Centre for Transport Studies

IMPERIAL COLLEGE UNIVERSITY COLLEGE

MSc in Transport and Sustainable Development

RELATIONSHIP BETWEEN LAND USE MIX AND TRAVEL DISTANCE.

Paula Ojea Fernández-Colmeiro

15th September 2010

A Special Study submitted as part of the requirements for
the degree of Master of Science of Imperial College
and the University of London

Abstract

In this study a disaggregate analysis of the relationship between land use mix and travel behaviour in Cherwell District, Oxfordshire, is performed. Data from a 7 day travel survey conducted in 2001 by TNS was used, and land use mix measures were calculated in ArcGIS with Address Layer 2 Ordnance Survey Data. Distance travelled per individual per week for non-work purposes is estimated using multivariate linear regression analysis, being the socio-demographic characteristics at the household and individual levels the control variables. Two models are estimated, one for distance travelled by drivers of private vehicles and another for distance travelled by bus or coach. The land use variables employed were Hansen type accessibility measures for different land uses, the entropy based Shannon index and diversity indices measuring the balance of housing, office, retail and other social land uses. The results show that even though the levels of explanation were moderate (R^2=0.14), their effect on distance travelled by private car were proportionally higher than the socio-demographic factors. This didn't apply to the bus/coach model, for which the land use mix variable was not significant. The results imply that if the degree of land use mix increases by 10%, private vehicle distances travelled would be reduced by 3.35%, with the consequent reduction in transport energy consumption.

Acknowledgements

I am deeply thankful to my supervisor Dr Helena Titheridge for her guidance and help and to the Fundación Pedro Barrié de la Maza for their financial support. I would also like to thank all of those who helped me in any respect during the completion of the project, particularly Fer and Niovi.

Table of contents

List of Tables

List of Figures

1 Introduction

The concept of sustainable mobility is gaining interest due to many factors. Global issues such as climate change and fossil fuel scarcity are urging the need to reduce energy consumption and CO_2 emissions. The transport sector contributes to a substantial proportion of the energy consumed in the UK, being responsible for 34% of total energy consumption (DTI, 2002), and contributes to global warming by producing 29% of total CO_2 emissions due to the use of fossil fuels (Banister, 2005).
Land use distribution and transport are two important factors that have the potential to significantly alter energy consumption and enable more sustainable mobility. Even though the existence of a relationship between land use mix and travel is intuitive, its nature and extent is a matter of controversy. It is thus important to have a better understanding of the relationship existing between them and how they influence energy demand.
In this study the relationship between distance travelled and land use mix in Cherwell District is explored, aiming to help enlighten the existence or not of a link between travel behaviour and land use mix, and to contribute to the debate on whether or not there is a significant contribution of the mixing of land uses to sustainable mobility.

The study is structures as follows. First, the literature concerning the relation between land use, particularly land use mix, and travel is reviewed in Section 2. Section 3 includes the description of the data used for the analysis, whereas the methodology is explained in Section 4. The results, presented in Section 5, are followed by a brief explanation of their consequences on energy consumption. Finally, Section 7 includes the concluding remarks.

2 Background

2.1 Land use and travel patterns

There are many characteristics of the built environment that are presumed to impact personal travel, including population density, street connectivity, proximity to the urban centre, land use mix, settlement size, etc.

Among all these land use factors affecting travel behaviour, a great amount of research has been done in order to find the impact of residential density on transport energy consumption. The first analysis was done by Newman and Kenworthy (1989). They used data from 32 cities in the world and calculated correlations between fuel consumption and density variables obtaining a strong negative relationship. The key criticisms of their work included the method of analysis, which didn't include a multivariate analysis and the consequent lack of control for variables affecting fuel consumption such as fuel price and income (Gómez-Ibáñez, 1991). Mindali *et al* (2004) used the same data to test the effect of total urban density and land use mix on fuel consumption and found that there is no direct impact of total urban density, however several other relationships between energy use and land-use patterns exist.

In terms of land use mix, the literature is inconclusive about the extent to which it can alter travel behaviour (Banister, 2005). In their review of studies on urban form and travel patterns, Stead and Marshall (2001) found that there is a relatively small number of studies analysing the impact of land use mix, and that their results apparently are contradictory. But they suggest this might be due to the fact that different measures are used in the models, particularly for travel patterns, which have been measured in terms of distance travelled, travel time, modal share and energy consumption.

Crane and Crepeau (1998) developed an analysis of household travel diary and GIS data for San Diego and concluded that there is little role for land use in explaining travel behaviour and could not demonstrate the environmental benefits of urban design. Cervero and Kockleman (1997) used data from 50 neighbourhoods in the San Francisco Bay Area. The aggregate measure of land use mix they employed, named diversity, was found to be linked to non-work travel distance but not to total distance travelled.

In Van and Senior's (2000) empirical study, travel behaviour in 3 neighbourhoods of Cardiff with different degrees of land use mix were compared to conclude that mixed land uses discourage car use only for light shopping trips, but have no influence in commuting patterns. They agree with the way forward Handy (1996) suggests, of including in disaggregate analysis a combination of a variety of measures of urban form and different aspects of travel that could be affected by them.

Stead and Marshall (2001) criticised the accuracy of the data used by a number of studies. They mention the inaccuracy of trip distances calculated from trip zone data, assuming centroid to centroid distances.

In their critique of the empirical methods of analysis, Stead and Marshall (2001) also highlight the difficulties of holding socio-economic variables constant in order to identify a link between land uses and travel patterns due to the usual association between land use factors and socio-economic factors. In the studies they reviewed, they found two main ways to sort this out: multiple regression analysis and case study areas with similar socio-economic characteristics (however, these imply the strong assumption that the differences found in travel patterns are due to land use differences). One of the conclusions drawn from their study is that the definition of unambiguous measures capturing different aspects of land use might have an increased explanatory power for travel demand, and they also highlight the importance of differentiating between the effects of land use and socio-economic characteristics on travel patterns.

Ewing *et al* (1994) analysed six communities with different land use patterns by eliminating from the data the households with very low or very high incomes to control for socio-economic variation. They found out that travel differences between communities were significant, but smaller than expected.

To represent travel behaviour many different measures have been chosen. Authors such as Cervero and Kockelman (1997) and Boarnet and Sarmiento (2004) predicted personal vehicle miles travelled. In his study about the various measures of passenger travel patterns in Britain and transport emissions, Stead (1999) chose travel distance as the dependent variable because it is a reasonable representation for environmental impacts of transport including emissions and energy consumption. Other studies (e.g. Dargay and Hanly, 2004; Van and Senior, 2000; Cervero and Kockelman, 1997) predicted the probability of choosing private car or public transport for a trip.

Handy (1996) has shown that the land use mix has a negative effect on car use but also emphasized the complexity of this finding.

Vance and Hegel's (2007) results from a disaggregate two part model involving ordinary least square estimators using travel diary data collected in Germany between 1996 and 2003 suggest that urban form variables, including mixed use, density, public transit provision, etc plus instrumental variables to control for potential endogeneity, suggest that urban form has a causative impact on car use. Among the recommendations they make for further research, is the inclusion spatial lag variables, namely measures of land use patterns in rings around each household. This is meant to weight the influence of the built environment on the household depending on how distant the feature is from the household.

As explained Van and Senior (2000), the contribution of land use mix on the reduction of travel is based on the hypothesis that mixed land uses help to reduce probability of using the car, increase the frequency of short trips, thus reducing long distance shopping, and encourage lower levels of car ownership.

2.2 Modelling the relationship between land use and travel

Many types of analysis have been used in the literature to explore the relationship between the different characteristics of land use and travel patterns. Examples are logistic regression to predict mode choice (Dargay and Hanly, 2004; Badland *et al*, 2008), Poisson regression to analyse the number of cars owned and frequency of trips (Van and Senior, 2000), or probit regression to study the propensity of an individual to choose a given mode of transport (Crane and Crepeau, 1998; Schwanen and Mokhtarian, 2005b). Many studies used OLS regression, at different levels of aggregation, to explore travel patterns represented by continuous variables such as distances travelled per trip or per person or household.

It is in the interests of this study to be able to translate the results into impact on energy consumption, for which distance travelled per individual per week is a good proxy[1].

---

[1] A more detailed explanation of this decision is provided in section 4.

Hence, this section will be exploring some of the OLS regression analysis performed in the literature to gain insight on best practices.

In their aggregate international comparative analysis of relationships between car ownership, daily travel and urban form, Giuliano and Dargay (2006) employed two measures of land use: metropolitan size and population density to predict the total amount of daily travel per individual by an OLS regression model. The resulting model had a low level of explanation (15%) but indicated that distance travelled was inversely related to residential density, and that a high proportion of the variation between the US and UK was due to country characteristics not included in the model.

Souche (2010) estimated two urban travel demand models using an econometric method, both for car demand and public transport demand to make international comparisons for 100 of the world's cities. The explanatory variables she used include cost of travel by car and public transport, length of the roads, public transport vehicle kilometres, income and urban structure (derived from urban density). She used several regression methods with logarithmic relationship between travel demand and the explanatory variables that helped avoid heteroskedasticity, and the results showed that the coefficients obtained from more sophisticated methods didn't differ significantly from the ones obtained with the OLS model.

Also with aggregate city level data, Karathodorou *et al.* (2010) estimated models for car ownership per capita, fuel consumption per km and annual distance driven per car, using both ordinary least square (OLS) and seemingly unrelated regression equations (SURE) (to avoid correlation problems). However, they couldn't reject the hypothesis that there existed no correlation across equations by performing a Lagrangian multiplier test, meaning that OLS might have been be more efficient than SURE.

Stead (2001) acknowledges that land use characteristics are often associated with socioeconomic factors, increasing the difficulty in establishing the effects of land use on travel patterns. As stated in his paper, the two main ways in which literature has attempted to hold them constant are multiple regression analysis and the selection of areas with similar demographics but different land use patterns. For his study the NTS and two Local Authority travel surveys were examined. The dependent variable used in the stepwise multiple regression analysis was travel distance for being a reasonable representative for many impacts of transport such as energy consumption.

Cervero and Kockelman (1997) used multiple regression analysis to predict personal VMT and binomial logit analysis to model the probability of a person travelling by car. They controlled for car ownership. The hypotheses addressed in their paper where that the built environment has a negative influence on vehicle trip rates and a positive influence on non single occupancy vehicles and on non-personal vehicle mode choice both for work and non-work trips. They employed factor analysis for combining collinear variables to reveal the relative contributions of different attributes of the built environment in explaining travel demand. They concluded that the effects of the built environment on travel were moderate, and showed a stronger influence on non-work travel than commuting trips.

One of the conclusions of Dargay and Hanly's (2004) analysis is that their results from the multinomial logit model predicting mode choice didn't differ much from those based on single equation OLS.

As distance travelled is a continuous variable, OLS is a good approach that has been yielding good results in the literature when compared to more sophisticated models.

In the next sections some hints extracted from these studies will be taken into consideration during the modelling process.

3 Data

The area of analysis is Cherwell District, belonging to the County of Oxfordshire in the South East of England. According to the Oxfordshire Data Observatory (2010) and the 2001 Census, Cherwell District has an area of 587 square kilometres and a population of 131785 inhabitants living in 53268 households, and based in three major settlements: Banbury, Bicester, and Kidlington, with populations of 43867, 28672 and 13719 respectively.

3.1 Travel survey

The travel data used in this study is a travel survey in the Cherwell District commissioned by the North Oxfordshire consortium (NOC) to Taylor Nelson Sofres (TNS) in 2000. The aim was to gain understanding of the movement of people in this area.

As stated in the travel survey report (TNS, 2001), the survey focused on the collection of information regarding where people were travelling to and from, the modes of transport, the main journey purposes, the distances travelled and the times of journeys.

The survey comprised a household interview and a travel diary. A total of 1805 households in Cherwell District took part in the survey, with the main interviewing areas being the urban centres of Banbury, Bicester and Kidlington. Around 80% of the households returned at least one travel survey.

One copy of the Travel Diary was provided for each household member over 4 years old at each household. The data recorded details including journey start location and time, purpose, completion location and time, mode of travel and distance travel for every trip over a period of 7 days.

The Household Interview consisted on an interview with an adult representative of the household, collecting information including age, sex and working status of all household members, vehicle ownership, and walk and cycle trips made the previous day.

As explained in TNS's (2001) report, a random sample of addresses was chosen across the Cherwell District, using the Post Office Address File. In order to determine if a household was suitable to participate in the survey, telephone surveys were performed collecting socio-demographic information regarding age, gender and working status of every member of the household as well as car ownership and number of people in the household with driver license.

The travel surveys took place in two waves, one around November 2000 and the other around April 2001, because of a rail disruption until around February 2001. Even though the second wave was not affected by the rail disruptions, some people might have been affected by the movement restrictions due to the outbreak of the Foot-and-Mouth Disease in the United Kingdom in 2001. However, these disruptions are thought to have a negligible effect in the survey.

The data used in this analysis corresponds to the households where the address was successfully geocoded, a total of 1397.

3.2 Sample characteristics versus population characteristics

The socio demographic factors of the selected households were checked against the 2001 Census (ONS, 2001) corresponding to Cherwell District. The characteristics compared were the percentages of individuals per gender, age group and employment status category; percentages of households with each household size, structure and vehicle ownership.

Figure 1 shows very similar percentages of males and females between the sample and the 2001 Census records, with differences of only 0.1%.

In Figure 2 it is observed that for the selected households, the percentage of people in each age group doesn't differ from the census data more than 2.5%. Even though the difference is not very substantial, it could be explained by the fact that adults in non pensionable age are more likely not to be home when the telephone surveys were performed that adults in pensionable age.

