Article:

Clark, S.D. (2007) Estimating local car ownership models. Journal of Transport Geography, 15 (3). pp. 184-197. ISSN 0966-6923

https://doi.org/10.1016/j.jtrangeo.2006.02.014

Institute of Transport Studies
University of Leeds

Published paper

Clark, S. (2007) Estimating Local Car Ownership Models, Journal of Transport Geography, 15(3), pp.184-197

Estimating Local Car Ownership Models

Stephen D. Clark (Monitoring Officer),
Transport Policy Monitoring, Development Department,
The Leonardo Building, 2 Rossington Street,
Leeds City Council, LEEDS, UK, LS2 8HD.
Stephen.Clark@leeds.gov.uk
Tel +44 113 39 50429
FAX +44 113 24 76386

Keywords : Regression; Car Ownership; Income; Local Estimation.

1 Introduction

The decision by a household to own a first or additional vehicle can be based on a variety of factors, with commonly cited reasons including household location, workplace location, lifestyle commitments and personal status. (Dargay, 2002, Karlaftis and Golias, 2002). Coupled with these factors is whether the household has sufficient income to purchase, maintain and accommodate the vehicle. Given the importance of car ownership in determining individual and household travel behaviour, there are many published studies that attempt to assess the impact of these various factors on car ownership. Rather than attempt to review this vast range of studies, four recent or well maintained resources are mentioned here that provide an authoritative review of the relevant studies (on-line versions of these resources are also available, see the reference list).

The first resource is the Travel Demand Management Encyclopaedia maintained by the Victoria Transport Policy Institute – which has a section devoted to Transport Elasticities. This section of the encyclopaedia begins with an introduction to the topic area followed by a few key mathematical definitions. The bulk of the section then contains a list of over 20 factors that impact on travel demand – using review information obtained from other studies. These travel demands include car ownership, car use and transit use whilst the influential factors include price, income and provision of infrastructure and services. This resource is an ideal introduction to this area of research and for off-line reading a pdf document version is available (Litman, 2005).

The second study was produced by the Rand Europe Corporation (de Jong et al, 2004) as a prelude to the updating of the Dutch national car ownership prediction model. In this paper the authors review a range of domestic Dutch and international studies on car ownership. The review is relatively contemporary, covering studies conducted since the early 1990's, and it outlines nine modelling methodologies used to model the relationship between car ownership and factors such as income, fixed/variable costs, quality of car stock and license holding. Since the paper concentrates heavily on technical descriptions of modelling techniques it is ideal for those who wish to gain a thorough understanding of the variety of work in this field of estimation from just one coherent, independent source. The paper does, however, contain little by way of quoted results from the models described in the paper. The interested reader is therefore required to refer to the individually cited papers for such information.

The third study was conducted by the World Bank (Ingram and Liu, 1999) and is an international study of both motorization and road provision. Whilst the previous mentioned study by the Rand Corporation concentrated on the modelling methodologies, this study concentrates mainly on the outcomes from models – primarily the elasticities. From this study it is clear that one of the most quoted determinants of car ownership is income – either at the national (GDP) or personal (household) level. In their table 2 the authors provide a list of income elasticities estimated from cross-sectional, time-series and panel type data. They suggest that elasticities estimated using cross-sectional data and at the national level are larger than those from time series data and for just urban areas. A discussion is then provided of other determinants quoted in the literature such as vehicle purchase costs, running costs (petrol, insurance and maintenance) and population density. The remaining parts of this report are primarily concern with road space and transport provision (the ratio of vehicles to roads).

The final study comprises two similar reports commissioned by the United Kingdom Department for Transport (Goodwin, Dargay and Hanly, 2004 and Graham and Glaister, 2004). The Goodwin paper only considers results based on United Kingdom (or similar countries) data, but even so they were able to gather the impact of price and income impacts on car travel from 69 published studies – providing nearly 500 elasticity estimates. They also provide some assessment of the possible causes of and policy implications of different elasticity estimates. The review within the Graham paper initially covers similar ground to that in the Goodwin paper but devotes the bulk of its content to the consideration of freight elasticities.

Many of the models referred to so far use a regression model or a refinement of such a model to represent the relationship between car ownership and factors such as income. As outlined in the de Jong et al (2004) paper some of these models use cross-sectional data, where data is from a single period of time but from many locations, whilst some others use time-series data, where the data is from many time periods but usually only one location. In addition some models use highly dissaggregate data where detailed information is known about individuals or households whilst some are based on information about the characteristics of an area in aggregate form. The model presented in this paper is a cross-sectional model on aggregate data.

2 Data Source

The data that is used in this study is provided by the United Kingdom Office for National Statistics (ONS) and is freely available in electronic format, either from the internet or on CD’s and DVD’s that ONS are able supply on request.

2.1 Car Ownership Counts

The United Kingdom Government conducts a national census of its population every ten years. The Census includes questions on household composition, housing stock, general health, education and, importantly for this study, car ownership and travel to work behaviour. This information is then used for policy and planning decisions at all levels of central and local government as well as by many bodies outside government.

ONS has now completed its release of tables of statistics from the 2001 Census. The data is provided at a range of geographical levels, from the small Census Output Area (COA) level (consisting on average of 150 households), through Super Output Areas (500 to 1000 households), electoral wards, districts and on to the constituent nations of the UK.

The primary table used in this study is Key Statistic 17, residential car and van ownership, produced at the ward level. In this table, information is recorded on how many households do not own a car, own just one car, two cars, three cars and finally, four or more cars. The total number of households and cars in the ward is also provided and this enables an average car ownership level per household to be calculated for the ward.

Figure 1 shows the geographical distribution of car ownership, expressed as the number of cars per household. Areas of the country that have good public transport provision and nearby access to services (shops, schools, medical, etc) would be expected to have low levels of car ownership whilst rural areas without this level of provision would be expected to have higher levels of car ownership. Looking at the map it is clear that car ownership is lowest in the urban areas of northern England, south Wales, the central West Midlands and the centre of London. Elsewhere car ownership is high, particularly in the less urban parts of the South East and Eastern England.

