**SEEDS**
**S**urrey **E**nergy **E**conomics
**D**iscussion paper **S**eries

SURREY
ENERGY
ECONOMICS
CENTRE

# Food Superstores, Food Deserts and Traffic Generation in the UK: A Semi-Parametric Regression Approach

Colin Black, David Broadstock, Alan Collins
and Lester C. Hunt

March 2006

SEEDS 112
ISSN 1749-8384

Department of Economics
University of Surrey

The **Surrey Energy Economics Centre (SEEC)** consists of members of the Department of Economics who work on energy economics, environmental economics and regulation. The Department of Economics has a long-standing tradition of energy economics research from its early origins under the leadership of Professor Colin Robinson. This was consolidated in 1983 when the University established SEEC, with Colin as the Director; to study the economics of energy and energy markets.

SEEC undertakes original energy economics research and since being established it has conducted research across the whole spectrum of energy economics, including the international oil market, North Sea oil & gas, UK & international coal, gas privatisation & regulation, electricity privatisation & regulation, measurement of efficiency in energy industries, energy & development, energy demand modelling & forecasting, and energy & the environment. SEEC also encompasses the theoretical research on regulation previously housed in the department's Regulation & Competition Research Group (RCPG) that existed from 1998 to 2004.

SEEC research output includes **SEEDS** - **S**urrey **E**nergy **E**conomic **D**iscussion paper **S**eries (details at www.seec.surrey.ac.uk/Research/SEEDS.htm) as well as a range of other academic papers, books and monographs. SEEC also runs workshops and conferences that bring together academics and practitioners to explore and discuss the important energy issues of the day.

SEEC also attracts a large proportion of the department’s PhD students and oversees the MSc in Energy Economics & Policy. Many students have successfully completed their MSc and/or PhD in energy economics and gone on to very interesting and rewarding careers, both in academia and the energy industry.

Enquiries:

***Director of SEEC and Editor of SEEDS*:**
Lester C Hunt
**SEEC,**
Department of Economics,
University of Surrey,
Guildford GU2 7XH,
UK.

Tel: +44 (0)1483 686956
Fax: +44 (0)1483 689548
Email: L.Hunt@surrey.ac.uk

**www.seec.surrey.ac.uk**

University of Surrey

Surrey Energy Economics Centre (SEEC)
Department of Economics

SEEDS 112
ISSN 1749-8384

# FOOD SUPERSTORES, FOOD DESERTS AND TRAFFIC GENERATION IN THE UK: A SEMI-PARAMETRIC REGRESSION APPROACH.

Colin Black, David Broadstock, Alan Collins
and Lester C. Hunt

March 2006

**ABSTRACT**

This study contributes another route towards explaining and tackling 'food desert' effects. It features the estimation of a (semi-parametric) trip attraction model for food superstores in the UK using a composite dataset. The data comprises information from the UK Census of Population, the NOMIS (National Online Manpower Information System) archive and traffic and site-specific data from the TRICS (Trip Rate Information Computer System) databases. The results indicate that traffic to a given food superstore, ceteris paribus, increases with household car ownership, store parking provision, site size (floor space), and distance to the nearest competitor. Furthermore, increases in public transport provision are shown to be associated with increasing car trips. This latter effect is discussed in the light of planning policy for development control purposes and a role linked to the reinforcement of 'food deserts'. The results also reveal activity-specific household economies of scope and scale. It is suggested how these may also further perpetuate unsustainable development and 'food desert' characteristics.

*JEL Classification Numbers:* C31, R12, R41.

*Key Words*: Traffic Generation, Food Superstores, Food Deserts, Activity Based Travel, Sustainable Development, Modelling.