Figure 3 shows the percentages of people in each employment status category. As expected, proportionally, there is less people working full time in the survey sample than in the census, but there is a higher percentage of retired people in the sample than in the census. An explanation for this could be that full time workers have less spare time than retired people hence they are less likely to complete a travel survey. The "Looking after home or relative" category in the census data could not be matched with any of the categories identified by the survey questionnaire. But the 4.4% of people falling in this category could correspond to the "not working" or "retired" categories in the survey sample, hence reducing the difference between the sample and the total population.

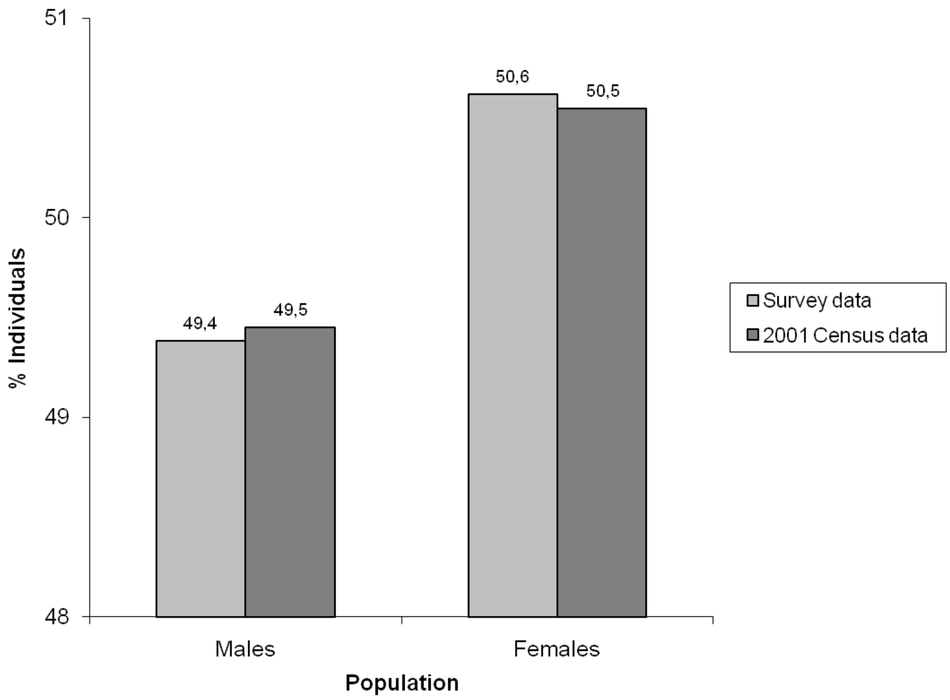


Figure 1. Comparison between Survey and Census data of population per gender.
Census output is Crown copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland. Source: 2001 Census, Output Area Boundaries. Crown copyright 2003.

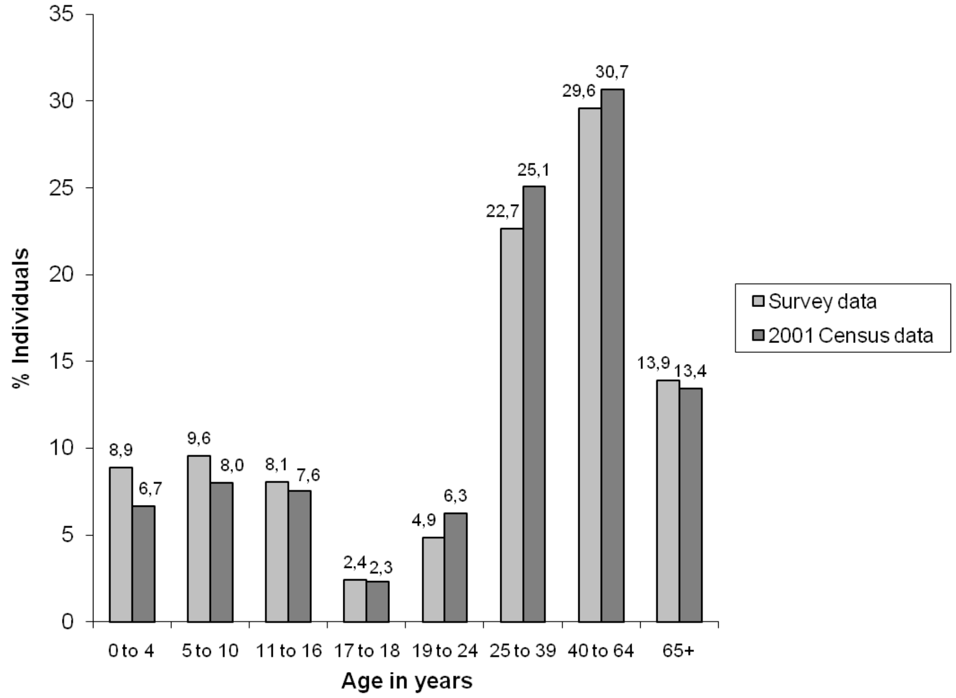


Figure 2. Comparison between Survey and Census data of population per age group.
Census output is Crown copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland. Source: 2001 Census, Output Area Boundaries. Crown copyright 2003.

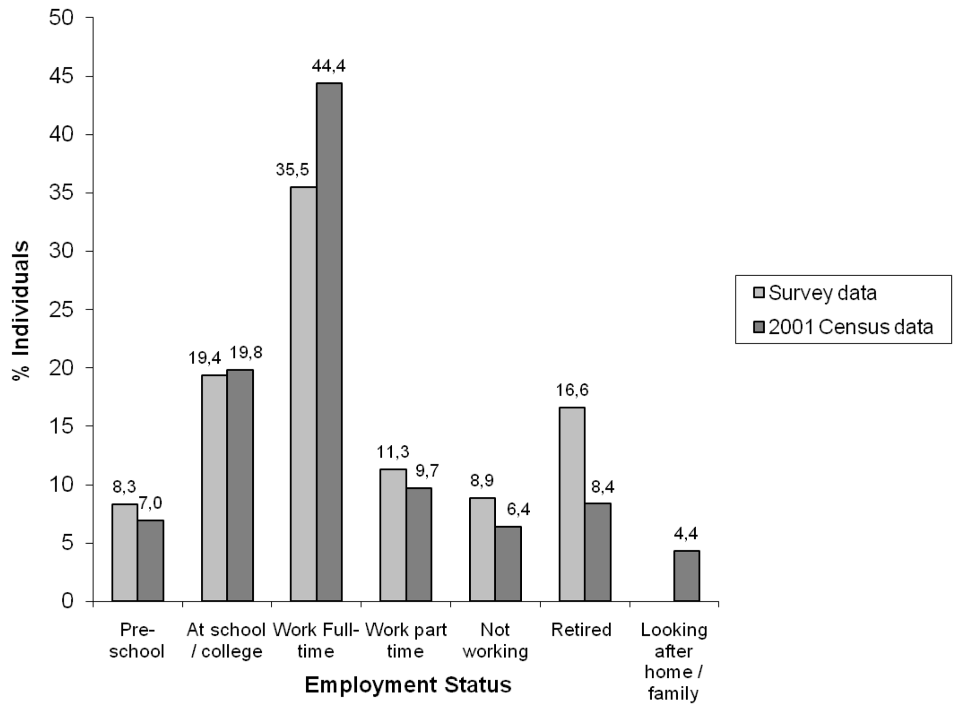


Figure 3. Comparison between Survey and Census data of employment status.
Census output is Crown copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland. Source: 2001 Census, Output Area Boundaries. Crown copyright 2003.

Figure 4 shows the comparison between the percentages of households with each household structure. In order to compare the household structure classifications, the census data table used was the Univariate Household Composition (alternative classification) (ONS, 2004) because its categories could be better matched with the ones obtained from the sample survey given the questionnaire questions regarding age. However, there still were some mismatches: in the 2001 census the term adult is used to refer to any person aged 16 and over, and children to individuals aged 15 and under. In the questionnaire, the age ranges are 11-16 and 17-18 so those with 16 years of age don't fall within the same category. Therefore, more children should be expected in the sample than in the census. However, the HH categories including children are less represented in the survey than the ones with children, maybe because people in households with children may be less likely to participate in the survey. Something similar happens with the definition of "pensionable age". According to the ONS (2004) Report, pensionable age is 65 and over for males and 60 and over for females. This causes a mismatch with the sample. affecting the "one adult at pensionable age" category comparison, since the survey sample includes males and females aged 65 or more, whereas the census additionally includes females aged 60 to 64. Thus in this category a larger percentage of people should be expected in the census than in the sample, but this is counterbalanced by the fact that pensionists are more likely to return travel diaries, as mentioned before, which gives a reasonable difference of 4.6%.

Figure 5 shows that household sizes closely match for all categories but the "1 person" category, where there are 8.9% less households in the sample. Also, the categories with 3 or more household members are over-represented in the sample. The reasoning behind this could be that, when the telephone survey is done, it is more likely that someone answers the phone in the households where more people live.

Finally, Figure 6 shows that the sample represents accurately the vehicle ownership levels registered in the census.

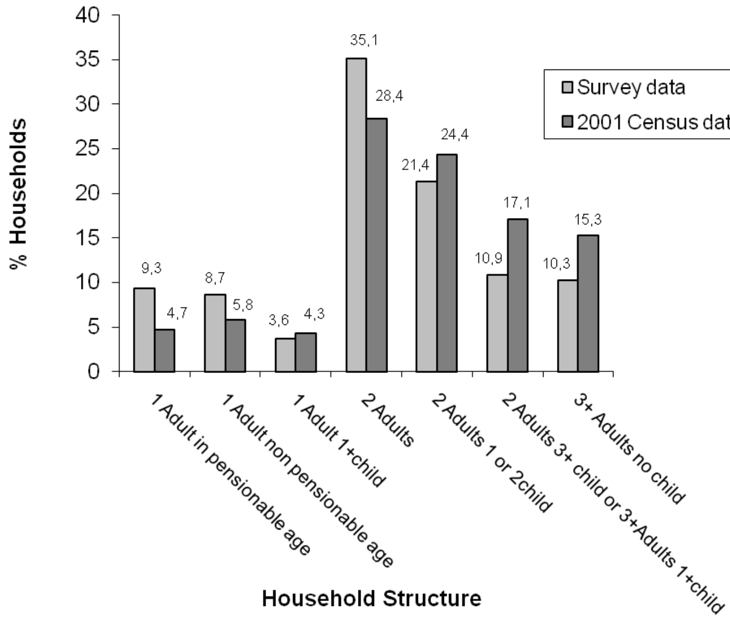


Figure 4. Comparison between Survey and Census data of household structure.
Census output is Crown copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland. Source: 2001 Census, Output Area Boundaries. Crown copyright 2003.

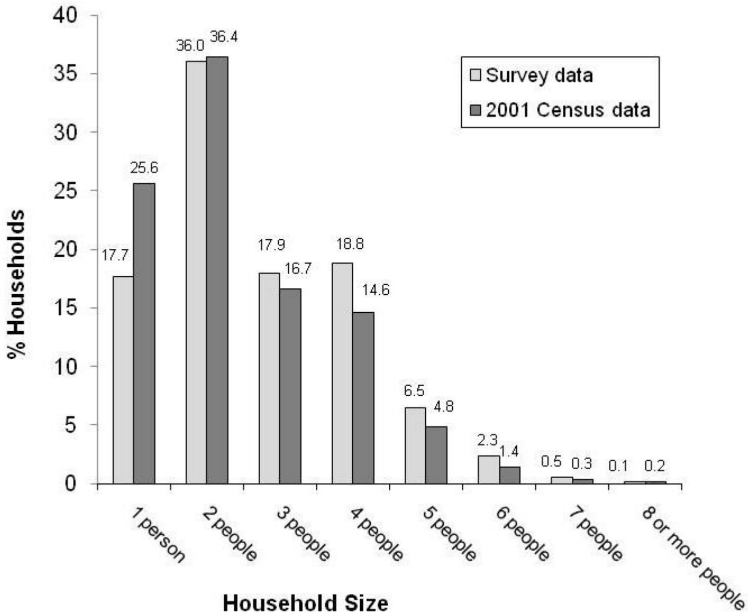


Figure 5. Comparison between Survey and Census data of household size.
Census output is Crown copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland. Source: 2001 Census, Output Area Boundaries. Crown copyright 2003.

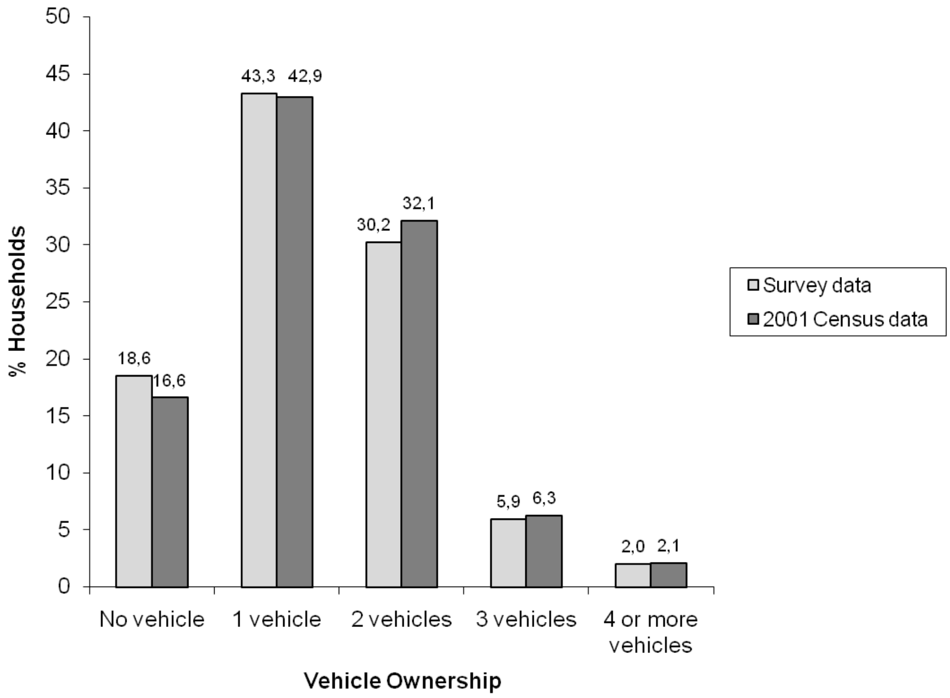


Figure 6. Comparison between Survey and Census vehicle ownership.
Census output is Crown copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland. Source: 2001 Census, Output Area Boundaries. Crown copyright 2003.

