unemployment outside Greater London are the signs intuitively suspect. The Durbin-Watson statistic is high at 2.348, but is within the region for acceptance of the hypothesis of no first order autocorrelation. There is also no evidence of multi-collinearity in this model formulation. The R^2_{adj} figure is mediocre.

3.2.3 Separate Regression

The reason for conducting separate regression exercises for each region is to overcome the problem of multi-collinearity seen in the sectorised approach, but still retain region specific elasticity estimates. The drawback to this approach is that the number of observations in the regression is reduced to only 13. With such few observations it is advantageous to only include significant variables in the regression so that the degrees of freedom for the residual can be maximised. Thus rather than including all 9 variables plus a constant in the model, only the fare, bus kilometres and the significant variables were included. These elasticities and (t-ratios) are reported in table 7.

	Greater London	Mets	Non-mets	Wales	Scotland
FARE	-0.408 (-9.801)	-0.581 (-6.325)	-0.066 (-0.486)	-0.341 (-2.437)	-0.336 (-5.726)
BUS KILOMETRES	-0.202 (2.368)	0.127 (-2.062)	0.020 (-0.257)	0.258 (-2.911)	0.160 (-1.432)
GDP	0.186 (4.383)	0.412 (5.696)	0.430 (5.297)	0.288 (2.548)	0.394 (4.718)
VEHICLE OWNERSHIP	-0.696 (-2.894)	-1.042 (-9.038)	-1.231 (-9.242)	-2.012 (-9.797)	-1.347 (-8.793)
POPULATION DENSITY	6.844 (15.694)				
DEREGULATION	N/A	-0.042 (-2.615)			
Constant	-4.764 (-4.475)	0.946 (1.417)	1.996 (2.137)	-1.322 (-0.966)	-0.062 (-0.046)
R^2_{adj}	0.990	0.999	0.993	0.976	0.993
Durbin-Watson	1.605	2.358	1.884	2.360	2.503

Table 7 : Simple separate regressions

Application of the lagged dependant variable approach produced negative or very insignificant estimates for the parameter associated with the lagged dependant variable. This approach was therefore not pursued.

The first differenced approach produced models with low R^2_{adj} figures in the non-metropolitan (13%), Welsh (32%) and Scottish (38%) regions. Only respectable R^2_{adj} were achieved for the Greater London and metropolitan areas and these elasticities and (t-ratios) results are given in table 8.

	Greater London	Mets
FARE	-0.384 (-11.857)	-0.819 (-10.995)
BUS KILOMETRES	-0.181 (1.787)	0.171 (-1.692)
GDP	0.226 (4.375)	
POPULATION DENSITY	7.157 (10.575)	
DEREGULATION	N/A	
Constant (TIME)	-0.0113 (-2.346)	-0.012 (-2.895)
R^2_{adj}	0.971	0.909
Durbin-Watson	2.368	1.366

Table 8 : First differenced separate regressions

These results appear less credible than those in table 7.

3.3 RECOMMENDATIONS

The results from the first differenced, sectorised model in table 6 appear to be the more focused of those presented in this section. All the significant parameters are of the correct sign (with the exception of the London Bus kilometres) and there are none of the statistical problems of autocorrelation or multi-collinearity present. If a lower level of disaggregation is acceptable, then the results in table 2 are to be prefered.

4 RAIL MARKET

4.1 DATA SERIES

The primary source of rail passenger volume in the United Kingdom is the rail passenger kilometres carried by rail (Table 5.11 in Transport Statistics GB, 1996) which has been tabulated for 1972 to 1995. The supply measure is loaded train kilometres (Table 5.12). No direct measure of real fare is available but a division of receipts by passenger kilometres yields a price per passenger kilometre statistic. This is deflated into constant money prices by the application of an RPI deflator. These statistic are all disaggregated by three sectors : InterCity; Network SouthEast and Regional Railways.

The other socio-economic variables considered are Gross Domestic Product per head and private vehicle ownership. Sector specific Gross Domestic Products are difficult to establish. A reasonable assumption is to use a South East England Gross Domestic Product for the Network SouthEast sector, a national Gross Domestic Product figure for the InterCity sector and a rest-of-GB (excluding South East England) for the Regional Railways sector. A similar reasoning was used for the vehicle ownership measure.

Palomo concentrated exclusively on the sectorised model formulation. Advances on the Palomo study are the consideration of alternative model formulations; one year of additional data and the re-definition of the fare and Gross Domestic Product per head variables as described above.

The behaviour of these series are given in figures 8 to 12. The rail passenger volume in figure 8 has shown erratic behaviour over the study period, with a tendency to mirror periods of prosperity and recession in the general economy. Rail train kilometres are static except for the Regional Railways sector which showed strong growth in kilometres from the mid 80's until a plateau was reached in the early 90's. A strong fare rise is seen between the years 74 to 75, and a strong dip in 1982, most pronounced for the InterCity sector. Gross Domestic Product has shown an increasing trend only moderated by periods of recession in the early 80's and 90's. Private vehicle Ownership has also shown an increasing trend with a tendency to flatten out at the start of the 90's.