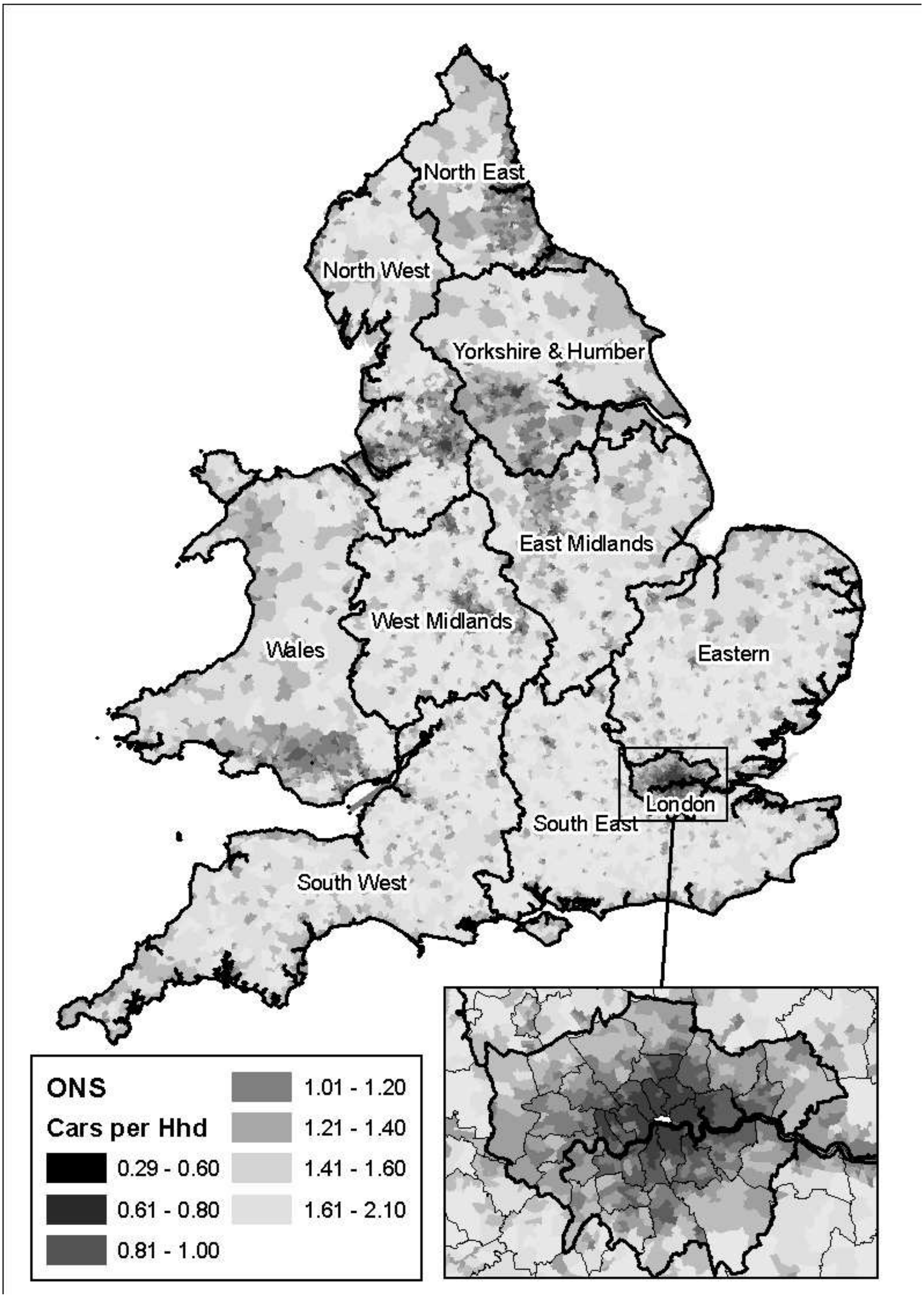

Figure 1 : Residential car and van ownership (cars and vans per household)

2.2 Income Estimation

One question that was not included in the 2001 Census was a question on the household's income. In April 1999 a trial Census was conducted and as part of this trial an income question was included on a number of the Census forms. Evidence from the response rate in those areas where an income question was included suggested that people were less likely to return a Census form if an income question was included. This, coupled with the difficulty in what to include as income, led the Government to drop the income question from the full Census.

There was still, however, recognition that estimates of household incomes for small geographical areas were required. To this end a project was established to produce income estimates at a level of geography as low the local authority ward (Williams, 2000). A brief summary of the methodology is presented here.

Information on household income is regularly collected as part of the Government's Family Resource Survey (FRS) (Department for Work and Pensions, 2005). The clustered coverage of this survey is not, however, sufficient to provide reliable estimates of household income for every UK ward. For those wards that have sufficiently reliable estimates of income from the FRS, a relationship was established between the income and a range of nationally available covariates. For the study, covariate data was taken from a number of sources: 2001 Census data; the Department of Works and Pensions claimant count; the Land Registry dwelling price data; Local Authority Council Tax data and regional indicators. Once this relationship was calibrated and validated for the subset of wards, it was then applied to the nationally available covariate data to provide income estimates for all wards. Whilst car ownership was included in this process as a candidate national covariate it was not selected for inclusion in the final model (Goldring et al, 2005) thereby eliminating a possible "feedback" effect. This is important since car ownership is commonly used as a proxy for incomes, eg in Longley and Tobón, 2004, an attempt is made to explain the variability in household incomes in Bristol, UK, using a basket of socio-demographic indicators, one of which is the number of households owning two or more cars.

Returning to the ONS study, estimates (and the associated 95% confidence intervals) for four types of household weekly income were produced: Gross (unequivalised); Net (unequivalised); Net before housing costs (equivalised) and Net after housing costs (equivalised). The process of equivalisation adjusts the income to take account of the composition of households in the area. Thus three households with the same unequivalised weekly income of £200 could translate to £264 for a single person household, £200 for a couple and just £172 for a couple with two children.

There are 31 wards with unusually small resident populations (mainly in the City of London) and to protect the confidentiality of these residents' incomes no estimates are provided. This leaves 8 837 wards with an income estimate. Figure 2 shows how the net (unequivalised) income measure varies across England and Wales. Whilst it was perhaps once the case that incomes in the UK tended to be fairly homogenous, trends in the gap between the poorest and richest in society have been dynamic in the past 20 years and this has resulted in some wide income disparities. Effects that widen this gap include higher annual pay rises for skilled as opposed to unskilled labour, increases in both two earner and no-earner households and income tax cuts. On the opposite side, the increase in indirect taxation and the introduction of means-tested benefits tends to reduce this income gap. The result of these dynamics has tended to concentrate high income households in the South East and Eastern England along with some pockets in the rural north. Most of Wales, the extreme South West and the east end of London along with the inner wards of the larger cities and towns in northern England tend to have low incomes.

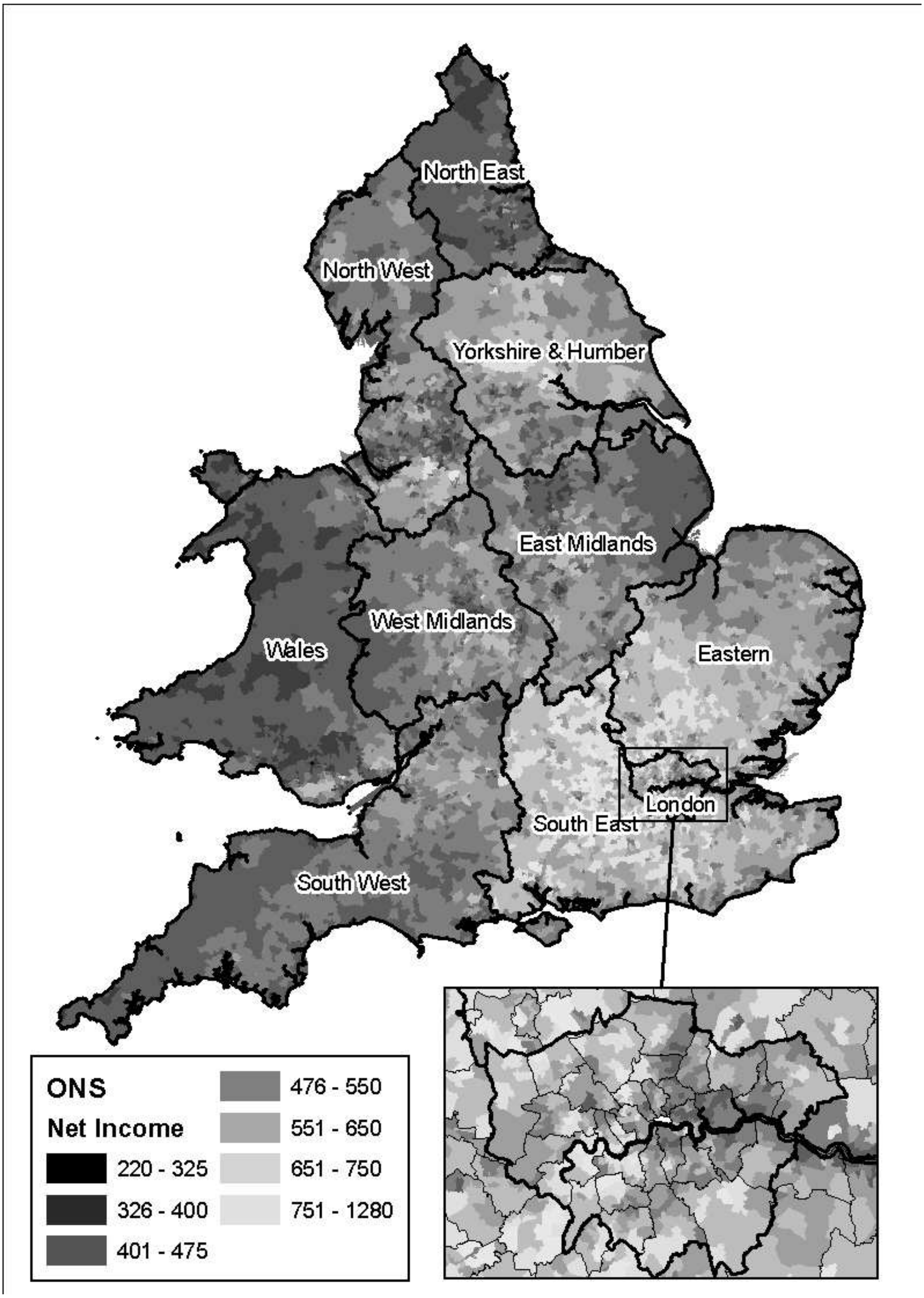

Figure 2 : Distribution of household net income (£ per week)

Comparing figures 1 and 2 immediately highlights some contrasting features. In the towns and cities of northern England, both incomes and car ownership are at low levels, whilst in the South West incomes are also low but car ownership is high. For most of the South East, both household incomes and car ownership are high but in west London, incomes are also high but car ownership is low. This suggests that the strength of the relationship between the two statistics varies across England and Wales and there is a need to recognise that these contrasts are becoming starker at ever finer geographical scales.