3.3 Ordnance Survey

The land use data was obtained from the Ordnance Survey. Specifically, the interest lies in the Address Layer 2 because of the spatial information regarding land uses. The Address Layer 2 contains for each address the NLUD (National Land Use Database) classification, plus the Ordnance Survey's field surveyor's allocation (OS Base function), which provides more detail than the NLUD. Additionally, the Topographic Area Layer was employed in order to measure the area of the buildings containing Address Layer 2 points.

The data was obtained in GML format and then converted into Geodatabase, readable by ArcGIS, the geographical information system software used for the study.

4 Methodology

The aim of the present study is to find out if energy consumption in transport is higher in areas where land uses are less mixed.

This hypothesis will be tested by measuring the influence of land use mix measures on distance travelled per individual per week, with two separate models, one for private vehicle travel (considering only distance travelled by vehicle drivers and not passengers, to avoid double counting vehicle distance travelled) and one for public transport distance travelled. Measures of travel such as trip frequency and distance travelled per trip would also be appropriate to calculate energy consumption. But for this study, the travel patterns will be measured in terms of the distance travelled per individual per week since, as explained by many researchers (e.g. Stead (1999b), cited in Stead, 2001, Boarnet *et al*, 2004), is a reasonable representation for transport energy consumption, and simplifies the problem by only needing to estimate one model for each mode.

4.1 Model Specification

Following previous research on this field, taking into account the disaggregate nature of this study, and given the dependent variable, distance travelled per individual per week, is continuous, multiple linear regression is used in this study.

The regression method used is Ordinary Least Square Regression (OLS).

$$y_i = \sum_{j=0}^{m} \beta_j x_{ij} + \varepsilon_i \tag{1}$$

Where:

y_i denotes the observation i on the dependent variable y

x_{ij} denotes the observation i on the j^{th} independent variable, with $x_{i0} = 1$.

β_j denotes the j^{th} parameter $0 \leq j \leq m$

ε_i denotes the error associated with observation i.

The definition of this model implies that both the observations y_i and the *independent variables* x_{ij} are independent normal variables.

Defining a normal error regression model means that regarding the error term ε_i, the assumptions that it is independently and identically distributed according to a normal distribution [$\varepsilon_i \sim iid\ N(0, \sigma_\varepsilon^2)$] are made. Hence, it needs to be checked whether the data available verifies these hypotheses, and whether or not is necessary to apply remedial measures.

To examine which ones among all the available potential explanators are needed in the model, a Stepwise regression method has been used, as in many studies (e.g. Stead, 1999a; Thideridge and Hall, 2006; Banister *et al*, 1997). This method develops a sequence of regression models, at each step adding or deleting an explanatory variable. The criterion for adding or deleting a predictor is stated in this study in terms of the F-statistic, choosing a value of p=0.05 for entry and p=0.1 for removal. This method also allows observing how stable the coefficients β_j of the predictors are through the estimation process.

4.2 Dependent Variables

In order to measure the travel patterns, trip distances were computed from the locations of each trip origin and destination given by the individuals. These locations were geocoded so that the total distance travelled per trip could be calculated in ArcGIS. The final value of the variable is the value of the estimated distance through the strategic road network plus the distances from the origin and destination points to the strategic road network as straight lines.

Most of the literature in this field explored commuting patterns. However non-work trips are more affected by land use characteristics than commuting trips (e.g. Giuliano and Dargay, 2006). This is because there is a limited choice on where to go to work or study, whereas for shopping, socialising and so on people are usually able to choose among different alternative destinations. As Crane and Crepeau (1998) suggest in their study, "it is the non-work trip that generates a majority of the local area travel" (p.228), and therefore is more likely to be influenced by the local land use mix.

Also, Erwing and Cervero (2002) indicated that land use variables exert a more significant influence on trip length than socio-demographic variables; however the latter are more significant when predicting trip frequency and mode choice. Following these insights, and the aim of exploring the less researched travel patterns, only distance travelled for non-work trips was taken into account.

The distinction of work and non-work trips was done according to the categories included in the questionnaire as answers to the question *Where were you going on this stage?* (TNS, 2001). Non-work trips include the following categories:

- Shopping/personal business (visit doctor, bank etc.)
- Visiting friends/family
- Other social (eating/drinking out, cinema)
- Taking/collecting someone (shopping / personal business)

The category "going home" was not included because it represented a very small percentage of the total number of trips (2.4%). Respondents were asked to choose it only when no other purpose applied, and it could not easily be recoded into a work or non-work category.

To determine the impact of land use mix on travel behaviour, the study will be comprised of two models, one analysing private car trips and another one analysing public transport trips. As it can be observed in Table 1, trips made by public transport are mainly bus or coach trips, because train trips only account for 0.4% of the number of non-work trips. Also, a single model containing both trips by bus/coach and train would not be appropriate to eventually estimate the effects on energy. Therefore, and given the small number of train trips, only the bus/coach trips were modelled.

	non-work trips	
	number of trips	percentage
Car / van / motorbike - (as driver)	12640	57.9%
Car / van / motorbike - (as passenger)	7229	33.1%
Bus / coach	1094	5.0%
Train	87	0.4%
Taxi/Minicab	240	1.1%
Bicycle	476	2.2%
Other	60	0.3%
TOTAL	21826	100%

Table 1. Number and percentages of non-work trips per mode.

For the model predicting private vehicle distance travelled, only trips made by private car as driver were analysed. Because the interest of the study lies in vehicle distance travelled to represent transport energy consumption, trips made by car as passengers are excluded to avoid double counting vehicle distance travelled.

After selecting all individuals who performed trips as vehicle drivers, it was observed that 5 respondents among the total of 1298 were individuals between 11 and 16 years of age. This small portion of the data was not taken into account for modelling. The cases probably correspond to people aged 16 years (since that is the minimum legal age in the UK) who drive a moped, vehicle with a level of energy consumption substantially lower than the most typically owned car, so the effect would not make a big difference in the results, and the bias of including extreme cases is avoided.

On the basis of the analysis of the distributions of each of the variables log transformations were used so that the dependent variable’s probability distributions were similar to normal and problems of heteroskedasticity were avoided. Histograms in Figure 7 show how this transformation affects the probability distributions.

As observed in Figure 7(a), the distribution of the variable is skewed; hence to make it more normally distributed natural logarithm transformations were applied, resulting in the histogram Figure 7(b), were the skewness is greatly reduced.

The same transformation was done to the non-work distance travelled by bus/coach and both distributions are represented in Figure 7(c) and (d).

In Table 2 the name and description of the dependent variables is presented, and Table 3 contains their descriptive statistics. It is noted that there is a good deal of variability in both values.

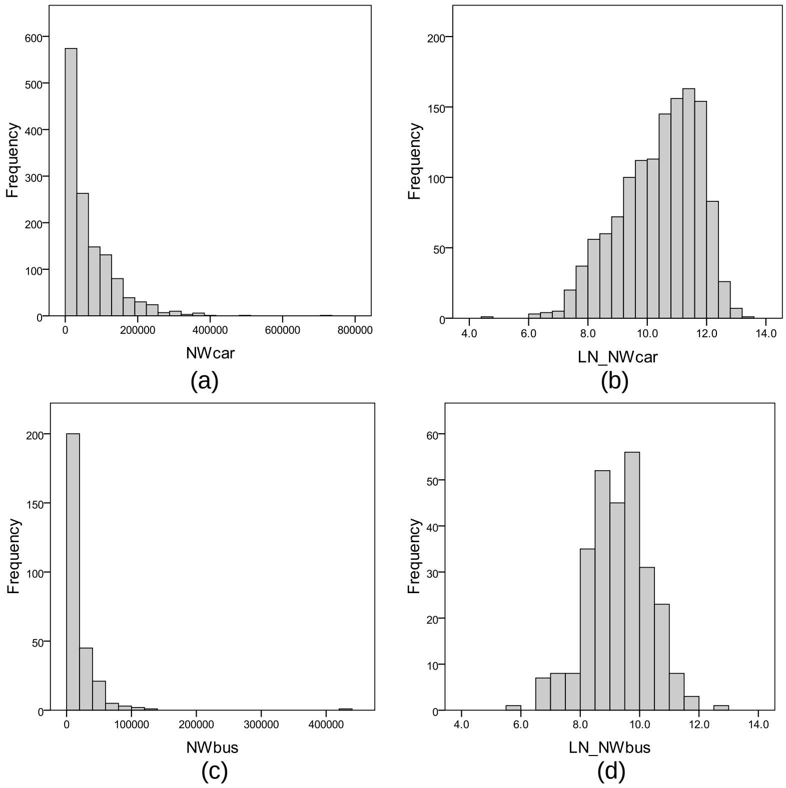


Figure 7. Probability distribution of the dependent variables with and without the logarithm transformation

Variable name	Description
NWcar	Distance travelled per individual per week in non-work trips by private vehicle as driver
NWBus	Distance travelled per individual per week in non-work trips by bus or coach
LN_NWcar	Normal Logarithm transformation of NWcar
LN_NWbus	Normal Logarithm transformation of NWbus

Table 2. Description of dependent variables, with and without log transformations.

Variable	Number of cases	Mean	S. D.	Range	
				Minimum	Maximum
NWcar	1318	65140.42	71309.60	107.87	719910.74
NWbus	278	18784.10	31384.90	345.62	428631.38
LN_NWcar	1318	10.40	1.34	4.68	13.49
LN_NWbus	278	9.27	1.07	5.85	12.97

Table 3. Descriptive statistics of the dependent variables.

4.3 Explanators

The explanatory variables considered for inclusion in the model are classified in two groups: the socio-economic factors and built environment characteristics, placing special interest in land use mix. They were chosen according to the relationships between travel patterns, socioeconomic factors and built environment characteristics found in the literature, the limitations of the available data and the limitations imposed by disaggregate nature of the analysis.

4.3.1 Socio Economic Factors

The socio economic factors initially included in the model can be classified into *individual characteristics* (gender, age range and working status) and *household characteristics* (car ownership, household size and household structure). They were chosen given the data available from the household surveys. These were introduced in the regression analysis converted into dummy variables. As explained in Dargay and Hanly (2004) "because of the perfect collinearity of groups of variables, the estimation requires that one in each group be omitted" (p.7). For this reason a **base case** was defined, which represents the individual with the attributes for which a dummy variable was not defined. The base case characteristics are shown in Table 4.

Table 5 describes the socioeconomic attributes considered and the dummy variables they were translated into, both at the individual and household level.

Individual	
Age	aged between 25 and 39
Gender	female
Working Status	Work Full-time
Household	
Household Structure	one Adult non pensionable age
Household size	1 person lives in the household
Car Ownership	owns one car

Table 4. Individual and household characteristics for the base model

Individual				
Age	A0_4	=1	if aged between 0 and 4	0 otherwise
	A5_10	=1	if aged between 5 and 10	0 otherwise
	A11_16	=1	if aged between 11 and 16	0 otherwise
	A17_18	=1	if aged between 17 and 18	0 otherwise
	A19_24	=1	if aged between 19 and 24	0 otherwise
	A40_64	=1	if aged between 40 and 64	0 otherwise
	A65+	=1	if aged 65 or over	0 otherwise
Gender	Male	=1	if Individual is male	0 otherwise
Working Status	work1	=1	if Pre-school	0 otherwise
	work2	=1	if At school / college	0 otherwise
	work4	=1	if Work part time	0 otherwise
	work5	=1	if Not working	0 otherwise
	work6	=1	if Retired	0 otherwise
Household				
Household Structure	HHstr1	=1	if one Adult pensionable age	0 otherwise
	HHstr3	=1	if one Adult 1+child	0 otherwise
	HHstr4	=1	if 2 Adults	0 otherwise
	HHstr5	=1	if 2 Adults 1 or 2 child	0 otherwise
	HHstr6	=1	if 2 Adults 3+ child or 3+ Adults 1+child	0 otherwise
	HHstr7	=1	if 3+ Adults no child	0 otherwise
Household size	Persons2	=1	if 2 people live in the household	0 otherwise
	Persons3	=1	if 3 people live in the household	0 otherwise
	Persons4	=1	if 4 people live in the household	0 otherwise
	Persons5	=1	if 5 people live in the household	0 otherwise
	Persons6	=1	if 6 or more people live in the hh	0 otherwise
Car Ownership	CO0	=1	if HH owns no car	0 otherwise
	CO2	=1	if HH owns 2 or more cars	0 otherwise

Table 5. Socio-economic variables included in the modelling process.

These socio-demographic variables are typically used as control variables. However, other variables frequently employed include income and possession of driver's license. The only data available at disaggregate level provided information on how many licensed drivers each household for each vehicle owned had. This was one of the questions in the household survey. As this was at the household level, not individual level, it is understood that this variable would be correlated to vehicle ownership also at the household level; hence the number of people with driver's license per household was not calculated from the survey. There was no disaggregate information concerning income, but it would not be expected to have a big influence in the model, and by not introducing it into the model the problem of collinearity between it and car ownership is avoided, as explained by Paulley *et al* (2006).