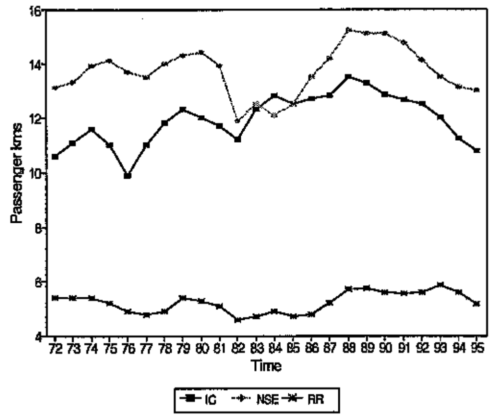

Figure 8 : Rail Passenger Kilometres

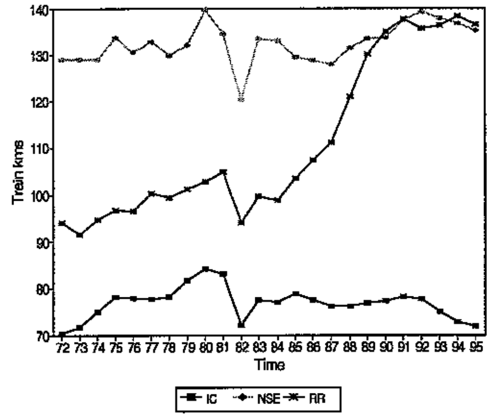

Figure 9 : Rail Train Kilometres

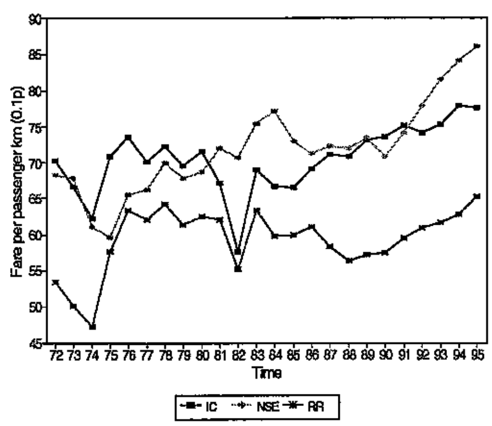

Figure 10 : Train Fare

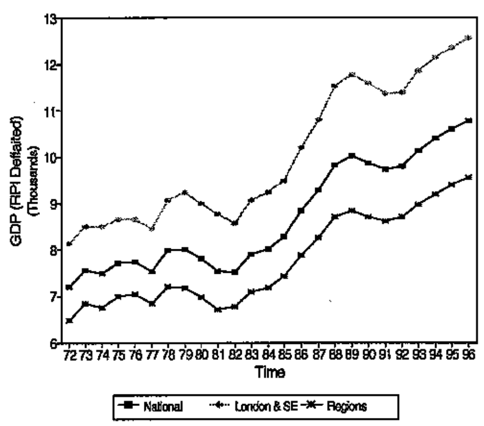

Figure 11 : Regional GDP

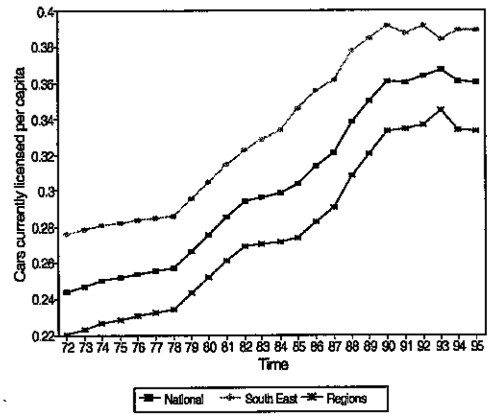

Figure 12 : Regional Vehicle Ownership

4.2 ANALYSIS

A fitting of the dummy variable model suggested once again the presence of heteroscedasticity in the residuals from the model. The double-log form was once more adopted.

4.2.1 Dummy Variable Regression

The first model presented is a stepwise dummy variable regression. The variables included are listed in table 9 along with the associated elasticity and t-ratio. All the elasticities have the correct, intuitive, interpretation and are of a plausible magnitude. The dominance of the InterCity and Network SouthEast dummy variables in explanatory power is clear from table 9. The variables not included are time and vehicle ownership. The value of the Durbin-Watson statistic is 1.0228, which is in the region where H_0 is rejected, signifying a problem with autocorrelation in the residuals. There was no evidence of multi-collinearity between the variables.