2.4 Other Explanatory Variables

To help account for some of the apparently contradictory features reported above, consideration was given to whether the inclusion of additional variables would help to improve the quality of the relationship between household car ownership and incomes. An obvious candidate is a variable that takes account of the demographic nature of the ward – either population density or household size. With population density, it could be hypothesised that car ownership is likely to be higher in a rural ward because there is a poorer provision of public transport and also people will need to travel further to access shops and services. A variable that specifies the population density of the ward (in people per hectare) is selected as one to help capture this accessibility effect.

Household size and structure is also commonly found to be a significant explanatory factor when examining car ownership. Larger households tend to have more cars. The concern here is that the counts that are used to define this variable – number of people and number of households are already present in two other model variables and may therefore present "co-linearity" issues. For this reason a household size variable is not included in this model.

Other candidate variables usually used in estimating time series car ownership relationships include purchase price, fuel price and taxation. Since the model proposed here is entirely cross-sectional in nature the usefulness of these additional variables will be limited – in the local area there will be little variation in the values of these variables.

3 Global Relationship

Looking at a scatter plot of these data, a single logarithm relationship appears to be the most appropriate. A consequence of this model formulation is that the income elasticity decreases as the number of cars per household increases, an entirely natural result since it is to be expected that households with high numbers of cars are less likely to use income to purchase further cars. Also income and population density values tend to be positively skewed, so by taking the logarithm of these variables the distribution becomes less skewed and more "normal-like". Bringing both the net income estimate and the population density variables into a single logarithmic regression model gives the following equation (with standard errors):

$$\mathit{cars\ per\ household} = \underset{(0.046)}{-3.075} + \underset{(0.007)}{0.744}\ln(\mathit{net\ income}) - \underset{(0.001)}{0.114}\ln(\mathit{population\ density}) \tag{1}$$

This rather parsimonious model has a relatively large R^2_{adj} value of 75.3%.

The average car ownership level in the UK is 1.223 cars per household which suggests a mean car ownership elasticity with respect to income of η_{co} = (0.744 / 1.223) = 0.608. This value predicts that a rise of 10% in incomes will produce a corresponding 6.1% increase in car ownership.

Whether this cross-sectional elasticity is best compared with a short or a long-run time series elasticity is a matter of debate. The estimate is between the rule of thumb values suggested in Goodwin, Dargay and Hanly, (2004) for short-run (η_{co} = 0.4) and long-run (η_{co} = 1.0) elasticites. Goodwin et al (2004), however, refuse to be drawn on the debate as to which type of elasticity is the best comparator, saying that "(they) do not support the common practice of using phrases 'short term' and 'long term' as legitimate labels for either cross-section equilibrium modes or unlagged time series models…". Both Fearnley and Bekken, JT (2005) and Ingram and Liu (1999) do, however, suggest that cross-sectional elasticities are best compared with long-run time-series elasticities. Not withstanding these opinions, here comparable elasticities of both types are provided and it is left to the reader to decide which of the two elasticity types they believe to be the best comparator.

In Table 5, Goodwin et al (2004) publishes relevant results from 15 studies, with the range of short run elasticities being ($0.08 \leq \eta_{co} \leq 0.94$) and long-run elasticities being ($0.28 \leq \eta_{co} \leq 1.62$) (results are not quoted from Table 6 of their paper since the number of studies on which the relevant results are based is small). The value of 0.608 sits comfortably within both these ranges. The population density elasticity is $\eta_{pd} = -0.093$, which is similar to the value suggested in Ingram and Liu (1999) of $\eta_{pd} = -0.1$ for a national data set.

Figure 3 shows a map of the residuals in each ward from this global model. Inspection of the map suggests that there may be some spatial correlation in these residuals. The East and West Midlands region appears to show consistently positive residuals whilst central London and areas of Yorkshire and Humber show negative residuals. The computed values of Moran's I statistic (Moran, 1950) for various distance bands using the ROOKCASE EXCEL addin (Sawada, M, 1999) are displayed in figure 4. Large values of this statistic indicate high spatial correlation and this graph shows that there is significant spatial correlation in the residuals (ignore the lines labelled GWR and SEM for the time being).