4.3.2 Measures of Land Use mix

4.3.2.1 Measures of Land Use mix used in the literature

Land use mix is defined as the degree to which different land uses, complementary of each other, coexist in a certain area. Many different types of measures have been used in the literature to account for land use mix. They differ from each other both in the characteristics of the built environment they measure and the level of aggregation. Next, the most common measures of land use mix are briefly described.

One widely used measure is the **employment/population ratio**. At the city level of aggregation, Banister *et al* (1997) used this ratio to analyse the link between urban form and transport energy consumption. This ratio was used by Boarnet and Sarmiento (1998), which was obtained by dividing total employment by total population in the census tract containing the individual's household. However, Giuliano (1992) and Cervero (1995) suggest that jobs/housing balance does not play a very significant role in shaping travel patterns.

The **number** of different types of facilities or land uses in an area is a measure used by some authors (eg Hanson and Schwab, 1987; Cervero and Kochelman, 1997; etc). Similarly, Van de Coevering and Schwanen (2006) employed measures of **intensity** given by jobs or residential densities. Vance and Hedel (2007) chose commercial density and Crane and Crepeau (1998) used area proportions of different land uses. Dummy variables representing the presence or absence of a type of facility within a certain travel time or distance have been employed in some studies as a proxy of land use mix. Van and Senior (2000) used this type of variable setting a 400m radius around each household because they considered it was a reasonable walking distance.

Another measure of the mixing of land use is **accessibility**, defined as the "ease of access to trip attractions"(p.3) by Erwing and Cervero (2010). It has been measured in several ways. Some papers measure it as a **distance to opportunities**, including studies measuring the distance to the closest facility of a certain type (Kitamura, Mokhtarian *et al*, 1997; Handy, 1996; Van and Senior, 2000; Dargay and Hanly, 2004; etc), or the **distance to the central business district** (e.g. Muñiz *et al*, 2005; Boarnet *et al*, 2004; Hensher, 2008; Glaeser and Kahn, 2010; Crane and Crepeau, 1998). These two types of accessibility where explained by Handy (1992), cited in Handy (1996), named them local accessibility and regional accessibility and used them to explain shopping travel patterns. Her results suggested that both levels of accessibility where inversely proportional to mean distance travelled for shopping purposes.

Hansen (1959) defined accessibility as the "potential of opportunities for interaction" (p.73), considering that the distance or time necessary to reach an opportunity plays a major role in accessibility of a point. He suggested that the accessibility at a certain point to a particular type of activity is directly proportional to the size of the activity and inversely proportional to a function of the distance separating the point from the activity. His measure of accessibility can be obtained with the following equation:

$$A_{ij} = \frac{S_j}{T_{ij}^x} \tag{2}$$

Where

A_{ij} = relative measure of the accessibility at zone *i* to an activity located within zone *j*

S_j = size of the activity in zone *j*

T_{ij} = travel time or distance between zones *i* and *j*

x = exponent describing the effect of the travel time or distance between zones

The total accessibility of zone *i* to a particular type of activity located in an area divided into *n* zones is thus defined by:

$$A_i = \sum_{j=1}^{n} \frac{S_j}{T_{ij}^x} \tag{3}$$

This type of gravity measure has been widely used (e.g. Cervero and Kockleman, 1997; Bhat and Guo, 2007) also with variations such as the cumulative-opportunity index employed by Hanson and Schwab (1986).

More sophisticated measures have also been used, such as the **Gini** coefficient (e.g. Shim *et al*, 2006), which measures the degree of equality between the concentration of two land uses along an area. But, as explained by Song and Rodriguez (2005), "it is not a very discriminating indicator" (p.20) because it can yield identical values for very different areas.

Diversity measures capturing in one single index the variety of land uses within an area have been used in several studies. Two measures of land use mix diversity used by Cervero and Kockelman (1997) were **entropy**, to capture the degree of mixing across land use categories, and a **dissimilarity index** to account for the proportion of dissimilar land uses among grid cells within an area. They used the following equations to obtain these measures:

$$Dissimilarity\ Index = \left\{\left[\sum_{j}^{k}\sum_{l}^{8}\left({}^{X_l}/_{8}\right)\right]/K\right\} \tag{4}$$

Where:

K= number of actively developed hectare grid-cells in tract, and

X_1=1 if land-use category of neighbouring hectare grid-cell differs from hectare grid-cell j (0 otherwise).

$$Entropy = \left\{\sum_{k}\left[\sum_{j} P_{jk}\ln(p_{jk})\right]/\ln(J)\right\}/K \tag{5}$$

Where:

p_{ij}= proportion of land-use category j within a half-mile radius of the developed area surrounding hectare grid-cell k;

j = number of land-use categories; and

K = number of actively developed hectares in tract.

The value of the dissimilarity index ranges between 0, corresponding to perfectly balanced land uses, and 1, representing a uniform land use. Other formulations of the dissimilarity index have also been used in the literature, as explained in Song and Rodriguez (2005).

Following Cervero and Kockelman's (1997) study, Vance and Hedel (2007) used an entropy measure of land use diversity, the **Shannon Index** (Shannon and Weaver, 1949 cited in Vance and Hedel, 2007), defined as:

$$H = -\sum_{j}^{s} p_j ln p_j \tag{6}$$

Where:

s= total number of land uses in an area (retail, service, entertainment)

p_j= fraction of land uses corresponding to the j^{th} land use

They claimed this measure was desirable because it takes into account both the number and the abundance of each type of land use relative to the other two types in the area.

Many other studies accounted for land use mix using entropy-based measures (e.g. Frank and Pivo, 1994; Greenwald, 2006).

Finally, the diversity index defined by Bhat and Gossen (2002) combines proportions of areas in each category of land use, and for 3 different categories of land uses they defined equation (7).

$$\text{Land Use mix diversity} = 1 - \left\{ \frac{\left|\frac{\mathrm{r}}{\mathrm{L}} - \frac{1}{3}\right| + \left|\frac{\mathrm{c}}{\mathrm{L}} - \frac{1}{3}\right| + \left|\frac{\mathrm{o}}{\mathrm{L}} - \frac{1}{3}\right|}{4/3} \right\} \tag{7}$$

Where:

$\mathrm{L} = \mathrm{r} + \mathrm{c} + \mathrm{o}$

r= zonal acreage in residence use

c= acreage in commercial/industrial use

o= acreage in other uses

This measure takes values between 0 and 1, the former representing a single use zone and the latter representing zones with richer land use mix.

The land use mix measure chosen for a specific study depends on the type of data available, the level of aggregation of the study and the level of sophistication of the analysis.

4.3.2.2 Definition of Measures of Land Use mix used in this study

In the present study, to avoid the “modifiable area boundary problem” (as described by Hess et al (2001): p.17) neither the use of buffers around each participant’s household, nor the definition of neighbourhoods or use census tract areas were chosen. For each respondent’s household unique values of land use measures were computed and spatially weighed indicators were used.

Most of the measures presented in section 4.3.2.1 have been computed in the literature with a certain degree of aggregation, varying from neighbourhoods, tracts, zones and up to cities. The ones selected for this study are adaptations of the accessibility measure defined by Hansen (1959), the entropy-type Shannon Index used by Vance and Hedel (2007) and the Diversity Index used by Bhat and Gossen (2004).

These measures were chosen because they can be calculated from the data available, and without the need to define sizes of neighbourhoods or limit the area of influence for each respondent’s residence.

4.3.2.2.1 Accessibility

In order to measure the intensity of the different land use types around each residence, ArcGIS was used to obtain the value of the area of the buildings containing such land uses. An exponential distance decay function was employed to spatially weight the areas depending on how distant the facility or land use type is from the respondent's household. The resulting measures are similar to a Hansen-type accessibility measure explained above. Following equation (3), in this project's case, *zone i* is each of the households participating in the survey, and *zones j* are the address points corresponding to the land use category the accessibility is being calculated for. Hence, the term T_{ij} is the distance between each household and each address points, which was obtained in ArcGIS measuring the Euclidean distance. The term S_j representing the size of the activity in *zone j* in this case is the footprint of the building containing the address point corresponding to the land use category the Accessibility is being calculated for.

The distance decay function T_{ij}^{x} spatially weights the area of buildings (or the number of dwellings) to take into account the fact that as the facilities are further away, they are less accessible. The values for *x* used in the literature vary but for the purpose of this study a value of 2 has been chosen, and coincides with Hansen's (1959) value for shopping trips.

So for this study, the accessibility measures for each respondent's household are defined as:

$$A_{ik} = \sum_{j=1}^{n} \frac{S_{jk}}{D_{ij}^{2}} \tag{8}$$

Where:

A_{ik}= Accessibility of household *i* to land use *k*

S_{jk} = Area of the building containing address point *j* with land use $k, j \leq n$

n= number of buildings in the district containing address points with land use *k*

D_{ij}= distance between the household *i* and address point *j*

The values of the accessibility obtained from equation (8) can also be considered the densities of the different land uses spatially weighted using the square of the distance as the decay function.

For the particular case of residential accessibility, it is worth mentioning that even though it is called "residential accessibility" in this study, it is not actually a land use mix measure but a measure of residential density. It was assumed that the number of dwellings was a good proxy of the residential area and it was calculated based on address point counts weighted by the distance decay function. This assumption was made both for ease of calculation but also because it is thought that the number of residential address points is a better representative of population density than the footprint of the buildings containing these address points (Address Points from the Ordnance Survey). To translate this value into a value "compatible" with the measures of accessibility computed for the other land uses, it was multiplied times the average dwelling area. The mean residential building footprint area is 72.4m^2 (obtained from ArcGIS using OS, 2010 data).

The other land use categories (Other Social; Office and Schools/Colleges; Shopping and Personal Business) were measured in terms of the area of the buildings containing OS address points corresponding to those categories. In Table 6 the NLUD groups considered in each category are presented. The land use types shown were grouped to match the trip purpose categories identified in the questionnaire.

Land use Category	**NLUD Group**	
Residential	U071	Dwellings
Working Places and Schools or colleges	U101	Manufacturing
	U102	Offices
	U104	Wholesale distribution
	U083	Education
Shopping/ Personal business	U081	Medical and health care services
	U082	Places of worship
	U084	Community services
	U091	Shops
	U092	Financial and professional services
		General commercial (*)
Other social (eating/drinking out, cinema)	U042	Amusement and show places
	U043	Libraries, museums and galleries
	U093	Restaurants and cafes
	U094	Public houses and bars

Table 6. Land use Categories used and the corresponding NLDU Group.
(*)Ordnance Survey's field surveyor's allocation (OS Base function)

For the calculation of each of the Accessibility measures, the distance between each sample household and address point included in the corresponding land use category was measured using ArcGIS. When the distance between the address point of a facility and the household's coordinates was less than 10m, it was substituted by 10m to avoid unrealistic high values of the accessibility measure that would bias the results. The threshold of 10m was chosen because it is small enough not to alter travel behaviour (very small walking distance) and high enough to account for possible inaccuracy of geocoding or for overlaps between the address points and the households.

Table 7 contains the descriptive statistics of the accessibility measures described above.

Accessibility measure	Name	Mean	S. D.	range	
				Minimum	Maximum
Residential (1)	A_Resi	5.01	2.59	0.02	27.05
Shops and Personal Business (2)	A_Shop	0.36	0.82	0.01	11.67
Other Social (3)	A_Social	0.02	0.15	0.00	3.16
Working Places & Schools/Colleges (4)	A_WoEd	0.11	0.29	0.00	7.68
Shops and Personal Business and Other Social	A_ShSo	0.38	0.90	0.01	11.74
Sum of (1), (2), (3) and (4)	A_Total	5.51	3.05	0.02	34.14

Table 7. Descriptive Statistics of the accessibility measures

Once the accessibility measures are obtained, they were combined to obtain a measure of land use mix diversity.

4.3.2.2.2 Diversity Index

Adapting equation (7) to a number n of different land use categories yields the formula:

$$\mathrm{Diversity_n} = 1 - \frac{\sum_{i=1}^{n}\left|\frac{A_i}{T} - \frac{1}{n}\right|}{(2n-2)/n} \tag{9}$$

Where:

A_i= accessibility of land use type *i* in the unit of analysis, as previously calculated

n = number of land use types

T= sum of the accessibilities for each residential area, $T = \sum_{i=1}^{n} A_i$.

The study distinguishes between 4 land use categories [(1) to (4) as per Table 6], hence using *n*=4 equation (9) takes the following shape:

$$\mathrm{Diversity_4} = 1 - \frac{\sum_{i=1}^{4}\left|\frac{A_i}{T} - \frac{1}{4}\right|}{6/4} \tag{10}$$

A diversity index using *n*=3 has also been calculated, joining the *Shopping/Personal business* category and the *Other social (eating/drinking out, cinema)* category together:

$$\mathrm{Diversity_3} = 1 - \frac{\sum_{i=1}^{3}\left|\frac{A_i}{T} - \frac{1}{3}\right|}{4/3} \tag{11}$$

As explained before, a value close to 0 indicates that there is no mixing of land uses whereas a value close to 1 indicates evenness in the distribution of land uses. Table 8 contains the descriptive statistics of these two diversity indices.

	Name	Mean	S. D.	range	
				Minimum	Maximum
Diversity Index using 3 types of LU	Diversity3	0.17	0.10	0.07	0.81
Diversity Index using 4 types of LU	Diversity4	0.09	0.09	0.01	0.67

Table 8. Descriptive Statistics of the diversity index

4.3.2.2.3 Shannon Index

The Shannon Index used in this study was adapted from equation (6) and takes the form:

$$Shannon_n = -\sum_{j=1}^{n} p_j ln p_j \qquad (12)$$

Where:

n= total number of land uses in an area

p_j= fraction of land uses corresponding to the j^{th} land use, which in terms of the variables defined in section 4.3.2.2.2 is defined as $p_j = A_j/T$.