R^2_{adj}=0.986	η	t-ratio
FARE	-0.562	(-5.721)
TRAIN KILOMETRES	0.185	(-2.332)
GDP	0.382	(5.729)
InterCity	0.945	(25.530)
Network SE	0.941	(40.104)
Constant	-0.341	(-0.744)

Table 9 : Simple dummy variable regression

To account for the presence of autocorrelation a lagged dependant variable was introduced into the expression. The resultant long and short run elasticities are given in table 10.

R^2_{adj}=0.991	SR η	t-ratio	LR η
FARE	-0.287	(-3.163)	-0.586
TRAIN KILOMETRES	0.151	(-2.205)	0.308

GDP	0.406	(3.649)	0.829
TIME	-0.005	(-2.348)	
InterCity	0.466	(5.239)	
Network SE	0.399	(4.402)	
Constant	-2.297	(-2.243)	

Table 10 : Lagged dependant variable regression

The parameter on the lagged dependant variable is 0.510 with a t-ratio of (5.593). The Durbin-Watson statistic has increased to 1.415, although this value is unreliable in the presence of a lagged dependant variable. The ACF of the residuals gives a correlation of 0.254 at lag 1 with a standard error of 0.116. There is now also a suggestion of multi-collinearity between the two dummy variables and the lagged dependant variable. An alternative approach is to form a first difference model. This formulation produces a sparse model containing only the fare, kilometre's and Gross Domestic Product variables.

R^2_{adj}=0.46	η	t-ratio
FARE	-0.363	(-4.418)
TRAIN KILOMETRES	0.724	(-5.937)
GDP	0.722	(4.643)
Constant (TIME)	-0.015	(-2.821)

Table 11 : First differenced variable regression

The R^2_{adj} figure is low. The Durbin-Watson is high at 1.688 and there is no evidence of multi-collinearity.

4.2.2 Sectorised Regression

The basic sectorised model rejected the inclusion of the InterCity Gross Domestic Product, Network SouthEast train kilometres and Regional Railways Gross Domestic Product variables due to high, induced, multi-collinearity. The results which are available are given in table 12.

R^2_{adj}=0.990	InterCity		Network SouthEast		Regional Railways	
	η	t-ratio	η	t-ratio	η	t-ratio
Fare	-0.259	-1.755	-0.661	-3.956	-0.382	-2.898
Train Kilometres	1.025	-5.842			0.684	-5.222
GDP			0.618	6.530		
Vehicle ownership	0.817	3.618	0.315	1.325	0.086	0.361
Time	-0.009	-2.022				
Constant	0.227	0.306				

Table 12 : Simple sectorised regression

The fare elasticity for InterCity is insignificant at the 5% level. Vehicle ownership fails to provide any significant explanatory power for the Network SouthEast and Regional Railways sectors. The parameter values all have the correct, intuitive, sign. As with all "un-treated" sectorised models there is high multi-collinearity amongst all the variables. The Durbin-Watson statistic is low at 1.1015. The introduction of a lagged-dependant variable retains the same set of included variables.

R^2_{adj}=0.991	InterCity		Network SouthEast		Regional Railways	
	SR η	LR η	SR η	LR η	SR η	LR η
Fare	-0.235	-0.361	-0.525	-0.807	-0.257	-0.395
Train Kilometres	0.748	1.150		0.000	0.466	0.716
GDP			0.453	0.696		
Vehicle ownership	0.521	0.801	0.188	0.289	0.044	0.068
Time	-0.006	-0.009				
Constant	0.067					

Table 13 : Lagged dependant variable sectorised regression

The parameter estimate on the lagged dependant variable is 0.349 with a t-ratio of (3.120). The first order autocorrelation is 0.261 with an standard error of 0.116. Multi-collinearity is still a feature of the models. The final set of results in this section are of a first differenced formulation. A first attempt produced a model with all insignificant private vehicle ownership parameters and a poor R^2_{adj} value (41%). A set of revised results omitting this variable are presented in table 14.

R^2_{adj}=0.427	InterCity		Network SouthEast		Regional Railways	
	η	t-ratio	η	t-ratio	η	t-ratio
Fare	-0.480	(-2.870)	-0.395	(-2.140)	-0.299	(-2.363)
Train Kilometres	0.900	(-3.582)	0.649	(-3.013)	0.715	(-3.354)
GDP	0.840	(2.966)	0.866	(3.623)	0.511	(1.910)
Constant (TIME)	-0.015	(-2.709)				

Table 14 : First differenced sectorised regression

The Durbin-Watson statistic is high at 1.725 and there is no evidence of multi-collinearity. The R^2_{adj} statistic is low.

4.2.3 Separate Regression

The use of a separate regression model for each operation sector reduced the number of observations in each model to 24. Given this reduced sample size more emphasis than normal should be placed on producing a parsimonious model formulation. Thus only significant variables will be included in the model form.