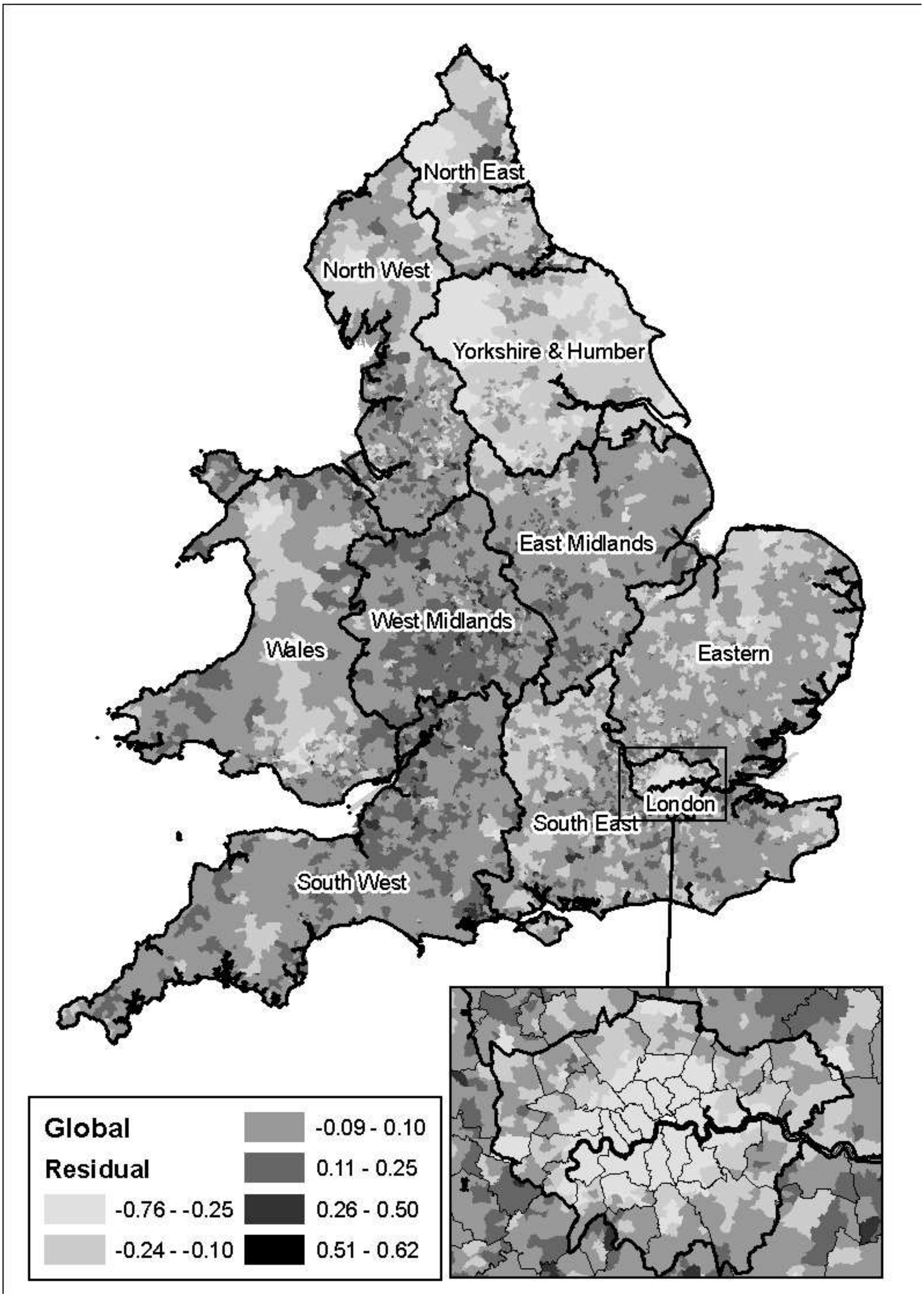

Figure 3 : Residuals from the global regression model

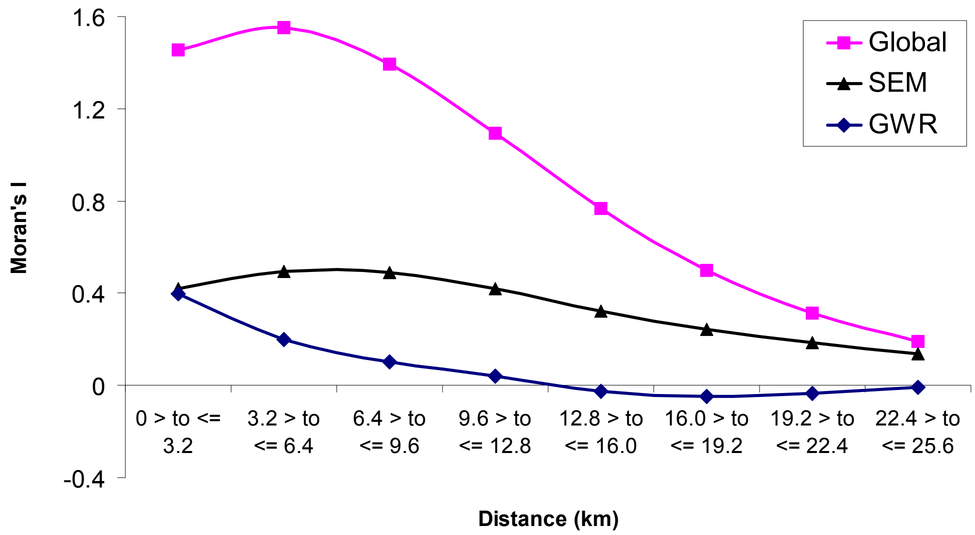

Figure 4 : Moran's I statistic for the residuals from the global, SEM and GWR models

The presence of such correlations is of concern for two reasons. Firstly it violates the assumption of randomly distributed and independent error terms in regression modelling. Any significance tests arising from the interpretation of this model are therefore suspect. Secondly the residuals clearly contain some geographic information which the global model is unable to incorporate in its formulation and is therefore lost.

A possible model enhancement is the use of Spatial Error Models (SEM) which extend the traditional regression formulation to include an explicit term to take account of the spatial structure in the regression's error term. Whilst this change deals with the statistical concern over the nature of the error term, it does not provide a measure of the strength of the relationship between variables at a local level. To see an illustration of the application of a SEM with these data, see Appendix 1.

One way to incorporate this geographical information at the local level is to introduce region specific dummy variables or regional interaction terms into the regression equation. Whilst this will go someway towards addressing the geographical issues the discrete nature of this approach is a little arbitrary. It does erroneously assume, for instance, that a ward in the North West, but on the border with the West Midlands, will have more in common with a ward to the far north of the North West, some 250 kilometres away, than a neighbouring West Midlands ward. A series of more local or overlapping dummy variables can help to minimise this impact but the question then arises as to how to define these areas.

4 Geographically Weighted Regression

In the global model reported above each observation is taken as an independent observation, contributing the same amount of information to the relationship at each data point. The data does, however, have a geographical context and this information is not used. In particular, there is no reason to believe that the strength of the income relationship will be the same across the whole of England and Wales. The global model is therefore a compromise that may not actually be appropriate for ANY area of England and Wales.

The technique of Geographical Weighted Regression (GWR) (Brunsdon, Fotheringham and Charlton, 1998) attempts to incorporate this geographical information into a regression model using a series of distance related weights. In effect, when estimating the local relationship between car ownership and income for a particular ward the corresponding information from neighbouring wards is given a higher weight than information from more distant wards. The outcome of this process is that rather than just having one estimate for the income parameter, each ward has its own estimate based on a specific set of observation weights. The model is summarised in the following equation.

$$y_i = \alpha_{(u_i,v_i)} + \sum_{j=1}^{k} \beta_{j,(u_i,v_i)}\, x_{j,i} + \varepsilon_i$$

Where y_i is the estimated car ownership in ward i;

$x_{j,i}$ is the value of explanatory parameter j for ward i;

$\alpha_{(ui,vi)}$ is the estimated intercept for ward i located at point (u_i,v_i);

$\beta_{j,(ui,vi)}$ is the estimated value for parameter j for ward i located at point (u_i,v_i);

k is the number of explanatory variables in the model;

ε_i is a randomly distributed error term, $\varepsilon_i \sim N(0,\sigma^2)$

In order to estimate the parameters in this model the following calculations are performed for each ward. Firstly the distances from the current ward to all the other wards is calculated and a decay function is applied to convert these distances into regression weights. This produces a 8 837 by 8 837 diagonal matrix. The regression equation is then estimated using these weights. The parameter estimates and ancillary statistics from this regression are taken as those for the current ward. The process then moves on by considering the next ward.

Initially such GWR models were used in a purely exploratory manner but recently attempts have been made to put the formulation, validation and interpretation of the models on a more robust statistical footing (Brunsdon, Fotheringham, and Charlton, 2000).

An initial scoping study along the lines of the national study reported here was conducted by the author (Clark, 2004). In this study only a limited geographical data set consisting of 220 wards was used in a GWR model but some interesting insights into local car ownership patterns were apparent. Other researchers' have also used GWR techniques in a transport context. Two studies have attempted to estimate the increase in local land values as a result of enhanced or differing transport accessibility, either by calculating land value up-lifts as a result of a new tram system in Croydon, south London (AWR et al, 2003) or by examining how the different patterns of accessibility in an urban area of northern England affect house prices (Mulley and Du, 2006). Two further studies have investigated how travel volumes and distances vary and if the use of GWR can provide additional insight – the first by modelling the relationship between commuting distance and socio-economic variables in Northern Ireland (Lloyd and Shuttleworth, 2005) and a second which attempts to discern patterns in a national Origin-Destination matrix for Japan (Nakaya, 2001). Other transportation studies using GWR include estimating the accident risk at traffic network locations in Toronto Canada (Hadayeghi et al, 2003) and explaining the demand for public transit ridership in Broward County, Florida, USA (Lee-Fang et al, 2006).

4.1 Fitting the GWR Model

The GWR model is fitted to the data using the be-spoke package, GWR3, produced at the National University of Ireland (2005). The package allows the use of a variety of calibration techniques to specify how quickly the regression weight diminishes as the distance from the estimation point increases and how to optimise a bandwidth parameter. For the model results reported below the adaptive kernel weighting function, where a constant number of neighbouring wards are used in the estimation process, is used in preference to the fixed distance kernel (which uses all wards within a constant distance). The rational behind this choice is that for rural wards the fixed approach would include too few neighbouring wards leading to an ill-defined estimation problem and for dense urban areas it would include too many wards and would potentially dilute the local nature of the relationship. The Akaike Information Criterion (corrected) (AIC_c) is used for selecting the optimal kernel size so that account is taken of the changes in the model specification – typically the number of effective parameters in the model.

4.2 Results

The GWR software package confirms the global model estimates provided in equation 1 and also calculates the global AIC_c as -7 088 with 3 parameters. The GWR model is fitted to the data, adapting the bandwidth of the Kernel to ensure that 454 observations (this is 5% of the total number of wards) are included in all the regression estimates. The AIC_c for the GWR model is -15 054 on 154.1 effective parameters which represents a large improvement in fit, even when the increase in the number of parameters is taken into account (the estimation of a fractional number of parameters is an unusual feature of GWR models – for an explanation of how this value is derived see appendix 2). The global R^2_{adj}

value has also increased from 75.3% to 90.1%. All the estimates for the income and population density parameters are significant at the 5% level and none of the standardised residuals are beyond ±3.0. Examination of the statistical distribution of the GWR parameter estimates and the results of a Monte-Carlo significance test suggests that there is spatial distribution in all three parameters. There is a strong linear correlation between the constant term and the income parameter, a not uncommon feature of GWR type models.

Figure 5 maps how the local R^2_{adj} value varies over England and Wales. The model fit is particularly good in the West Midlands Region (in excess of 90%) and poorest in central and south west London and the South East (as low as 62%).