As in 4.3.2.2.2, the Shannon index was obtained both for n=4 and n=3:

$$Shannon_4 = -\sum_{j=1}^{4} p_j ln p_j \qquad (13)$$

$$Shannon_3 = -\sum_{j=1}^{3} p_j ln p_j \qquad (14)$$

Where p_1 is the residential accessibility, p_2 is the Work/Education accessibility and p_3 is the Shopping/Personal Business and Other Social Accessibility.

Table 9 contains the descriptive statistics of these two Shannon indices.

	Name	Mean	S. D.	range	
				Minimum	Maximum
Shannon Index using 3 types of LU	Shannon3	0.27	0.17	0.05	1.09
Shannon Index using 4 types of LU	Shannon4	0.28	0.17	0.05	1.10

Table 9. Descriptive Statistics of the Shannon index

4.3.2.3 Land use variables to include in the regression analysis

The measures explained in previous sections were the ones introduced in the analysis. Nonetheless, it needs to be checked whether they are independent normal variables.
By representing the probability distributions of these variables, as done with the dependent variables in section 4.2, it is found that the Diversity and Shannon Indices, when log-transformed, show a distribution of probabilities closer to normal than without transformation. However, the residential density variable probability distribution is quite similar to a normal distribution both before and after being log-transformed, so both A_Resi and LN_A_Resi are kept as explanatory variables until their relationship with the dependent variables is analysed in section 0.
Appendix A contains the histograms for all land use variables considered, both before and after transformation.

The explanatory variables initially considered for the model are shown in Table 10.

	Variable name	Description
Accessibility	A_Resi	Residential Accessibility
	LN_A_Resi	Natural Log of residential Accessibility
	LN_A_Shop	Natural Log of Accessibility to shopping & personal business land uses
	LN_A_Social	Natural Log of Accessibility to other social land uses
	LN_A_WoEd	Natural Log of Accessibility to working places or schools/colleges
Diversity	LN_Diversity3	Natural Log of the Diversity Index for 3 types of land uses
	LN_Diversity4	Natural Log of the Diversity Index for 4 types of land uses
	LN_Shannon3	Natural Log of the Shannon Index for 3 types of land uses
	LN_Shannon4	Natural Log of the Shannon Index for 4 types of land uses

Table 10. Land use variables included in the modelling process.

4.4 Model estimation

On the basis of the analysis of the probability distributions of each of the continuous variables explained in the previous section, log transformations were used so that both the explanatory and dependent variables' probability distributions were similar to normal and problems of heteroskedasticity were avoided.

Before incorporating variables into the stepwise regression procedure, a correlation matrix was set up to examine land use variables multicollinearity by obtaining Pearson's bivariate correlation coefficients.

Pearson's correlation coefficients between explanators and dependent variables were also obtained. When two explanators were strongly correlated, the one explaining a higher correlation with the dependent variable was selected to avoid the insertion of two explanatory variables substantially correlated in the model.

Table 11 shows the correlation coefficients and their levels of significance.

		A_Resi	LN_A_ Resi	LN_A_ Shop	LN_A_ Social	LN_A_ WoEd	LN_ Diversity4	LN_ Diversity3	LN_ Shannon4
LN_A_Resi	Corr. Coeff	**.875****							
	Sig. (2-tailed)	.000							
LN_A_Shop	Corr. Coeff	**.601****	**.591****						
	Sig. (2-tailed)	.000	.000						
LN_A_Social	Corr. Coeff	**.454****	**.452****	**.628****					
	Sig. (2-tailed)	.000	.000	.000					
LN_A_WoEd	Corr. Coeff	**.602****	**.625****	**.799****	**.567****				
	Sig. (2-tailed)	.000	.000	.000	.000				
LN_Diversity4	Corr. Coeff	**.038**	**-.070***	**.709****	**.498****	**.574****			
	Sig. (2-tailed)	.183	.015	.000	.000	.000			
LN_Diversity3	Corr. Coeff	**.019**	**-.084****	**.701****	**.483****	**.539****	**.979****		
	Sig. (2-tailed)	.503	.003	.000	.000	.000	.000		
LN_Shannon4	Corr. Coeff	**.043**	**-.064***	**.703****	**.514****	**.597****	**.997****	**.965****	
	Sig. (2-tailed)	.135	.025	.000	.000	.000	.000	.000	
LN_Shannon3	Corr. Coeff	**.049**	**-.055**	**.708****	**.492****	**.608****	**.997****	**.964****	**.999****
	Sig. (2-tailed)	.085	.052	.000	.000	.000	.000	.000	.000

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

Table 11. Pearson Correlation Coefficients and 2-tailed level of significance for land use variables.

Pairs of variables having high levels of collinearity cannot be included in the model together. To select the appropriate predictors each variable was regressed against the dependent variables alone and chose to drop the variables providing a small coefficient of correlation. However, the fact that this study aims to explain the effect of land use

mix on travel behaviour implies that either one of the measures of land use diversity indices or one of the Shannon indices should be kept for inclusion in the model.
As expected, in Table 11 it can be observed that the diversity indices and Shannon indices have very strong correlation, all of them higher than 0.96 and significant at the 0.01 level. Thus only one of them can enter the model.
Table 11 also reveals that the accessibility measures less correlated to the diversity indices are A_Resi and LN_A_Resi, with the coefficients being smaller than 0.05 for all land use mix measures, which represents a negligible correlation. A_Resi and LN_A_Resi are strongly correlated with each other, with r=0.875 at 0.01 level, which means that only one of them can be included in the model.
Other accessibility measures (accessibilities to land uses other than residential) show moderate to strong correlation to the diversity measures, and this was also expected since the diversity measures were calculated by combining the accessibility measures, in the case of the diversity index a linear combination. Even though Van de Coevering and Schwanen (2006) for dropping a variable out of the model, used the criteria that the Pearson coefficient exceeds 0.8, the measures of accessibility LN_A_Shop, LN_A_Social and LN_A_WoEd are not included in the model because they have a moderate to strong correlation that could alter the relationship between the dependent variable in the model and land use mix measure, which is the major interest of the study.
The inclusion of the residential accessibility measure is important because combined with any of the diversity or Shannon indices provide a more realistic representation of the mixing of land uses. These indices only consider the mix, as they only measure the relative quantities of each land use relative to each other, not in absolute terms. So, combined with the residential accessibility measure, which represents the “amount of dwellings”, conform a stronger representation of the built environment.
Table 12 shows that, considering the variable distance travelled by private vehicle as driver, for all land use measures considered, the correlation is higher with the non-work trips than the work trips. This confirms the findings of previous studies mentioned in section 4.2. Table 12 also shows that for non-work private vehicle distances travelled correlations are significant at the p=0.01 level for all variables.

		LN_NWcar	LN_NWbus	LN_car	LN_bus
A_Resi	Corr. Coeff	**-.299****	**-.142***	**-.259****	**-.217****
	Sig. (2-tailed)	.000	.018	.000	.000
	N	1316	278	1480	408
LN_A_Resi	Corr. Coeff	**-.304****	**-.187****	**-.270****	**-.232****
	Sig. (2-tailed)	.000	.002	.000	.000
	N	1316	278	1480	408
LN_Diversity3	Corr. Coeff	**-.102****	**.012**	**-.054***	**.076**
	Sig. (2-tailed)	.000	.836	.037	.127
	N	1316	278	1480	408
LN_Diversity4	Corr. Coeff	**-.121****	**-.021**	**-.074****	**.022**
	Sig. (2-tailed)	.000	.732	.005	.657
	N	1316	278	1480	408
LN_Shannon3	Corr. Coeff	**-.133****	**-.033**	**-.085****	**.004**
	Sig. (2-tailed)	.000	.589	.001	.943
	N	1316	278	1480	408
LN_Shannon4	Corr. Coeff	**-.130****	**-.033**	**-.082****	**.004**
	Sig. (2-tailed)	.000	.579	.002	.940
	N	1316	278	1480	408

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

Table 12. Pearson correlation coefficients and 2-tailed level of significance for land use variables and dependent variables.

Bus and coach distance travelled, however, have significant levels of correlation only with the residential accessibility variables. The fact that the distance travelled by bus/coach is not significantly correlated to the diversity measures predicts that in the stepwise regression procedure the diversity variables are not likely to enter the model. This issue will be addressed in the next sections.

Taking into account the correlation coefficients shown both in Table 11 and Table 12, for the LN_NWcar regression model, the land use variables considered are LN_Shannon3 and LN_A_Resi, since they are not highly correlated with each other and have the largest Pearson's bivariate correlation coefficients with the dependent variables, r=0.133 and r=0.304 respectively, both significant at the 0.01 level.

Regarding the dependent variable for bus and coach trips LN_NWbus the same land use variables will be considered for consistency, even though none of the diversity indices show correlation.

The logic behind the sign of the correlation coefficients, for variables log-transformed, is explained in Appendix B. It will also be dealt with when interpreting the regression coefficients in the following section.

5 Results

5.1 Model Summary

Stepwise regression analysis was performed using SPSS v17, with the variables included in Table 4 as base model. The criterion for an explanator to be entered in the model is a probability of the F statistic of 0.05 and a probability of the F statistic of 0.10 to be removed. Hence, the final model only includes the variables that have high level of significance.

The outcome is shown in the model summary table (Table 13) and the ANOVA (analysis of variance) in Table 14, for both dependent variables, LN_NWcar and LN_NWbus.

The explanatory variables entering the model for **LN_NWcar** are LN_A_Resi, LN_Shannon3, CO0 (not cars owned by the household), Work5 (not working) and Male. For the model predicting **LN_NWbus** the explanators are HHstr3 (household comprised by one adult and 1+ child), LN_A_Resi, HHstr1 (household comprised by one adult at pensionable age), A40_64 (age between 40 and 64 years) and LN_Shannon3, which was included in the model even though the level of significance is very low (p=0.373) and would not have been automatically included in the model with the stepwise criteria. But, as it is the aim of the study to find out the effects of land use mix on distance travelled, it was manually included in the regression.

The model predicting LN_NWcar yields an adjusted R^2 of 0.143. This value is considerably low, meaning that only 14.3% of the variability in the dependent variable is captured by the predictors. Comparing the two models, Table 13 shows that for LN_NWbus, the value of adjusted R^2 is smaller (0.105).

Dependent Variable	R	R^2	Adjusted R^2	σ_{est}
LN_NWcar	0.382	0.146	0.143	1.260
LN_bus	0.348	0.121	0.105	1.013

R is the coefficient of correlation
$\mathbf{R^2}$ is the coefficient of determination
Adjusted $\mathbf{R^2}$ is the modification of R^2 to consider the number of explanators
$\boldsymbol{\sigma_{est}}$ is the standard error of the estimate

Table 13. Model Summary.

The high values of F shown in Table 14 indicate that the model is efficient in representing the variability in the data and also that the coefficients are significantly different from zero. This is not surprising since sample sizes are quite large in both cases.

		Sum of Squares	df	Mean Square	F	Sig.
LN_NWcar	Regression	332.601	5	66.52	44.140	0.000
	Residual	1941.069	1288	1.507		
	Total	2273.670	1293			
LN_NWbus	Regression	38.546	5	7.709	7.514	0.000
	Residual	279.071	272	1.026		
	Total	317.617	277			

R is the coefficient of correlation
$\mathbf{R^2}$ is the coefficient of determination
Adjusted $\mathbf{R^2}$ is the modification of R^2 to consider the number of explanators and observations
$\boldsymbol{\sigma_{est}}$ is the standard error of the estimate

Table 14. ANOVA.

5.2 Coefficients

5.2.1 Dependent Variable: LN_NWcar

Table 15 shows the values of the parameters, their standard errors and levels of significance for this model. At a glance, it is noticed that the larger standardised coefficients correspond to the land use variables, meaning that their effect on the dependent variable is larger than the effect of the socio-demographic variables included in the model.

	Unstandardised Coefficients		Standardised Coefficients	t statistic	Level of Signif
	β	Std. Error			
Constant	10.457	0.114		91.952	0.000
LN_A_Resi	-0.552	0.045	-0.322	-12.304	0.000
LN_Shannon3	-0.357	0.058	-0.160	-6.115	0.000
CO0	-0.596	0.171	-0.091	-3.485	0.001
Work5	0.370	0.114	0.086	3.244	0.001
Male	0.226	0.071	0.085	3.193	0.001

Table 15. Coefficients. Dependent Variable: LN_NWcar.

When inspecting the coefficients of the predictor variables in the intermediate models of the stepwise process they showed strong stability, with small variations between models. Next, the meaning of the values of the constant and of the coefficients of the dummy variables and land use variables will be explained.

The value of the constant is 10.457, and represents the logarithm of the distance travelled per week as vehicle driver by and individual whose socio-demographic characteristics match the reference case defined in Table 4, living in a household with residential accessibility and Shannon index (as defined in section 4.3.2.2) of 1.

The coefficient of the dummy variable Male, representing the gender of the individual, equals 0.230. This means that a male is likely to travel larger distances as car driver than a woman, an outcome that matches previous research regarding the influence of socio-demographic variables on travel behaviour.

The dummy variable CO0 has a negative coefficient (-0.593), meaning that individuals in households where no cars are owned (CO0=1) are likely to travel shorter distances by car for non-work purposes than people living in households owning 1 car. It should be noted that individuals in households owning 2 or more cars don't significantly travel larger distances than people in households with one car, since the variable CO2 didn't verify the entering criterion in the stepwise regression process.

The positive coefficient of the dummy variable Work5 indicates that individuals whose employment status is "not working" significantly travel larger distances by car for non-work purposes than people working full time (the employment status of the base model). This is reasonable since people not working generally have more spare time than people working to travel more for social or personal reasons. Neither students nor retired people or part time workers behave significantly different from full time employees for their corresponding dummy variables to enter the model.