	InterCity		Network SouthEast		Regional Railways	
	η	t-ratio	η	t-ratio	η	t-ratio
Fare	-0.440	(-2.226)	-0.850	(-7.150)	-0.321	(-2.980)
Train Kilometres	0.773	(-2.949)	0.706	(-2.864)	0.827	(-6.508)

GDP			0.480	(6.252)		
Vehicle ownership	0.410	(4.523)				
Time					-0.011	(-3.853)
Constant	1.488	(1.359)	-1.613	(-1.522)	-0.787	(-1.018)
R^2_{adj}	0.500		0.740		0.750	
Durbin-Watson	0.8478		1.096		1.506	

Table 15 : Separate regression

As expected, the magnitude of multi-collinearity in the data set has been considerably reduced. The R^2_{adj} statistics are low and the Durbin-Watson statistics suggest the presence of auto-correlation. The introduction of a lagged dependant variable only produced a sensible result for the InterCity market. The sparse results from this exercise are given in table 16.

R^2_{adj}=0.819	InterCity		
	SR η	t-ratio	LR η
Fare	-0.458	(-2.638)	-0.690
Train Kilometres	0.480	(-2.163)	0.723
GDP	0.613	(4.208)	0.923
Time	-0.007	(-1.872)	-0.011
Constant	-4.193	(-2.393)	

Table 16 : Lagged dependant variable separate regression

The parameter on the lagged dependant variable is 0.336 with a t-ratio of (2.814). The first order auto-correlation statistic is 0.347 with a standard error of 0.196. The final set of results are for a first differenced model and are presented in table 15.

	InterCity		Network SouthEast		Regional Railways	
	η	t-ratio	η	t-ratio	η	t-ratio
Fare	-0.273	(-1.453)	-0.403	(-3.047)	-0.268	(-1.840)
Train Kilometres	-0.750	(-2.439)	0.653	(-4.272)	0.664	(-2.798)
Gross Domestic Product			0.838	(4.546)		
Constant			-0.013	(-1.917)		
R^2_{adj}	0.147		0.697		0.228	
Durbin-Watson	1.552		1.803		1.588	

Table 15 : First difference separate regression

The R^2_{adj} statistics are very poor for the InterCity and Regional Railways sectors and only just respectable for the Network SouthEast sector. Given this lack of explanatory power little credibility can be given to these results.

4.3 RECOMMENDATIONS

The majority of these results are characterised by low fare and high service elasticities. This may be indicative of a degree of simultaneity between these two measures. The most plausible set of results are presented in table 14 for a first differenced, sectorised model formulation. This formulation contains no statistical problems but does possess a low R^2_{adj} statistic.

5 PRIVATE TRAVEL

Two measures of vehicle activity are of interest - vehicle ownership and vehicle use. To a large extent these two measures are inter-related, vehicle use requires ownership and anticipated use determines an ownership decision.

5.1 DATA SOURCE

The DoETR publishes vehicle ownership as motor vehicles currently licensed (Table 3.1). For this study all body-type cars are considered, which includes both private and light goods vehicles. The dependant variable for the **ownership** is a function of vehicles currently licensed per capita. The explanatory variables are Gross Domestic Product per head of population, a price basket for vehicle ownership (composed of purchase, maintenance, petrol and oil and tax and insurance costs) and a time trend. The behaviour of the series over time are show in figures 13 to 15.

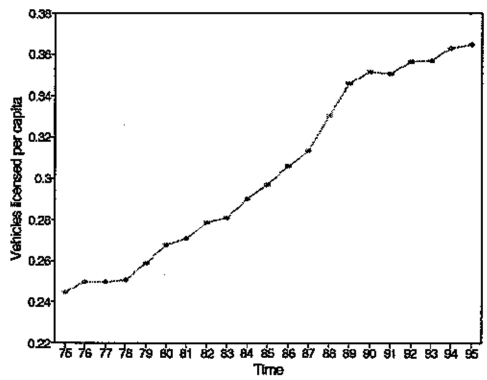

Figure 13 : Vehicle Ownership

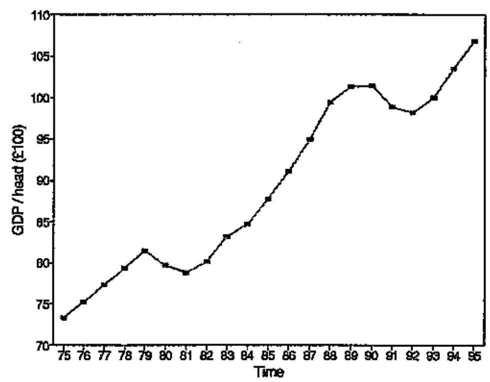

Figure 14 : GDP per capita

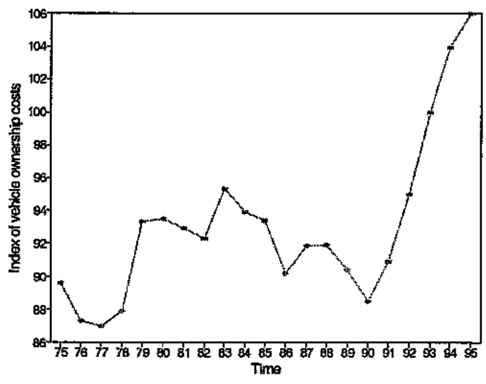

Figure 15 : Vehicle ownership cost index

The only addition to the work considered by Batley is the addition of two years worth of data at the end of the series.