The residuals from the GWR model are displayed in figure 6. Here the appearance is much more “dappled” than in figure 3 with no blocks of neighbouring wards showing a similar rendering. The values of Moran’s I statistic for these GWR residuals is shown in figure 4 and this shows that there is much less evidence of spatial correlation in the GWR model than for the global model and the SEM.

The residuals from the GWR model are displayed in figure 6. Here the appearance is much more “dappled” than in figure 3 with no blocks of neighbouring wards showing a similar rendering. The values of Moran’s I statistic for these GWR residuals is shown in figure 4 and this shows that there is much less evidence of spatial correlation in the GWR model than for the global model and the SEM.

The values for the estimated income parameter vary across England and Wales and are shown in Figure 7. The parameter is lowest in central and west London but highest in the West Midlands and areas on the North West and the North East. In central London the model outputs suggest that a high level of income is required to own a vehicle whilst in the West Midlands and the North West, high car ownership happens at much lower levels of income.

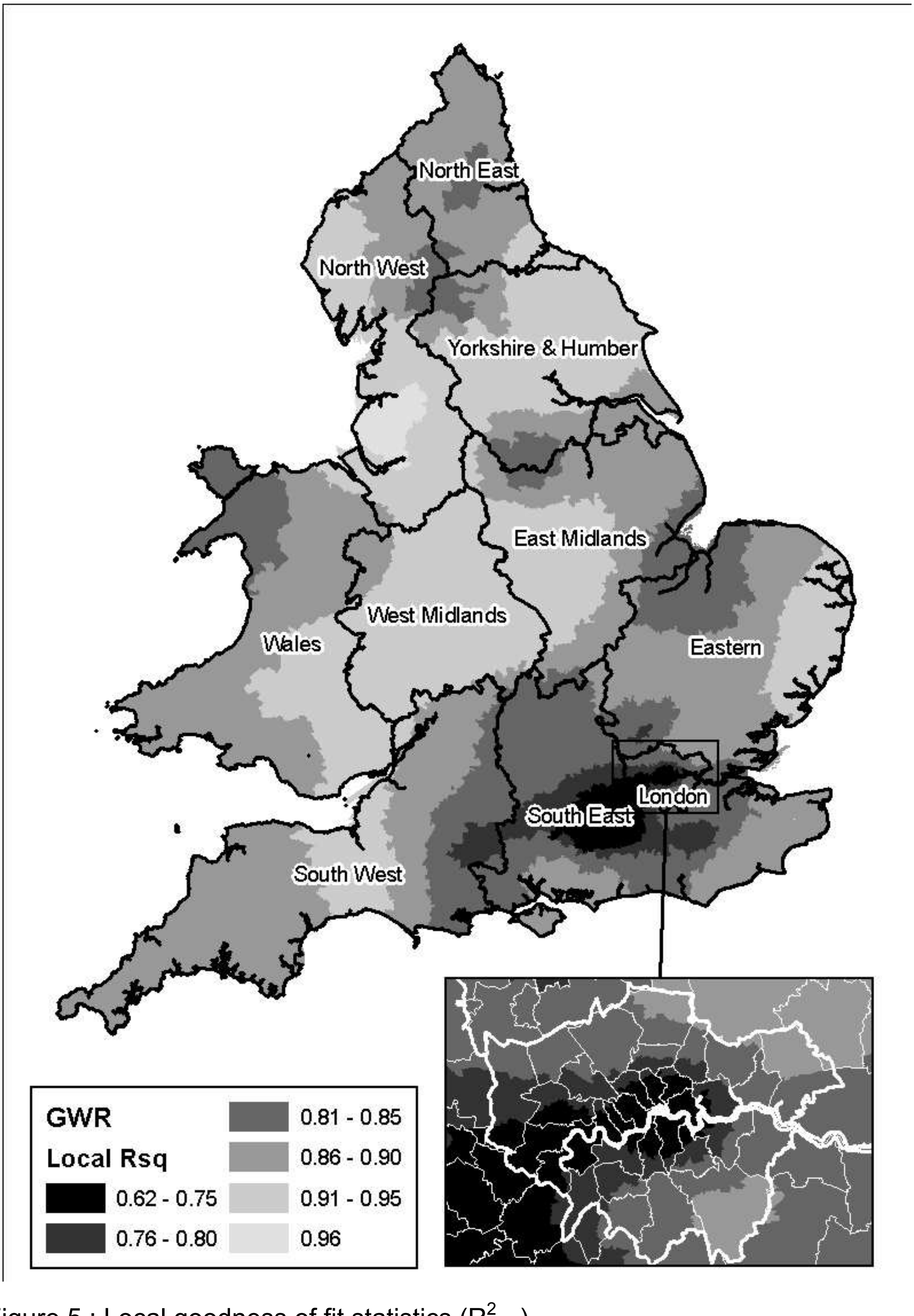

Figure 5 : Local goodness of fit statistics (R^2_{adj})