These three socio demographic dummies show similar standardised coefficients in absolute terms (ranging from -0.091 for CO0 to 0.086 for Work5 and Male) meaning that their influence in the dependent variable is proportionally similar.

As explained in Appendix B, the effect of a dummy variable taking values 0 or 1 with coefficient β_n on the percentage of variation (p_y) in the dependent variable without the logarithmic transformation (y) is given by equation (15).

$$p_y = \frac{1}{e^{-\beta_n}} - 1 \tag{15}$$

Table 16 contains these values for the binary variables included in the model. In a nutshell, all other things being equal, for non work purposes a person in a household with no cars travels 44.9% less distance by car as driver than a person in a house with 1 or more cars. A person not working travels 44.7% longer distances than a person with another occupation, and males travel by car for non work purposes distance s 25.3% longer than females.

Dummy Variable	Coef predicting LN_NWcar	Percentage of change in distance travelled*
CO0	-0.596	-44.9%
Work5	0.370	44.7%
Male	0.226	25.3%

* when dummy variable changes from 0 to 1.

Table 16. Effect of the dummy variables.

With regard to the land use variables, it is interesting to mention that the coefficients represent the elasticities since both the predictors and the dependent variables are log transformed. This is helpful to compare the results with previous findings.

The coefficient obtained for the explanatory variable representing the Shannon Index appears to be negative (-0.357), which is logical since land use mix is expected to affect negatively the distance travelled. The algebraic reasoning in Appendix B, from which equation (16) was obtained, derives from the coefficients β_n the percentage of change of the distance travelled for a given change in an explanator if both are log transformed

.

$$p_y = (1 + p_x)^{\beta_n} - 1 \tag{16}$$

Where:

p_y is the percentage of change in y,

p_x is the percentage of change in x and

β_n is the unstandardised coefficient of the variable *Ln(x)* in the model predicting *Ln(y)*.

Assuming that Shannon index increases (p_x>0), and taking into account that β_n<0 Table 15the value of $(1 + p_x)^{\beta_n}$ is positive below one, then p_y results negative. In other words, the model shows that when the land use mix increases, the distance travelled by car for non-work purposes decreases.

An equivalent reasoning can be made for the variable LN_A_Resi, which yielded a negative parameter of -0.552. This means that the predicted value of the distance travelled by private vehicle drivers per week for non-work purposes increases as the residential accessibility decreases.

Applying equation (16) to find the influence of the land use measures the values shown in Table 17 were obtained. These results imply that an increase in the Shannon index of 10% yields a 3.35% reduction in distance travelled for non-work trips, whereas the same percentage of increase of the accessibility to dwellings (or spatially weighted residential density) produces a decrease of -5.12% in distance travelled by private vehicle by driver per week for non-work purposes.

Variable	Coef predicting LN_NWcar	% of change in explanators, no log	Percentage of change in distance travelled*
LN_A_Resi	-0.552	10%	-5.12%
LN_Shannon3	-0.357	10%	-3.35%

* when explanatory variable increases 10%.

Table 17. Effect of the land use variables on LN_NWcar.

5.2.2 Dependent Variable: LN_NWbus

Table 18 contains the values of the parameters, their standard errors and levels of significance.

The coefficients of the predictor variables in the intermediate models did not vary during the stepwise process, showing strong stability, in this model too.

	Unstandardised Coefficients		Standardised Coefficients	t statistic	Level of Signif
	β	Std. Error			
Constant	9.589	0.197		48.656	0.000
HHstr3	-0.887	0.246	-0.209	-3.611	0.000
LN_A_Resi	-0.260	0.076	-0.198	-3.424	0.001
HHstr1	0.501	0.218	0.134	2.296	0.022
A40_64	-0.302	0.139	-0.128	-2.175	0.031
LN_Shannon3	-0.098	0.109	-0.052	-0.893	0.373

Table 18. Coefficients. Dependent Variable: LN_NWBus.

The value of the constant (9.589), slightly above the mean value (9.27, according to Table 3) corresponds to the value of LN_NWbus for individuals in the reference case defined in Table 4 with LN_A_Resi=0 and LN_Shannon3=0. These values correspond to a low value of the residential accessibility and a high value of the land use mix.

The sign of the coefficient of the dummy variable A40_64 indicates that the distance travelled by bus or coach is likely to be shorter if the person has between 40 and 64 years of age. This is a reasonable outcome because middle age people tend to travel more by car instead of public transport, as explained by Hanly and Dargay (2004), who found that the age group with higher car shares is 35 to 64.

With regard to the binary variable HHstr3, the negative coefficient (-0.887) implies that individuals living in households where one adult and one or more children live travel less by bus/coach than an adult living alone (base case), all other things equal. The explanation could be that lone parents' travel behaviour is less flexible than single adults' or couples' since they need to look after children alone, and public transport may not be the most convenient way for them to travel because it is not as flexible as driving a private vehicle. This also agrees with Dargay and Hanly's (2004) conclusion that people in households with children are less likely to travel by public transport than those without children. The other household structure variable that was significant enough to verify the enter criteria is HHstr1, one adult at pensionable age living in the household. The coefficient (0.501) indicates that retired people are likely to travel longer distances by bus/coach than other adults, for non-work trip purposes.

As done for the previous model, the effect of the dummy variables in absolute terms is presented in table Table 19, revealing that in comparison with the reference case, individuals in households with one adult and at least one child travel 58.8% less distance, those in households where only one pensionist lives travel 65.0% more and those ages 40 to 64 travel distances 26.1% shorter, all else being equal.

Dummy Variable	Coef predicting LN_bus	Percentage of change in distance travelled*
HHstr3	-0.887	-58.8%
HHstr1	0.501	65.0%
A40_64	-0.302	-26.1%

* when dummy variable changes from 0 to 1.

Table 19. Effect of the dummy variables.

With respect to the land use variables, as mentioned earlier, the variable representing land use mix doesn't have a high level of significance (it is significant at the 63% level). The standardised coefficient is also low, explaining only 5.2% of the variation in LN_NWbus explained by the model. However, the residential accessibility is significant at the 99.9% level, and this difference may be because bus and coach use depends more on the level of supply, which is usually linked to the level of demand, which depends on the population density. As observed in Table 20, the model suggests that a change of +10% in the residential accessibility of an individual's home produces a change of -2.45% in distance that he or she travels by bus or coach. This percentage is smaller than the percentage previously shown for private vehicle distance travelled, which means that changes in residential accessibility affect more private vehicle use than public transport use. This result suggests that people living in high residential density areas travel shorter distances by public or private transport, probably because they walk and cycle more, or because they do shorter trips. But the decrease is steeper in the case of private car use, which is reasonable since the residential density also affects bus and coach service level, by increasing accessibility to bus stops, hence making bus and coach travel more attractive. Hence, the effect of some bus trips being substituted by walking and cycling trips is counterbalanced, which explains why bus and coach distances travelled are less affected by changes in residential density than distance travelled by private vehicle. As described in Balcombe *et al* (2004), public transport trip frequency increases with density, but average trip distance decreases.

With regard to the lack of influence of land use mix on LN_NWbus, an explanation is that given equal residential densities, people living in areas with a more balanced mix of land uses may choose to travel less by car, and walk or cycle instead to local facilities and services. Some people may travel less by bus or coach since they can walk or cycle, but others may choose to use more public transport to access nearby areas with a range of facilities available. People in high land use mix areas may travel more frequently by public transport but do shorter trips, and the overall effect may not be identifiable when looking at distance travelled per week as in the present model.

Variable	Coef predicting LN_bus	% of change in explanators, no log	Percentage of change in distance travelled*
LN_A_Resi	-0.248	10.00%	-2.45%
LN_Shannon3	-0.098	10.00%	-0.93%

* when explanatory variable increases 10%.

Table 20. Effect of the land use variables on LN_NWbus.

5.2.3 Comparison with the literature

The low level of explanation of the models (R^2=0.145 and R^2=0.105) is consistent with disaggregate models that have been estimated in the literature, with values of R^2 generally below 0.25 (Handy, 1996). Stead (2001) obtained values of R^2 in the range of 0.22 to 0.24 when regressing all the land use and socio-demographic variables to predict the distance travelled per person at the individual level of analysis. As Banister (2007) states, for land use-travel models, "the level of explanation decreases as the level of disaggregation increases" (Banister, 2002, p3). Schwanen and Mokhtarian (2005) obtained a level of explanation on their model predicting VMT of 0.391, which is larger than for this study; however they also introduced individual preferences as explanators.

Results can also be compared in terms of elasticities. Erwing and Cervero's (2010) recently presented in their meta-analysis elasticity values from a large number of studies in the field. The elasticity of the distance travelled by private vehicle against land use mix in the present project is -0.37 (directly the coefficient due to the log-log transformations), whereas the values presented in Erwing and Cervero's meta-analysis vary from -0.02 and -0.27 for entropy-based land use mix measures's relationship with non-work distance travelled.

Concerning the residential accessibility, this study's results provide elasticities of -0.55 and -0.26, which fall within Erwing and Cervero's (2010) range of values collected (from -0.04 to -0.58 for non work distance travelled per person or household).

In this study elasticities are higher than the average values. The difference can be due to the way the Shannon Index was obtained in this study, using disaggregate accessibility measures where areas of different land uses were spatially weighted. In most studies the index is calculated with a certain degree of aggregation, after a boundary is set around the households or cells or neighbourhoods are defined.

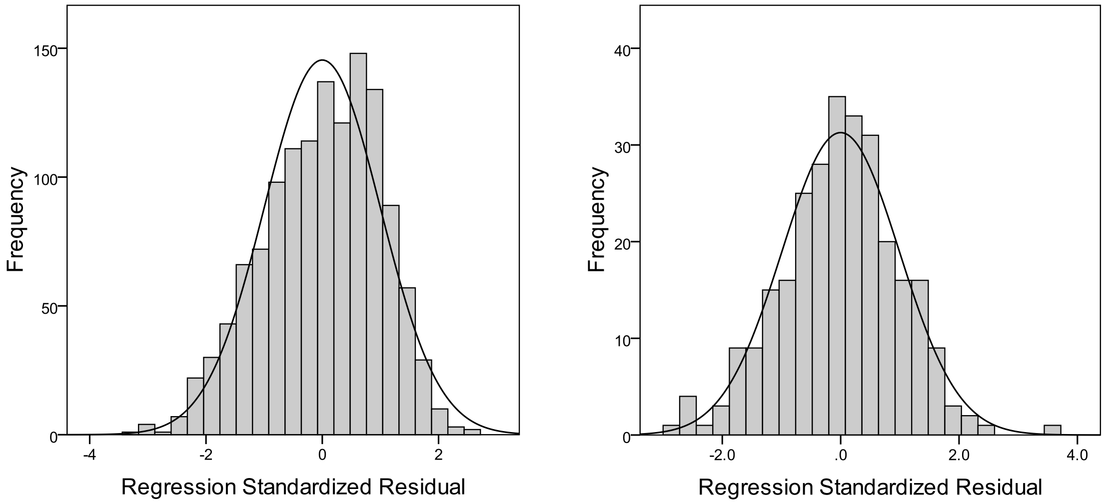
5.3 Residuals

To avoid problems of heteroskedasticity and to test whether the hypothesis used in order to estimate the OLS model can be rejected, the residuals were visually inspected, as per Van de Coevering and Schwanen (2006).

Figure 8 shows the residuals ε_i against the frequency of occurrence. A normal curve with mean 0 and standard deviation 1 is superimposed, and the values of ε_i look reasonably close to normally distributed for the dependent variables in both models.

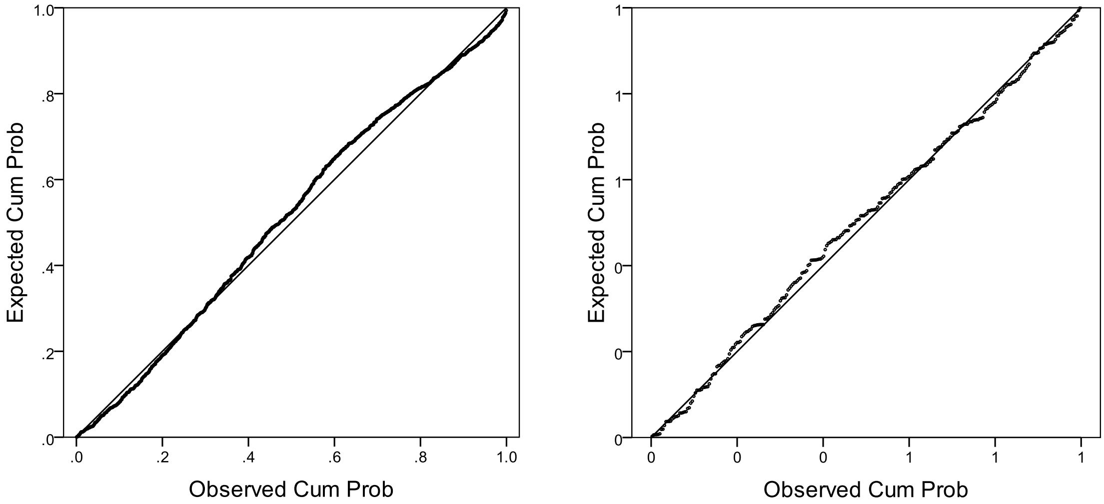
Figure 10 shows a plot of the residuals versus the predicted *Ln(y)*. The plot for the dependent variable LN_NWcar shows a pattern that indicates that there are no problems with the assumption for ε_i, since they seem to be randomly distributed and with similar variance for all values of the dependent variable. It should be noted that the scatter plot for LN_NWcar shows a slight heteroskedasticity, since the variance of the error term seems to decrease slightly as LN_NWcar increases. However, this difference is not substantial enough to worry about.