The dependant variable for **vehicle use** is passenger kilometres (Table 1.1). The independent variables are road kilometres (Table 3.17), Gross Domestic Product per head, a basket cost of road travel (composed of maintenance, petrol and oil and tax and insurance costs), an index of bus fares and a time trend. Figure 16 gives a bi-variate scatter plot of all these variables. The top row of plots all have car km's on the vertical axis, with the first graph having GDP per head on the horizontal axis, the second, running cost on the horizontal axis and so on. The second row have GDP per head on the veritcal axis.

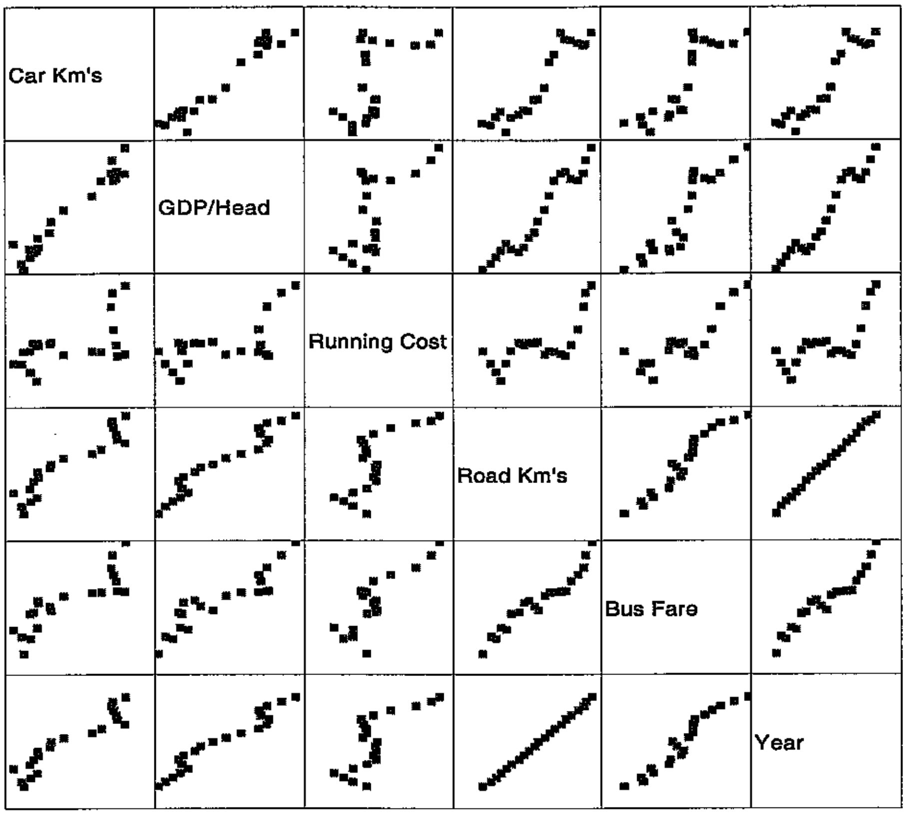

Figure 16 : Bi-variate scatter plot of vehicle use variables

Once again only two years of additional data were available.

5.2 FUNCTIONAL FORM FOR OWNERSHIP

There is a need to determine the saturation level of ownership, S, from the relationship:

$$\ln\left(\frac{O_t}{S-O_t}\right) = \beta_0 + \beta_1 \ln(GDP_t) + \beta_2 \ln(COST_t) + \beta_3 t + \varepsilon$$

Where O_t is the level of vehicle ownership (vehicles currently licensed per captia);
S is the saturation level of vehicle ownership (vehicles licensed per captia);
GDP_t is Gross Domestic Product per capita;
$COST_t$ is a composite of the cost of vehicle ownership;

t is a time trend variable;
β_0, β_1, β_2, β_3 are parameters to be estimated.

The appropriate level of ownership can be inferred from the value of S which gives the best "goodness-of-fit" statistic (R^2_{adj}) from a range of possible values for S.