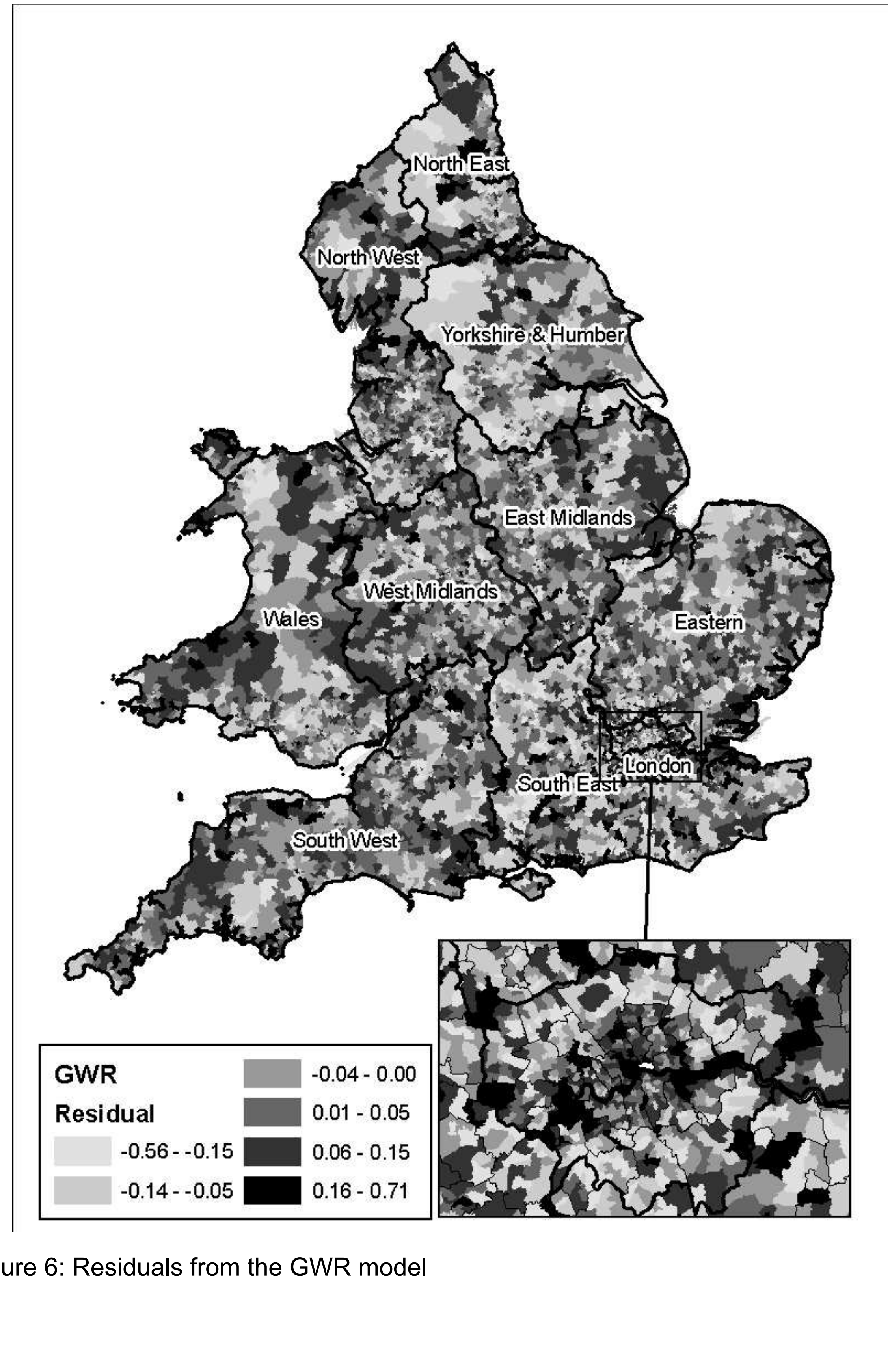

Figure 6: Residuals from the GWR model

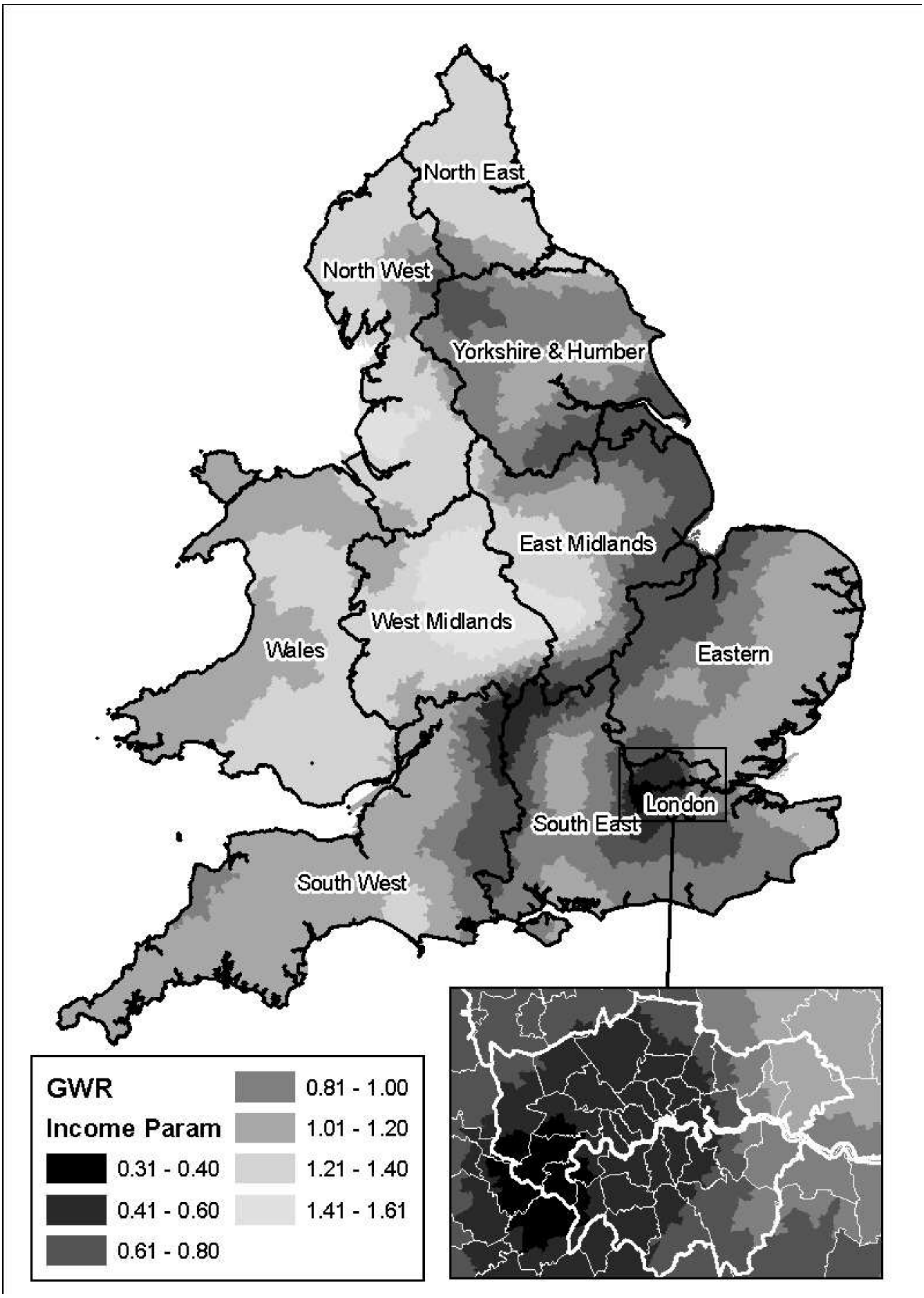

Figure 7 : Local model estimates for the income parameter

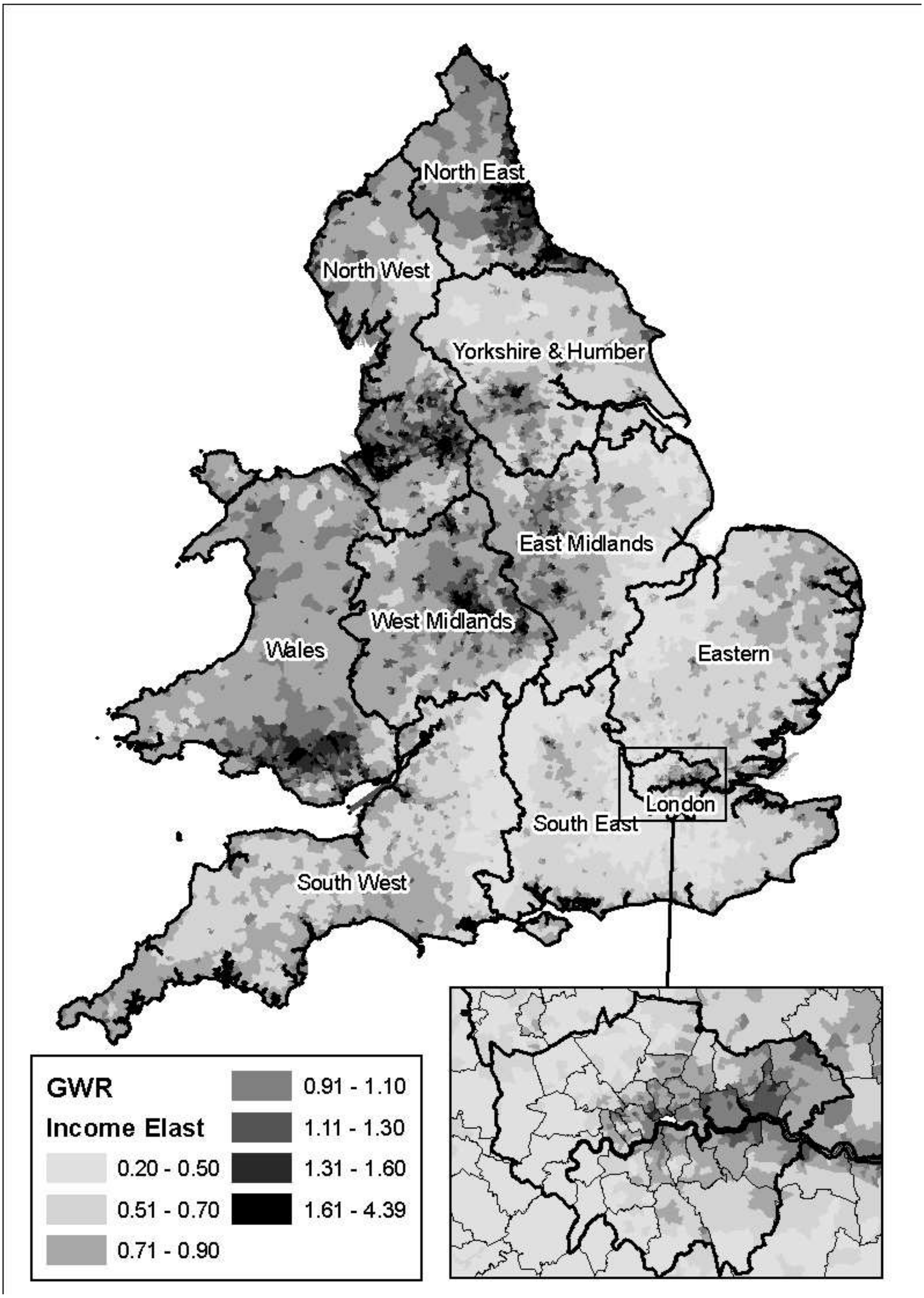

Figure 8 : Implied values for the elasticity of car ownership with respect to income

Figure 8 shows the variation in the implied elasticity of car ownership with respect to income. The median elasticity is 0.71 and the 95% interval is (0.345, 1.872). The elasticity is highest in the city centres of northern England, the West Midlands and south Wales. In these high elasticity areas car ownership levels are traditionally low (see figure 2) so this effect may be a desire to increase car ownership levels using a large proportion of available income. The elasticity is low in the South East of England suggesting that car ownership aspirations here are already largely satisfied (note that the car ownership levels are high here in figure 2) and that income is used for other household purposes.