# Food Superstores, Food Deserts and Traffic Generation in the UK: A Semi-Parametric Regression Approach[#]

**Colin Black*, David C. Broadstock**, Alan Collins*** and Lester C. Hunt****

* JMP Consultants Ltd., London, UK

** Surrey Energy Economics Centre (SEEC), University of Surrey, UK

*** Department of Economics, University of Portsmouth, UK

## 1. Introduction

The growth and spatial placement of food superstores, alongside society's increasing reliance upon them is the subject of considerable controversy and debate in the UK and elsewhere (see, for example, Yim 1992, Clarke *et al*, 2000, Smith and Sanchez 2003). This issue has brought to the fore concerns relating to the encouragement of traffic growth on local road networks with all the attendant negative externalities of emissions, congestion, higher accident rates etc. Concern has also been pointed to possible deleterious effects upon traditional city centre trading vitality and the flow of investment into the physical fabric of traditional inner urban shopping centres. The change in consumer trends toward food superstores in parts of the UK has also been highlighted as being a central element in the development of 'food deserts' (Wrigley 2002, Guy and David 2004). Such phenomena are locations where access to food shops is made difficult for low income (and low mobility) households as a by-product of the lack of

[#] **Acknowledgements**
Support from the UK Economic and Social Research Council (Project Number PTA-033-2004-00035) is gratefully acknowledged, as is permission from the JMP Consultants Limited and the UK TRICS Consortium, to use the TRICS database. Thanks goes also to John Nankervis for comments received on an earlier draft of this paper.

nearby retail outlets willing and able to meet a demand for healthy, affordable food. Thus, the problems identified with food superstores relate both to their role in traffic generation and to their role in the development of a less sustainable urban spatial structure. Arguably a key element of such a structure is the increasing concentration of food provision towards food superstores, since smaller retailers may often not be able to compete on cost grounds for mobile consumers (a significant fraction of which are needed in order to retain economically viable retail customer bases). Moreover, such concentration may further perpetuate less sustainable lifestyles and also militate against transport-poor, lower income households, who become much more reliant on higher cost, less healthy food from nearby 'convenience' stores (see Wrigley 2002; Wrigley *et al* 2002 and Cummins and Macintyre 2002 for a detailed consideration of the 'food deserts' issue and some policy responses).

This study estimates a semi-parametric food superstore trip generation model, applying a bootstrap algorithm to re-sample the data matrix. It is accompanied by some sensitivity analysis to assess the robustness of the results and the policy implications drawn from them. The results contribute to the body of evidence by means of the estimation of a trip attraction model specifically for food superstores in the United Kingdom. The data held within the Trip Rate Information Computer System (TRICS) database from the period 1986-2003 is augmented with the UK Census and other data sources, to provide a rich dataset from which the model estimates are generated.

At a conceptual level, this study operates in a general transport and development framework that is broadly consistent with a *'predict and manage'* strategy as opposed to the '*predict and provide*' approach of historic notoriety (see Berry 1960), or the more contemporary focus on 'Transport Demand Management' (TDM) solutions (see for example, Meyer 1997, 1999). Given that much of the literature generally neglects the potential synergies between these schools of thought, an eclectic approach is taken here. Trip attraction models have been developed for various land-uses, including food superstores and similar enterprises elsewhere (see, for example, Vickerman and Barmby 1984, Goldner and Portugal 2002, Tan and Fan 2003). Yet it is accepted that the growth of food superstores may well, in part, be a product of the diffusion of a more car-orientated culture and that the relationship between land use, spatial structure and mobility is undoubtedly complex and varied (see, for example Mackett 1993; Badoe and Miller 2001; Meurs and Haaijer 2001).

The paper is organised as follows. Section 2 outlines the modelling strategy developed for this application, followed by a description of the data in section 3. The following section presents and discusses the results, with section 5 offering some concluding remarks.

## 2. Modelling Strategy and Estimation

It is widely understood that the demand for transport primarily stems from the desire to participate in activities or purchase goods (see, for example, Ettema and Timmermans 1997, for a more comprehensive discussion on the role of 'activity' in transport analysis). Transport is a necessary component in the production, delivery and consequent consumption of any good or service, thus its demand is nearly always derived from the demand for these other goods (Ortuzar and Willumsen 1990, Hensher and Button 2000 and Hibbs 2003).