5.3 VEHICLE OWNERSHIP RESULTS

An immediate problem with the above functional form was that auto-correlations existed in the residuals from the model. The introduction of a lagged dependant variable did produced residuals with insignificant first order auto-correlations. The results for a range of S values are given in table 16.

S	COST	GDP	t	Const	Lag DV	R^2_{adj}
0.40	-0.09 (-0.210)	1.334 (2.457)	0.0699 (0.423)	-5.201 (-1.425)	0.7109 (6.904)	0.9905
0.45	-0.269 (-0.949)	0.812 (2.369)	0.0141 (1.179)	-2.234 (-0.962)	0.6159 (5.217)	0.9919
0.50	-0.299 (-1.275)	0.611 (2.242)	0.0159 (1.513)	-1.325 (-0.711)	0.5624 (4.284)	0.99218
0.55	-0.304 (-1.453)	0.504 (2.139)	0.0163 (1.686)	-0.927 (-0.571)	0.5272 (3.720)	0.99223
0.60	-0.302 (-1.562)	0.437 (2.058)	0.0163 (1.785)	-0.729 (-0.496)	0.5024 (3.348)	0.9222
(0.565)	-0.304 (-1.491)	0.481 (2.113)	0.016 (1.721)	-0.8535 (-0.543)	0.5189 (3.593)	0.99223

Table 16 : Vehicle ownership results

The above results would suggest an S value of between 0.50 and 0.60. A finer grid for S shows that 0.54,0.55,0.56,0.57,0.58 and 0.59 all have an R^2_{adj} value of 0.99223. In the absence of other information, the value of S will be (0.54+0.59)/2 = 0.565.

These estimates can then be used to produce both short and long run elasticities (see appendix for the expressions used to calculate the elasticities). These elasticities are presented in table 17.

	SR η	LR η
GDP	0.233	0.432
Cost	-0.147	-0.273
Time	0.670	1.233

Table 17 : Summary vehicle ownership elasticities

The evolution of these elasticities over time is shown in figure 18. Both the short and long run Gross Domestic Product elasticities are decreasing over time. In absolute terms, the cost elasticity is also decreasing. The short run time elasticity is increasing whilst the long run is decreasing. These results tend to suggest an eventual convergence between short and long run behaviour.

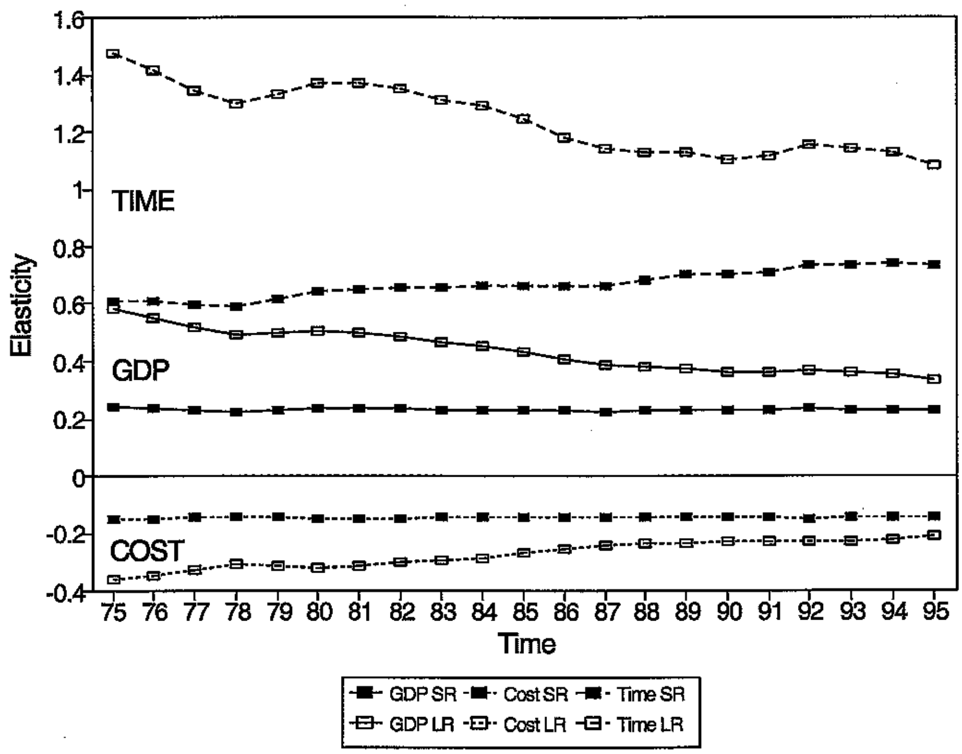

Figure 17 : Elasticities through time

5.4 FUNCTIONAL FORM FOR PRIVATE VEHICLE USE

The basic model form is:

$$\ln\left(\frac{D_t}{V_t}\right) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 \ln(COST_t) + \beta_4 \ln(GDP_t) + \sum_{i=1}^{n} \beta_{i+4} X_{i,t} + \varepsilon$$

Where D_t is the total distance travelled by all vehicles;
V_t is the number of vehicles;
$COST_t$ is a measure of vehicle running costs;
GDP_t is the Gross Domestic Product per capita;
$X_{i,t}$ are socio-economic variables.