To demonstrate the impact of using different elasticity estimates the change in car ownership of a 10% rise in real incomes is estimated. One scenario is to assume a constant global elasticity of $\eta_{co} = 0.608$, this assumes that the car ownership behaviour is uniform across the whole of England and Wales. The second scenario is to assume ward specific elasticities from $\eta_{co} = 0.744 / y_i$ which takes account of local circumstances but still assumes a uniformly estimated value for the strength of the relationship between car ownership and incomes. The third scenario is the most flexible, it uses the locally estimated GWR income parameters when estimating the elasticity, $\eta_{co} = \beta_{\text{Income }(ui,vi)} / y_i$. The predicted size of the car and van fleet under these three scenarios is show in table 1. On a total 2001 vehicle stock of 24 million cars and vans, the increase derived from the ward specific GWR elasticities is 781k more cars than that from the global elasticity.

Scenario		Number of cars and vans
2001 Census		23 933 690
Car and van fleet size after a 10% increase in real incomes	$\eta_{co} = 0.608$	25 388 858 (+ 6.1%)
	$\eta_{co} = 0.744 / y_i$	25 544 841 (+ 6.7%)
	$\eta_{co} = \beta_{(ui,vi)} / y_i$	26 169 943 (+ 9.3%)

Table 1 : Estimated number of cars and vans under three elasticity scenarios

5 Discussion

Other similar studies of the relationship between incomes and car ownership have only provided a rudimentary allowance for the possibility that the strength of the relationship will vary over such a wide and diverse area as the UK. The attempts to allow for some spatial variation have included introducing co-variate data such as population density (as this study has) or area specific dummy variables. These models will produce a reliable estimate for the income elasticity (in that the income term is freed from having to represent other features in the data) but they still produce a single income parameter and elasticity estimate. This GWR model has allowed for a truly spatially varying estimate of the income parameter and hence the calculation of a local income elasticity.

One potential concern is whether the GWR model is over-fitting the data. The increase in the effective number of parameters from just 3 in the global model to 155 in the GWR model does appear large. This increase needs to be placed in the context of the size of the whole data set consisting of nearly 8 900 observations, and there is therefore a 1 to 50 ratio between parameters and data. Even if the alternative (and arguably less appropriate) approach of using an additional dummy variable for each local authority district so as to modify just the intercept term in a regression is used, there would initially be more than 370 additional dummy variables in the global model. The analyst would then have the task of eliminating insignificant and combining together similar dummy variables. Even if the number of dummy variables were halved, there would still be more parameters than in the GWR model.

Another concern is that the variability in the parameters is really “picking-up” the existence of another important variable which has been omitted from the model. The strength of the relationship in the global model suggests that there is little need for additional variables in the model. Also, the range of GWR income parameter estimates is roughly what one might expect from such regression type models suggesting that no “outside influence” has caused them to be modified in some unreasonable manner.

Artificial edge effects were a concern in the scoping study for this paper (Clark, 2004). Unlike the region studied in the earlier article, England and Wales is a highly segregated landmass, largely bounded by coastlines, thereby enabling an effective and natural border to be drawn around the area of study. The only locations used in the model which are likely to be affected by artificial edge effects are those which border Scotland in the north. These bordering wards should technically use income and car ownership values from neighbouring Scottish wards but these have not been included in this model. This border, is however, quite short in relation to the entire England and Wales coastline.

In terms of further analysis, the approach outlined in this study could easily be applied to the other travel-related question asked in the 2001 Census, “Method of Travel to Work”. The mode share of each of the various modes (Key Statistic 15) or the distance travelled (Standard Table S121) could be related to income and other explanatory variables and any variability in this relationship mapped across England and Wales. Since travel to work is an individual response (rather than a household response) an equivalised income measure may be more appropriate than the net income used here. Initial explorations suggest that equivalised income has a negative exponential relationship with the proportion of people who travel to work as a passenger in a car or van. Beyond the subject area of transport, similar analysis could be conducted against the other question domains within the Census such as health, educational attainment and living conditions.

6 Summary

In establishing that incomes are a significant and influential factor in explaining levels of car ownership, this study is in line with the body of previously published evidence in this field. The estimated global value for the elasticity of car ownership with respect to income is, maybe, on the low side, but is within the range found in other studies.

The residuals from this global model were, however, found to contain spatial correlation which raised a number of concerns. Primary among these was that the local information or variability in the relationship was being ignored by the global regression model. This led to the development of a geographically weighted regression model that was able to capture the spatially varying nature of the relationship. An illustration of the use of the spatially varying elasticity estimates was demonstrated with a scenario of a 10% rise in real incomes.

This prediction exercise demonstrated that the use of global estimates is unlikely to predict local changes well. In areas of the country where this elasticity is low, eg the South East of England and parts of London, the estimates from the use of the global elasticity value will tend to over predict changes in car ownership. Similarly in areas where the elasticity is high, such as the urban areas of northern England, the global estimate will under predict car ownership. Even in aggregate, at a national level of England and Wales, these under and over estimates do not cancel out. For local planners who need to predict future car ownership levels, for example to plan future road and parking provision or forecast the use of public transport, the local GWR estimates will be more appropriate.

Acknowledgements

Any views expressed in this paper are those of the author only and should not be taken to be those of Leeds City Council or its Agencies. The author would like to thank the anonymous referees for their comments and observations on an earlier draft of this paper.

References

ARW, Geofutures, University College London, and Symonds Group (2003). Land and Property Value and Public Transport Investment — Mapping Changes in Property Values around the Croydon Tramlink. Report prepared for the Royal Institute of Chartered Surveyors, Office of the Deputy Prime Minister and Department for Transport, London, October.

Brunsdon, C, Fotheringham, S and Charlton, M (1998). Geographical Weighted Regression – Modelling Spatial Non-Stationarity. The Statistician, Vol 47, No 3, pp431-443.

Brunsdon, C, Fotheringham, S and Charlton, M (2000). Geographical Weighted Regression as a Statistical Model. Paper presented at AEA Meeting, Palm Springs, 2001.

Chow, LF, Zhao, F, Min-Tang, L, Liu, X and Ubaka, I (2006). Transit Ridership Model Based on Geographically Weighted Regression. Paper 06-2348 presented at the Transportation Research Board Annual Meeting, 2006.

Clark, SD (2004). Estimating Car Ownership using Geographically Weighted Regression. Traffic Engineering and Control, Vol 45, No 12, pp416-420.

Dargay, JM (2002). Determinants of Car Ownership in Rural and Urban areas: a Pseudo-Panel Analysis. Transportation Research Part E: Logistics and Transportation Review, Vol 38, No 5, pp 351-366.

de Jong, G, Fox, J, Pieters, M, Vonk, L, Daly, A (2004). Comparison of Car Ownership Models. Transport Reviews, Vol 24, No 4, pp379-408. See also, Audit of Car Ownership Models. http://www.rand.org/publications/MR/MR1664/

Department for Work and Pensions (2005). Family Resources Survey. http://www.dwp.gov.uk/asd/frs/

Du, H and Mulley, C (2006). Relationship Between Transport Accessibility and Land Value: a local model approach with Geographically Weighted Regression. Paper 06-2009 presented at the Transportation Research Board Annual Meeting, 2006.

Fearnley, N and Bekken, JT (2005). Short- and long-run demand effects in transport: A literature survey. TØI report no. 802/2005. Summary in English at http://www.toi.no/attach/a1274456r371471/Summary_802_05.pdf

Goldring, S, Longhurst, J and Cruddas, M (2005). Model-Based Estimates of Income for wards in England and Wales, 2001/02. Technical Report. United Kingdom Office for National Statistics.