An Activity-Based Trip Model is therefore specified under the same auspices as a standard demand relationship for consumer behaviour, derived from standard micro-economics so that the demand for a good or service is determined by the desire and ability to purchase the good (e.g. characteristics of the good, and individual capacity to consume), as well as interactions with substitute markets/goods and/or effects. This can be summarised as follows:

$$T_{m,o.d} = f\left(E_o, SE_o, P_d, G_d\right) \tag{1}$$

$$m \ni M, o \ni O, d \ni D$$

Where ‘M’ is the set of travel modes, ‘O’ is the set of origins and ‘D’ is the set of destinations.

Equation (1) implies that the desire to travel to a site (by any chosen mode of travel, for instance cycling, walking or taking public transport), is determined by factors that influence the ability and/or wish to partake in (or consume) the activity (or service) which that site offers. In particular, ‘E’ represents the economic characteristics of the local population, thus capturing the ability of individuals within that area to consume. ‘P’ is site specific attributes, which may be considered as features of that site which may serve to attract more trips. The variable ‘G’ suggests that wider geographic information may have some bearing on trip levels including site accessibility, and potential resistance offered by the existing geography. Finally ‘SE’ are socio-economic characteristics, which reflect (in part) lifestyles, and consequently, consumption choices within a given area. Previous multivariate trip generation models adopting similar relationships have been advocated for instance in Brady and Betz (1971), Hensher and Dalvi (1978) and Washington (2001). Although the inclusion of the physical characteristics of a site is not often introduced, probably due to the nature of the datasets used in previous work. The dependant variable ‘T’ in this instance, is the one way hourly average traffic flow.

For empirical tractability a number of simplifying assumptions are required to ensure that activity specific elasticties are revealed, as follows;

- ‘M’ is constrained to passenger vehicle traffic only.

- ‘O’ is not known with certainty, and is therefore assumed to be a function of the surrounding areas characteristics.

- ‘D’ is constrained to one type of destination, in this instance food superstores. i.e. the model estimates the levels of trips for only one individual type of activity.

The traffic count data is typically from established sites (i.e. not new developments), it is therefore assumed that the general customer base has levelled out after any initial opening 'boom', where the customer levels are essentially stable. Furthermore, given the model features no dynamic elements there is no distinction between long-run and short-run effects. It is not difficult to formulate hypotheses to suggest that initial opening levels could be (and are likely to be) higher OR lower than the estimates produced by the static model. Although initial opening levels should, on average, converge towards the model estimate over time. The estimated model does not differentiate between trips to the store from linked (or chained) trips or single purpose journeys. The framework therefore focuses on the fundamentals of the determinants of trip 'attractions' to a particular site.

Given the discussion, the results of this study are based upon the following general specification;

$$\begin{aligned}
\ln FLOW_i = {} & \beta_0 + \beta_1(\ln CAR_i) + \beta_2(\ln ACCESSIBILITY_i) + \beta_3(DATE_i) + \\
& \beta_4(\ln FLOORSPACE^*) + \beta_5(\ln RESISTANCE_i) + \beta_6(\ln SOCIOECON^*) + \\
& \beta_7(\ln EMP_i) + \beta_8(\ln PARKING_i) + \beta_9(PFS_i) + \beta_j(LU_{m,i}) + \\
& \beta_{17}(ECONOMY_i) + \beta_{18}(SAT_i) + \beta_{19}(SUN_i) + \beta_{20}(MON-THURS_i) + \mu_i
\end{aligned} \qquad (2)$$

For $j=(m+8)$ where $m=1,\ldots,7$ and $i=1,\ldots,N$. $N$ is the number of observations in the dataset.