This formulation is of recognisable format. The inclusion of a t and t^2 term is novel. Care is required in the choice of the t series. Whether t is counted as 1975, 1976, ... or 75, 76, ... or 1, 2, ... will have a direct effect on all the parameter estimates.

5.5 PRIVATE VEHICLE USE RESULTS

	CONST	t	t^2	GDP	COST	DW	R^2_{adj}
Full model	11.60 (6.52)	-0.048 (-2.10)	0.00033 (2.22)	0.236 (1.33)	-0.289 (-2.36)	1.57	0.9450
WO time	7.61 (32.9)			0.527 (13.83)	-0.068 (-1.033)	1.50	0.9353

Table 18 : Private vehicle use

None of the socio-economic variables are significant factors. The Gross Domestic Product, t and t^2 terms are very co-linear so the final row presents a model containing only Gross Domestic Product and running COSTS. This clearly shows that Gross Domestic Product (or indeed time) is the only driving force in the relationship (see also the multiple scatter plots in figure 16).

6 TWO STAGE LEAST SQUARES MODELLING OF RAIL DATA

All the models considered up to now have been single equation models. In such equations there is a clear distinction between the dependant (endogenous) and independent (exogenous) variables. The relationship is such that the independent variables explain the behaviour of the dependant variable. This one-way relationship may not, however, be appropriate. It may be the case that some of the "independent" variables are determined by the dependant variable. A system is then built which contains a number of equations with possible recursive relationship built into them.

The two-way relationship invalidates the assumption in ordinary least squares that the error terms from the model are un-correlated with an "independent" variable. Failure to observe this assumption means that the estimates are biased and inconsistent.

6.1 A SYSTEM OF EQUATIONS FOR THE RAIL DATA

To explore the consequences of this result, the rail data is taken as an example of an econometric transport data set. The basic formulation will be the dummy variable approach. A prior hypothesis is that the simultaneity relationship may exist between the passenger and the service kilometres. The system of equations may be:

$$PK_t = \alpha_0 + \alpha_1 F_t + \alpha_2 TK_t + \alpha_3 GDP_t + \alpha_4 VCLPC_t + \alpha_5 TIME + \varepsilon_t$$

$$TK_t = \beta_0 + \beta_1 PK_t + \beta_2 GDP_t + \beta_3 IC + \beta_4 NSE + \varepsilon_t$$

Where PK_t is the passenger kilometres during year t;
TK_t is the train service kilometres during year t;
GDP_t is the appropriate Gross Domestic Product per capita;
$VCLCP_t$ is the vehicle currently licensed per capita;
TIME is a monotonically increasing time trend;
IC is an InterCity dummy variable;
NSE is a Network SouthEast dummy variable;
α_0, α_1, α_2, α_3, α_4, α_5, β_0, β_1, β_2, β_3, β_5 are parameters to be estimated.

To some degree the composition of the equations can be arbitrary so long as it meets certain structural requirements. The passenger kilometre equation (PK) is similar to those encountered in section 4. The train kilometres equation includes Gross Domestic Product and sector dummy variables. Where appropriate, all the variables in the two equations are in logarithm form.

6.2 A TEST FOR SIMULTANEITY

A test to establish whether simultaneity exists between a set of variables is given by Pindyck and Rubinfeld (1981). In our case this would involve the regression of PK on F, GDP, VCLPC, TIME, IC and NSE (step 1). The predicted values and the residuals from this regression are then regressed on TK (step 2). If the parameter value associated with the residuals term is significant then there is evidence for simultaneity.

STEP 1	F	GDP	VCLPC	TIME	IC	NSE	R^2_{adj}	Lag 1 ACF
PK	-0.453 (-3.957)	0.484 (10.449)	0.523 (2.206)	-0.012 (-2.242)	0.801 (20.728)	0.829 (13.218)	0.999	0.493 (0.114)
STEP 2	PRED	RESID	Constant				R^2_{adj}	
TK	-0.015 (-0.23)	0.428 (0.734)	4.677 (29.969)				-0.020	

Table 19 : Test of simultaneity

Of immediate concern is the presence of auto-correlation in the residuals from the STEP 1. The approach of introducing a lagged dependant variable is not appropriate when estimating simultaneous equations. The alternative approach of first differences is adopted. The inclusion of the TIME variable is no-longer appropriate since its function will now be carried out by the constant term. The sector dummy variables are also inappropriate since first differences remove the differing level effects which these dummy's capture.