For the second model the errors seem to be close to normally distributed in all the plots, verifying the hypothesis of $\varepsilon_i \sim iid\ N(0, \sigma_\varepsilon^2)$.



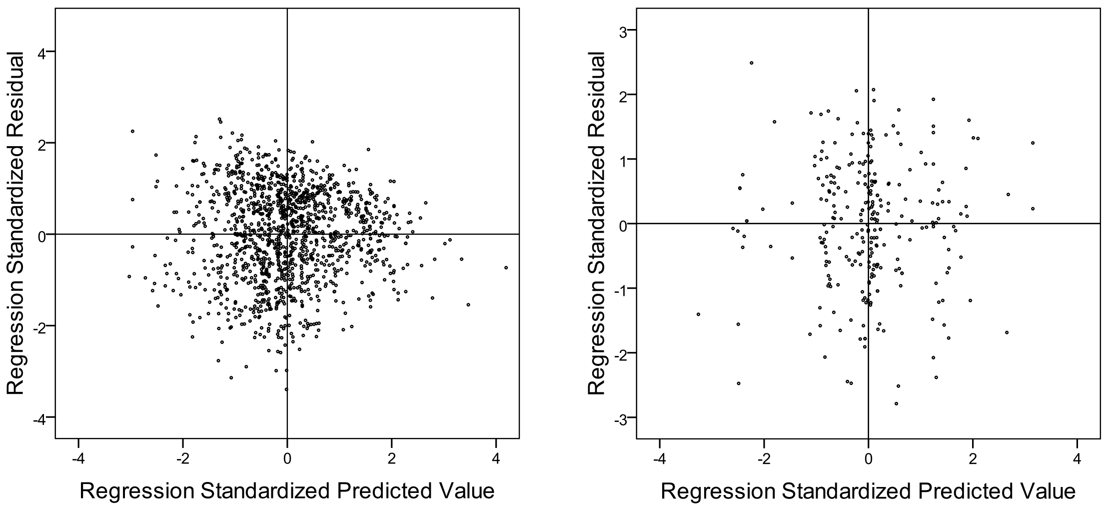
(a) LN_NW_car (b) LN_NWbus

Figure 8. Histogram of the regression standardised residuals.



(c) LN_NW_car (d) LN_NWbus

Figure 9. Normal probability plot of the regression standardised residuals.



(e) LN_NW_car (f) LN_NWbus

Figure 10. Scatter plot of the regression standardised residuals and predicted values.

6 Impact on energy consumption

It was mentioned in the first part of the study the interest on the impact of land use mix on transport energy consumption, as a rough indication of the environmental externalities of travel such as climate change and fossil fuel resource depletion.

For measuring the impact on energy consumption of land use mix, it is assumed that the transport energy consumption is directly proportional to the distance travelled per vehicle. This is a strong assumption since the energy consumed by a vehicle varies depending on several factors, such as the characteristics of the vehicle (type, size, weight, engine, etc), the type of fuel it consumes, the occupancy, the driving conditions, etc. However, distance travelled per vehicle has been defined as a good surrogate for energy consumption by many authors (e.g. Erwing and Cervero, 2010; Van de Coevering and Schwanen, 2006), and particularly as a very good indicator of transport emissions (Stead, 1999a).

The model predicting the distance travelled per individual by bus or coach doesn't include the land use mix variables at a significant level, it can therefore be concluded that it does not significantly alter the quantity of energy consumed by public transport use.

On the other hand, the model for the distance travelled per private vehicle shows that a change of 10% in the degree of land use mix [measured as a Shannon index per equation (12)] of an individual's residence location produces a change in private vehicle energy consumption of 3.41% in the opposite direction.

Similarly, a change of 10% in the accessibility to dwellings [as measured in equation (8)] produces a change of 5.24% private vehicle energy consumption in the opposite direction.

7 Conclusions

In summary, the results indicate that land use mix does play an important role in determining distance travelled by private vehicles with the consequent implications on energy consumption.

This implies that, for reducing transport energy consumption and achieving more sustainable mobility practices, policy should facilitate mixed use developments, with high residential densities, were complementary land uses coexist and are accessible by public transport, walking and cycling. This would make people be able to choose means of transport alternative to private vehicles, and would have a noticeable effect particularly on non-work trips patterns, which are less constrained than commuting trips.

It is important to keep in mind that Cherwell is a rural district and has less medium size population centres than other rural districts in the County of Oxfordshire (Melling, 2009). For this reason, the higher values of residential density in the district may not be very high when compared to larger urban agglomerations, so we cannot extrapolate the results of this study to areas with a very different urban structure.

It is worth mentioning that the level of explanation of the models, even though being normal when compared to other studies of the same sort, is quite low (R^2=15%) indicating that other factors not included in the models may have very relevant impacts on distances travelled, thus in energy consumption. These factors can be the level of transport provision, individual's preferences, transport prices, etc.

Other shortcomings that can be identified within the study, include that the travel questionnaire doesn't distinguish between the type of private vehicles used (including motorbikes, car and vans together) implying a further increase of inaccuracy when assuming that distance travelled is proportional to energy consumed travelling. Furthermore, accessibility measures were calculated using Euclidean distance, which is less computationally demanding, but provides less accurate accessibility measures than network distance. Care must be taken when generalizing the results from this model since they only apply to non-work trips, which for the sample under consideration account for slightly over half of the total distance travelled by private vehicle drivers.

Still, the present study has important strengths. One of the main advantages of this study is that GIS was employed to obtain measures of the land use characteristics without relying on the definition of neighbourhoods or boundaries. This accurate way of measuring at the disaggregate level provides a great deal of variation that helps identifying the nature of the very complex relationship existing between land use mix and travel. The large sample of data that was available, with detailed travel and socio-demographic information at disaggregate level also improves the quality of the results, and together with the fact that the characteristics of the district population were realistically represented by the sample constitute another important strength of the study.

References

Badland, H.M., Schofield, G.M. & Garrett, N., 2008. Travel behavior and objectively measured urban design variables: Associations for adults travelling to work. *Health & Place,* 14, 85-95.

Balcombe, R., Mackett, R., Paulley, N., Preston, J., Shires, J., Titheridge, H., Wardman, M. & White, P., 2004. *The demand of public transport: a practical guide*. [online] TRL Report 593. Available at http://www.demandforpublictransport.co.uk/TRL593.pdf [Accessed 20 July 2010]

Banister, D., 2005. *Unsustainable Transport: City Transport in the New Century*, London: Routledge.

Banister, D., Watson, S., & Wood, C., 1997. Sustainable cities, transport, energy, and urban form. *Environment and Planning B: Planning and Design*, 24 (1) 125-143.

Bhat, C.R. & Gossen, R., 2004. A Mixed Multinomial Logit Model Analysis of Weekend Recreational Episode Type Choice. *Transportation Research Part B*, 38 (9), 767-787.

Bhat, C.R. & Guo, J.Y., 2007. A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. *Transportation Research Part B,* 41 (5), 506–526.

Boarnet, M.G. & Sarmiento, S., 1998. Can Land-use Policy Really Affect Travel Behaviour? A Study of the Link between Non-work Travel and Land-use Characteristics. *Urban Studies,* 35 (7), 1155-1169.

Boarnet, M.G., Nesamani K.S., & Smith, C.S., 2004. Comparing the Influence of Land Use on Nonwork Trip Generation and Vehicle Distance Traveled: An Analysis Using Travel Diary Data. *In: 83rd Annual Meeting of the Transportation Research Board*, 11-15 January 2004, Washington, D.C.

Cervero, R. & Kockelman, K., 1997. Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment*, 2 (3), 199-219.

Cervero, R., 1995. Planned Communities, Self-containment and Commuting: A Cross-national Perspective. *Urban Studies*, 32, 1135-1161.

Crane, R. & Crepeau, R., 1998. Does Neighbourhood Design Influence Travel? A Behavioral Analysis of Travel Diary Data and GIS Data. *Transportation Research D,* 3 (4), 225-238.

Dargay, J. & Hanly, M., 2004. Land use and mobility. *In: World Conference on Transport Research*, 4-8 July 2004, Istanbul, Turkey.

Department of Trade and Industry. 2002. *Energy Consumption in the United Kingdom*. [online] London: Department of Trade and Industry. Available from: http://www.decc.gov.uk/assets/decc/statistics/publications/ecuk/file11250.pdf [Accessed 18 March 2010]

Directgov, 2009. *The vehicles you can drive or ride and minimum ages* [online]. © Crown Copyright. UK Government Available from: http://www.direct.gov.uk/en/Motoring/DriverLicensing/WhatCanYouDriveAndYourObligations/DG_4022547. [Accessed 09 September 2010].

Ewing, R. & Cervero, R., 2002. Travel and the Built Environment - Synthesis, *Transportation Research Record* 1780, TRB Washington.

Ewing, R. & Cervero, R., 2010. Travel and the Built Environment: A Meta-analysis. *Journal of the American Planning Association*, 76 (3) 265–294.

Ewing, R., Haliyur, P. & Page, G.W., 1994. Getting Around a Traditional City, a Suburban Planned Unit, Development, and Everything in Between. *Transportation Research Record*, 1466, 53-62.

Frank, L.D., & Pivo, G., 1994. Impacts of Mixed Use and Density on Utilization of Three Modes of Travel: Single-Occupant Vehicle, Transit, and Walking. *Transportation Research Record*, 1466, 44-52.

Giuliano, G. & Dargay, J., 2006. Car ownership, travel and land use: a comparison of the US and Great Britain. *Transportation Research Part A*, 40, 106-124.

Giuliano, G., 1992. Is jobs-housing balance a transportation issue? *Transportation Research Record*, 1305, 305-312.

Glaeser, E.L. & Kahn, M.E., 2010. The Greenness of Cities: Carbon Dioxide Emissions and Urban Development. *Journal of Urban Economics*, 67 (3), 404-418.

Gómez Ibáñez, J.A., 1991. A global view of automobile dependence. *Journal of the American Planning Association*, 57 (3), 376-379.

Greenwald, M.J., 2006. The relationship between land use and intrazonal trip making behaviors: Evidence and implications. *Transportation Research Part D*, 11 (6), 432-446.

Handy, S., 1992. *Regional Versus Local Accessibility: Variation in Suburban Form and the Implications for Nonwork Travel*. PhD Thesis, University of California at Berkeley.

Handy, S., 1996. Methodologies for Exploring the Link Between Urban Form and Travel Behavior. *Transportation Research D*, 1 (2), 151-165.

Handy, S.L., Cao, X. & Mokhtarian, P., 2005. Self-selection in the Relationship between the Built Environment and Walking? Empirical Evidence from Northern California. *Journal of American Planning Association*, 72 (1), 55-74.

Hansen, W.G., 1959. How accessibility shapes land-use. *Journal of the American Institute of Planners*, 25 (2), 73-76.

Hanson, S., & Schwab, M., 1987. Accessibility and Intraurban Travel. *Environment and Planning A*, 19, 735-748.

Hensher, D.A., 2008. Climate change, enhanced greenhouse gas emissions and passenger transport. What can we do to make a difference? *Transportation Research Part D: Transport and Environment*, 13 (2), 95-111.

Karathodorou, N., Graham, D.J. & Noland, R.B., 2010. Estimating the effect of urban density on fuel demand. *Energy Economics*, 32, 86-92.

Kitamura, R., Mokhtarian P.L. *et al*, 1997. A Micro-Analysis of Land Use and Travel in Five Neighborhoods in the San Francisco Bay Area. *Transportation*, 24, 125-158.

Melling, M., 2009. *Living in Cherwell* [online]. MM Consulting Limited for Cherwell District Council. Available from: http://www.cherwell.gov.uk/media [Accessed 06 September 2010].

Mindali, O., Raveh, A. & Salomon, I., 2004. Urban density and energy consumption: a new look at old statistics. *Transportation Research A*, 38, 143-162.

Muñiz, I., Galindo, A. & García, M.A., 2005. Urban form and the ecological footprint of commuting. The case of Barcelona. *Ecological Economics*, 55, 499-514.

Newman, P.W.G. & Kenworthy, J.R., 1989. *Cities and Automobile Dependence: An International Sourcebook.* Aldershot, UK: Gower.

Office for National Statistics (ONS), 2004. *Census 2001: Definitions*, *Office for National Statistics.* © Crown copyright 2004. [online] Available from http://www.statistics.gov.uk/downloads/census2001/definitions_chapters_1_5.pdf [Accessed 24 August 2010].

Office for National Statistics, *2001 Census: Digitised Boundary Data.* (England and Wales) Crown copyright 2003 [computer file].

OS, 2010. MasterMap Address Layer 2 [GML geospatial data], Coverage: Cherwell District, updated Jul 2010. Ordnance Survey © Crown copyright. All rights reserved.

Oxfordshire data Observatory, 2010. *Sharing Local Data* [online]. Available from: http://www.oxfordshireobservatory.info [Accessed 20 July 2010].

Paulley, N., Balcombe, R., Mackett, R., Titheridge, H., Preston, J., Wardman, M., Shires, J. & White, P., 2006. The demand for public transport: the effects of fares, quality of service, income and car ownership. *Transport Policy*, 13, 295-306.

Schwanen, T. & Mokhtarian, P.L., 2005. What if you live in the wrong neighborhood? The impact of residential neighborhood type dissonance on distance traveled. *Transportation Research D*, 10, 127-151.

Shannon, C.E., Weaver, W., 1949. *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.

Song, Y., & Rodriguez,D., 2005. *The measurement of the level of mixed land uses: A synthetic approach* [online], Carolina Transportation Program White Paper Series. Available from: http://planningandactivity.unc.edu [Accessed 16 June 2010].