Goodwin, P, Dargay, J and Hanly, M (2004). Elasticities of Road Traffic and Fuel Consumption with Respect to Price and Income: A Review. Transport Reviews, Vol 24, No 3, pp275-292. See also, Hanly, M, Dargay, J and Goodwin, P . Review of Income and Price Elasticities in the Demand for Road Traffic. Contract No PPAD 9/65/93. http://www.dft.gov.uk/stellent/groups/dft_econappr/documents/downloadable/dft_econappr_033848.pdf

Graham, D and Glaister, S (2004). Road Traffic Demand Elasticity Estimates: A Review. Transport Reviews, Vol 24, No 3, pp261-274. See also, Review of Income and Price Elasticities of Demand for Road Traffic. Contract No PPAD 9/65/93. http://www.dft.gov.uk/stellent/groups/dft_econappr/documents/downloadable/dft_econappr_033849.pdf

Hadayeghi, A, Shalaby, AS and Persaud, BN (2003). Macrolevel Accident Prediction Models for Evaluating Safety of Urban Transportation Systems. Transportation Research Record 1850, pp 87-95.

Ingram, GK and Liu, Z (1999). Determinants of Motorization and Road Provision. World Bank Policy Research Working paper 2042. . http://www.worldbank.org/html/dec/Publications/Workpapers/wps2000series/wps2042/wps2042.pdf

Karlaftis, M and Golias, J (2002). Automobile Ownership, Households Without Automobiles, and Urban Traffic Parameters: Are They Related. Transportation Research 1792, Transportation Research Board, pp. 29-35.

LeSage, JP (1998). Spatial Econometrics. Department of Economics, University of Toledo, USA. http://www.spatial-econometrics.com/html/wbook.pdf

Lloyd, C and Shuttleworth, I (2005). Analysing Commuting Using Local Regression Techniques: Scale, Sensitivity, and Geographical Patterning. Environment and Planning, Vol 37, pp81-103.

Longley, PA and Tobón, C (2004). Spatial Dependence and Heterogeneity in Patterns of Hardship. An Intra-Urban Analysis. Annals of the Association of American Geographers, Vol 93, No 3 pp503-519.

Moran, I (1950). Notes on the Continuous Stochastic Phenomena. Biometrika, Vol 37 pp17-23.

Nakaya, T (2001). Local Spatial Interaction Modelling based on the Geographically Weighted Regression Approach. GeoJournal, pp347-358.

National University of Ireland (2005). GWR Software Specification. http://www.nuim.ie/ncg/GWR/

R Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Sawada, M. 1999. ROOKCASE: An Excel 97/2000 Visual Basic (VB) Add-in for Exploring Global and Local Spatial Autocorrelation. Bulletin of the Ecological Society of America, Vol 80, No 4 pp231-234.

Todd Litman (2005). Transportation Elasticities. How Prices and Other Factors affect Travel Behaviour. Traffic Demand and Management Encyclopaedia. http://www.vtpi.org/elasticities.pdf

Williams, P (2000). Income Imputation for Small Areas. Project web site at: http://pcwww.liv.ac.uk/~william/income/index.html

Appendix 1 Fitting a Spatial Error Model

Spatial error models (SEM) are an intermediate type of model, in that they retain the global nature of the parameter estimates associated with the explanatory variable but extend the model form to explicitly account for the spatial structure in the residuals. The model form is:

$$\mathbf{Y} = \mathbf{X}\,\beta + \mathbf{u} \qquad \text{(A1a)}$$

$$\mathbf{u} = \lambda\,\mathbf{W}\,\mathbf{u} + \varepsilon \qquad \text{(A1b)}$$

Where **Y** is a n x 1 vector of dependant data

X is a n x k matrix of explanatory data

β is a k x 1 vector of parameter estimates

u is a n x 1 vector of spatially correlated disturbances

λ is a scalar parameter measuring the spatial correlation

W is a n x n matrix which captures the spatial congruity in the data

ε is a n x 1 vector such that $\varepsilon \sim N(0, \sigma^2 I_{nxn})$

The λ parameter is analogous to the moving average parameters typically used in autoregressive and moving average (ARMA) time series models.

The estimation of the parameters in this model is done using the public domain *R* package (2005) and the spatial dependence add in package *spdep*. As a preliminary to the fitting of the model the *tri2nb()* function is used to create a triangulation grid from the wards' centriod co-ordinates. This grid essentially provides the structure of the **W** matrix in equation (A1b).

The actual SEM model is fitted using the *Gmerrorsar()* function.

$$cars\ per\ household = \underset{(0.051)}{-4.804} + \underset{(0.008)}{1.016}\ln(net\ income) - \underset{(0.001)}{0.077}\ln(population\ density)$$

with the spatial correlation parameter $\lambda = 0.747$. The likelihood ratio test from this model is highly significant (p-value < 2.22e-16) thereby rejecting the null hypothesis that there is no spatial correlation in the global model (confirming the conclusions from figure 4). Also looking at figure 4, it is apparent that the values of Moran's I statistic are lower for the residuals from a SEM than from the global model. This suggests that a revised estimate of the car ownership elasticity is now $\eta_{co} = (1.016 / 1.223) = 0.831$, a higher value than with the global model, where $\eta_{co} = 0.608$. The AIC value for this SEM model is -15 101 with 5 parameters. This compares well with the AIC_c value of -15 054 with 154.1 parameters from the GWR model.

Table A1 provides estimates of the total UK car ownership under two scenarios consistent with those in table 1 but using SEM estimated parameters. The estimate derived from using the second scenario of a global SEM parameter estimate with ward car ownership levels (26.13m) is very close to that of the third scenario in table 1, which uses both ward specific GWR parameter estimates and car ownership (26.17m). Whilst this similarity at a national level is remarkable, the SEM model does still not provide the best estimate of car ownership growth at the local ward level since it fails to estimate the strength of the relationship at a local level.

Appendix 2 Estimating the number of parameters in a GWR model

The goodness of fit of a statistical model should always be put in the context of the number of parameters used in its estimation. With GWR models each data point has its own estimate of each of the parameters in the model, so in theory the number of parameters could be n(k+1), where n is the number of observations - a highly over-parameterised model. There is, however, a relationship between neighbouring parameters (clearly visible in figure 7) reducing the freedom that the model has to estimate these parameters. In Brunsdon, et al (2000) an argument is proposed that allows an estimate to be made of the effective number of parameters in a GWR model. This argument is summarised here.

In the global regression model the expression for the expected values of the residual sum of squares (RSS) is given by

$$E[RSS] = (n - DoF)\,\sigma^2 \qquad \text{(A2a)}$$

Where DoF is the degrees of freedom within the model, equal to (k+1) (that is, k variables plus an intercept); and

and σ^2 is the variance of the error term.

In their paper Brunsdon propose the existence of an (n+1) by (n+1) transformation matrix, **S**, that transforms the observed y_i's to the fitted $\tilde{y}_i$'s

$$\tilde{\mathbf{y}} = \mathbf{S}\,\mathbf{y}$$

The fitted residuals are thus, $(\mathbf{I} - \mathbf{S})\,\mathbf{y}$, and

$$RSS = \mathbf{y}'\,(\mathbf{I} - \mathbf{S})\,(\mathbf{I} - \mathbf{S})\,\mathbf{y}$$

Using published results and assumptions with regard to the small size of a bias term when we have a large sample, the expectation of the RSS in the case of GWR equates to

E[RSS] = (n – (2 tr(**S**) – tr(**S**'**S**))) σ^2

In many cases, tr(**S**'**S**) is close to tr(**S**), thereby reducing the expected residual sum of squares to

E[RSS] = (n – tr(**S**)) σ^2 (A2b)

Comparing (A2b) and (A2a) suggests that the degrees of freedom, and hence number of effective parameters in a GWR model is given by the trace of the **S** matrix. This can clearly be an unorthodox, non-integer, value but it does give an indication of how heavily parameterised the model is.

This value of tr(**S**) is also used in the calculation of a GWR version of the AIC statistic:

$$AIC_c = 2n\log(\tilde{\sigma}) + \frac{n + tr(S)}{n + 2 - tr(S)}$$

Scenario		Number of cars and vans
2001 Census		23 933 690
Car and van fleet size after a 10% increase in real incomes	η_{co} = 0.831	25 921 967 (+ 8.3%)
	η_{co} = 1.016 / y_i	26 133 864 (+ 9.2%)

Table A1 : Estimated number of cars and vans using SEM estimates