STEP 1	F	GDP	VCLPC	Constant (TIME)	R^2_{adj}	Durbin-Watson
PK	-0.142 (-1.441)	0.750 (3.974)	0.608 (1.771)	-0.-022 (-2.608)	0.210	1.952
STEP 2	PRED	RESID	Constant		R^2_{adj}	
TK	0.164 (0.100)	0.438 (5.036)	0.006 (1.555)		0.256	

Table 20 : Test of simultaneity on differenced data

Since the parameter on the RESID term of step 2 is significant there is evidence for simultaneity between PK and TK. The low R^2_{adj} in step 1 is a disappointment. Step 2 was repeated with Gross Domestic Product and vehicle ownership replacing the PRED value but produced a similar poor fit in step 2. The low R^2_{adj} may invalidate the test results.

6.3 ESTIMATION OF MODEL PARAMETERS

The estimates for the parameters of the two equations are given in table 20.

	PK		TK	
PK			-0.126	(-0.293)
TK	0.145	(-0.158)		
F	-0.013	(-0.050)		
GDP	0.585	(2.216)	0.015	(0.042)
VCLPC	2.252	(1.355)		
IC			-0.015	(-1.146)
NSE			-0.014	(-1.072)
Constant (TIME)	-0.022	(-2.200)	0.016	(1.314)
R^2_{adj}	0.08		-0.03	
Durbin-Watson	1.902		2.229	

Table 20 : Two stage least square regression

Very little of significance emerges from this model formulation. Many of the parameters are insignificant and the R^2_{adj} are very much on the low side.

7 RECOMMENDATIONS

REFERENCES

Aleem, T (1995). Forecasting the Demand for Bus and Coach Travel. MA Thesis, Institute for Transport Studies, University of Leeds, Leeds, UK.

Batley, R (1996). New Directions in Forecasting Car Use. MA Thesis, Institute for Transport Studies, University of Leeds, Leeds, UK.

Employment Gazette. Department of Education. HMSO.

Gujarati, DN (1995). Basic Econometrics. Third Edition. McGraw-Hill, New York.

Palomo, AP (1996). Analysing the Demand for Passenger Rail Travel. MA Thesis, Institute for Transport Studies, University of Leeds, Leeds, UK.

Pindyck, RS and Rubinfeld, DL (1981). Econometric Models and Econometric Forecasts. Second Edition, McGraw-Hill, New York.

Regional Trends (1996). Department of the Environment, Transport and the Regions. HMSO.

Stewart, MB and Wallis, K (1981). Introductory Econometrics. Second Edition. Basil Blackwell Publisher, Oxford, UK.

Transport Statistics : Great Britain (1996). Department of Environment, Transport and the Regions. HMSO.

Wonnacott, RJ and Wonnacott, TH (1979). "Econometrics", Second Edition. John Wiley and Sons, New York.

APPENDIX : Expressions for Short and Long Run Elasticities

The appropriate expressions for the elasticities from the formulation for the relationship between car ownership are given below (from Preston, 1997)

The expression for explaining car ownsership is give by:

$$\ln\left(\frac{O_t}{S-O_t}\right) = \beta_0 + \beta_1 \ln(GDP_t) + \beta_2 \ln(COST_t) + \beta_3 t + \beta_4 \ln\left(\frac{O_{t-1}}{S-O_{t-1}}\right) + \varepsilon$$

Where O_t is the level of vehicle ownership (vehicles currently licensed per captia);
S is the saturation level of vehicle ownership (vehicles licensed per captia);
GDP_t is Gross Domestic Product per capita;
$COST_t$ is a composite of the cost of vehicle ownership;
t is a time trend variable;
β_0, β_1, β_2, β_3, β_4 are parameters to be estimated.

The β_0, β_1, β_2, β_3, β_4 parameters can be further expressed as -ln(b), fS, gS, aS, (l-1)/l respectively.

The expression for the Short-run elasticities are given by:

$$SR\,\eta_{GDP} = O_t fb(GDP^{-fS}\ COST^{-gS}\ e^{-aSt})$$

$$SR\,\eta_{COST} = O_t gb(GDP^{-fS}\ COST^{-gS}\ e^{-aSt})$$

$$SR\,\eta_{TIME} = O_t abt(GDP^{-fS}\ COST^{-gS}\ e^{-aSt})$$

The expression for the Long-run elasticities are given by:

$$LR\,\eta_{GDP} = (-l)(-f)O_t(bGDP^{-fS}\ COST^{-gS}\ e^{-aSt})^l$$

$$LR\,\eta_{COST} = (-l)(-g)O_t(bGDP^{-fS}\ COST^{-gS}\ e^{-aSt})^l$$

$$LR\,\eta_{TIME} = (-l)(-a)O_t(tbGDP^{-fS}\ COST^{-gS}\ e^{-aSt})^l$$