Souche, S., 2010. Measuring the structural determinants of urban travel demand. *Transport Policy*, 17 (3), 127-134.

Stead, D. & Marshall, S., 2001. The relationship between urban form and travel patterns: an international review and evaluation, *European Journal of Transport and Infrastructure Research*, 1 (2), 113-141.

Stead, D., 1999a. Relationships between transport emissions and travel patterns in Britain. *Transport Policy*, 6, 247-258.

Stead, D., 1999b. *Planning for Less Travel: Identifying Land Use Characteristics Associated with More Sustainable Travel Patterns*. PhD thesis, Bartlett School of Planning, University College London.

Stead, D., 2001. Relationships Between Land Use, Socio-economic factors and Travel Patterns. *Environment and Planning B: Planning and Design*, 28, 499-528.

Taylor Nelson Sofres (TNS), 2001. *Cherwell District Travel Survey*, commissioned by North Oxfordshire Consortium.

Thideridge, H. & Hall, P., 2006. Changing travel to work patterns in South East England. *Journal of Transport Geography*, 14, 60-75.

Van de Coevering, P. & Schwanen, T., 2006. Re-evaluating the impact of urban form on travel patterns in Europe and North-America. *Transport Policy*, 13, 229-239.

Van, U.P. & Senior, M., 2000. The contribution of mixed land uses to sustainable travel in cities. In: K. Williams, E. Burton, M. Jenks eds. *Achieving Sustainable Urban Form*. London: Spon, 139-148.

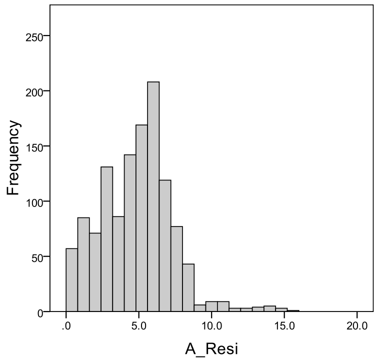
Vance, C. & Hedel, R., 2007. The Impact of Urban Form on Automobile Travel: Disentangling Causation from Correlation. *Transportation*, 34, 577-588.

Appendix A

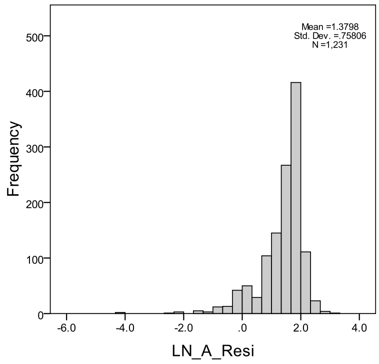
Accessibilty measures

Land Use type | **No transformation** | **Natural Logarithm**

Residential

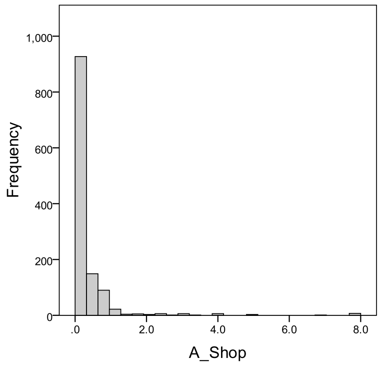


(a)

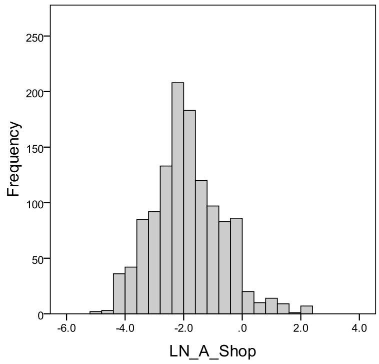


(b)

Shopping and personal business

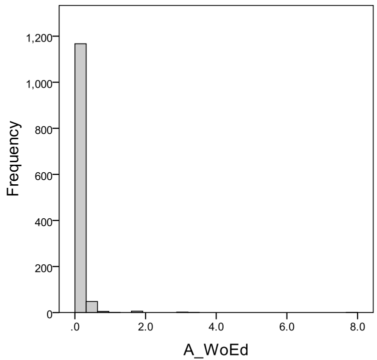


(c)

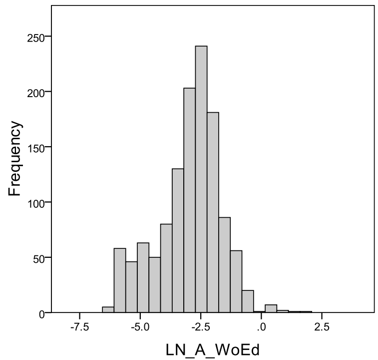


(d)

Working Places and Schools/Colleges

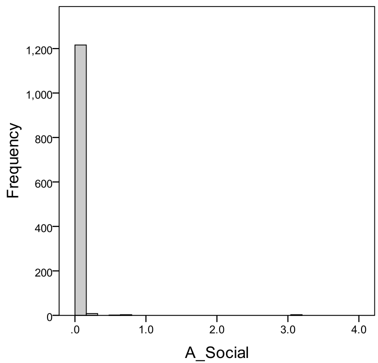


(e)

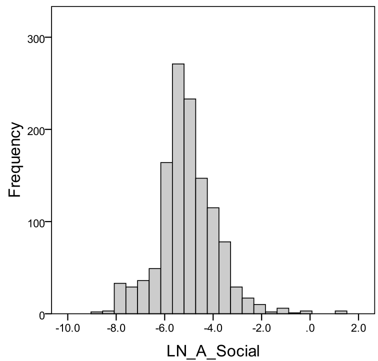


(f)

Other Social

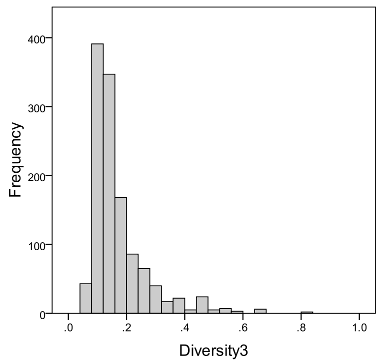
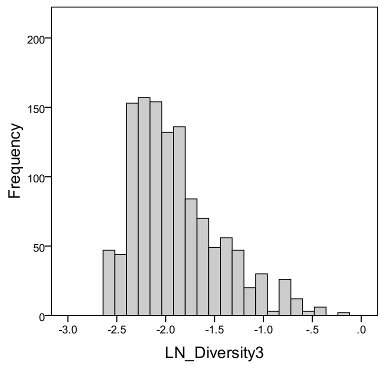
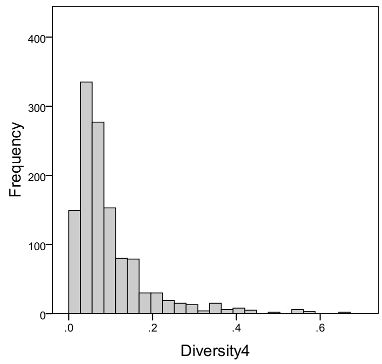
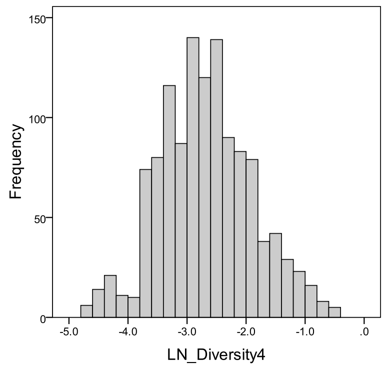


(g)

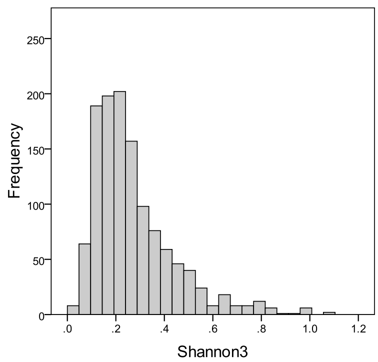
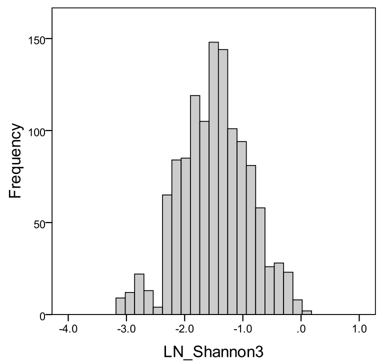
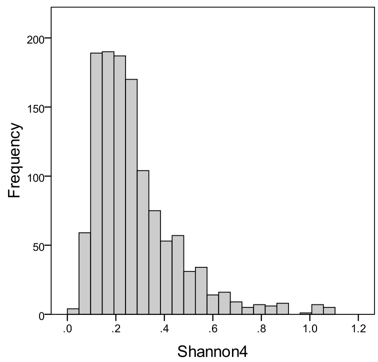
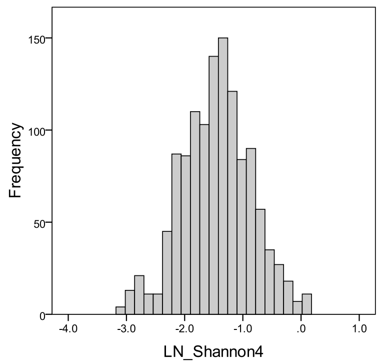


(h)

Diversity Index

	No transformation	Natural Logarithm
Diversity index for 3 types of Land uses	 (i)	 (j)
Diversity index for 4 types of Land uses	 (k)	 (l)

Shannon Index

	No transformation	Natural Logarithm
Shannon index for 3 types of Land uses	 (m)	 (n)
Shannon index for 4 types of Land uses	 (o)	 (p)

Appendix B

Interpretation of the coefficient of the Dummy variables when the dependent variable is log-transformed

In equation (17) the formulation of a model where the dependent variable (y) is log transformed and one of the independent variables (x_n) is a dummy is shown. It will be analysed how the sign of the coefficient β_n affects the relationship between the variables y and x_n.

$$Ln(y) = \beta_0 + \sum_{i=1}^{n-1} \beta_i \, x_i + \beta_n(x_n) \tag{17}$$

Let's assume all other explanatory variables remain constant, and that for values of x_n of 0 and 1 the model yields the $Ln(y_1)$ and $Ln(y_2)$ respectively:

$$Ln(y_1) = \beta_0 + \sum_{i=1}^{n-1} \beta_i \, x_i + \beta_n(0) \tag{18}$$

$$Ln(y_2) = \beta_0 + \sum_{i=1}^{n-1} \beta_i \, x_i + \beta_n(1) \tag{19}$$

The difference between equations (18) and (19) is:

$$Ln(y_1) - Ln(y_2) = -\beta_n$$

$$Ln\left(\frac{y_1}{y_2}\right) = -\beta_n$$

$$\frac{y_1}{y_2} = e^{-\beta_n} \tag{20}$$

With the relationship shown in equation (20), the influence of β_n on the dependent variable y can be calculated. Let's take the percentage of change of y (p_y%):

$$y_2 = (1 + p_y)y_1$$

$$\frac{y_1}{y_2} = \frac{1}{1 + p_y} = e^{-\beta_n}$$

$$p_y = \frac{1}{e^{-\beta_n}} - 1 \tag{21}$$

If $\beta_n > 0$ → $p_y > 0$, meaning that $y_2 > y_1$ (when x=1, y is smaller than when x=0).

If $\beta_n < 0$ → $p_y < 0$, meaning that $y_2 < y_1$ (when x=1, y is greater than when x=0).

Interpretation of the coefficient of the log-transformed variables when the dependent variable is log-transformed

In equation (22) the formulation of a model where both the dependent (y) and one of the explanatory variables (x_n) are log transformed is shown. It will be analysed how the sign of the coefficient β_n affects the relationship between the variables y and x_n.

$$Ln(y) = \beta_0 + \sum_{i=1}^{n-1} \beta_i x_i + \beta_n Ln(x_n) \tag{22}$$

Let's assume all other explanatory variables remain constant, and that for values of x_n of x_{n1} and x_{n2} the model yields the $Ln(y_1)$ and $Ln(y_2)$ respectively:

$$Ln(y_1) = \beta_0 + \sum_{i=1}^{n-1} \beta_i x_i + \beta_n Ln(x_{n1}) \tag{23}$$

$$Ln(y_2) = \beta_0 + \sum_{i=1}^{n-1} \beta_i x_i + \beta_n Ln(x_{n2}) \tag{24}$$

The difference between equations (18) and (19) is:

$$Ln(y_1) - Ln(y_2) = \beta_n[Ln(x_{n1}) - Ln(x_{n2})]$$

$$Ln\left(\frac{y_1}{y_2}\right) = \beta_n Ln\left(\frac{x_{n1}}{x_{n2}}\right)$$

$$\frac{y_1}{y_2} = \left(\frac{x_{n1}}{x_{n2}}\right)^{\beta_n} \tag{25}$$

Given the relationship shown in equation (25), assuming that $x_{n2} > x_{n1}$, the influence of the sign of β_n in the values of y:

If $\beta_n > 0 \rightarrow y_2 > y_1$, meaning that an increase in x produces an increase in y.

If $\beta_n < 0 \rightarrow y_1 > y_2$, meaning that an increase x produces a decrease y.

In terms of the absolute value of β_n, the influence can be calculated by assuming that for a percentage of change of the variable x, denoted p_x, variable y changes p_y%

$$x_{n2} = (1+p_x)\, x_{n1}$$

$$y_2 = (1+p_y)\, y_1$$

hence, from equation (25), the percentage of change in y given the percentage of change in x and the coefficient β_n can be obtained:

$$y_2 = (1 + p_x)^{\beta_n} y_1$$

$$p_y = (1 + p_x)^{\beta_n} - 1$$