Where

| | | | |
|---|---|---|---|
| Where | FLOW | = | The average hourly flow of passenger vehicle traffic to site $i$. |
| | CAR | = | Car ownership in the area of site $i$. |
| | ACCESSIBILITY | = | Public service provision at site $i$. |
| | DATE | = | Date of the survey for site $i$. |
| | FLOORSPACE* | = | A measure of floor-space for site $i$. |
| | Two alternatives are available; | | |
| | (I) GFA | = | Gross Floor Area. |
| | (II) RFA | = | Retail Floor Area. |
| | RESISTANCE | = | Proxy for competition. |

| | | |
|---|---|---|
| SOCIOECON* | = | Offers representations of the socio-economic characteristics of the area surrounding site $i$. |
| Three alternatives are available; | | |
| (a) AVHS | = | Average household size. |
| (b) AVLH | = | Proportion of households which are large. |
| (c) AVHC | = | Proportion of households with children. |
| EMP | = | Employment levels in the area of site $i$. |
| PARKING | = | On site parking at site $i$. |
| PFS | = | Petrol pump facilities. |
| LU1-LU7 | = | Land-zone indicators. |
| ECONOMY | = | Quality proxy. |
| SAT | = | Day identifier. |
| SUN | = | Day identifier. |
| MON-THURS | = | Day identifier. |
| ln | = | Natural logarithm. |
| $\beta_0$ | = | Constant or intercept term. |
| $\beta_{1,\ldots,20}$ | = | Estimated slope coefficients. |
| $\mu_i$ | = | Error term. |

and *a-priori* the following are expected for the slope coefficients, $\beta_0>0$, $\beta_1>0$, $\beta_2<0$, $\beta_3>0$, $\beta_4>0$, $\beta_5>0$, $\beta_6<0$, $\beta_7>0$, $\beta_8>0$, $\beta_9>0$, $\beta_{17}<0$, $\beta_{18}>0$, $\beta_{19}<0$, $\beta_{20}<0$. No prior expectations are made on the coefficients of the land zone dummies ($\beta_{10} \ldots \beta_{16}$).

Given all the continuous variables are in natural logs, this log-linear equation is an activity-based model of derived demand and the estimated coefficients represent constant elasticities. The error term ($\mu_i$) is likely to be influenced by random factors but possibly also by some potential direct-demand (which is not measurable) for travel (Mokhtarian, Salomon and Redmond (2001)), store offers etc. This is not likely to be a source of (noticeable) bias, as people who make journeys for the direct pleasure derived from driving, are far less likely to couple that journey with other activities. The decision to stop at a supermarket would be considered an 'outlier' (or rare event) in a journey where the primary activity is driving itself.

Previous literature offers mixed views as to which is the appropriate floor-space measure to use (gross or retail), mainly surrounding the relevance of the use of gross floor area which includes warehouse space. Retail floor area, on the other hand, arguably captures more directly the area of business operations (in terms of retail sites) that customers come into contact with. Ortuzar and Willumsen (1990, pp97-98) provide arguments in support of both approaches. Dasgupta, Raha and Sharman (1996) and Goldner and Portugal (2002) however, restrict their analysis to gross floor area. Tan and Fan (2003) in a study of peak hour trip rates to office and retail developments find mixed evidence mostly in favour of gross floor space. Given both the gross and retail are available, both are used in the estimation, denoted by models I and II.

The expected negative coefficient on the SOCIOECON* variables may not be so intuitively obvious without some clarification. The expectation is that household levels of economies of scale and scope exist such that "…the cost per person of maintaining a given material standard of living may fall as household size rises…" (Nelson 1988, p1301). Following Lazear and Michael (1980), these economies arise due to the nature of certain goods used within the physical confines of the house, as certain goods cannot avoid being 'public goods', i.e. goods that provide benefits to everybody (positive externalities), not just those who purchase them. Lazear and Michael (1980), term these goods 'family goods'. Examples include lighting in hallways and locks on shared entrances. It is not feasible to exclude people from the use of such goods.

Such economies of scale in a non-nuclear household could be considered as economies of scope. These arise when two independent parties (i.e. two separate individuals living under one roof, as in shared accommodation for professionals/students etc.) are able to pool together their resources and reduce the marginal cost faced by each in achieving the same level of utility. Lazear and Michael (1980) present an example-using door locks, though characterise this purely as a scale economy. Yet this scale economy is only realised through the recognition of scope. Thus in the case of trip making behaviour, two individuals within a shared house could car share and thus utilise space that may previously have gone un-used (assuming individual shopping load constraints are not biting). This arises at a cost which is lower to each individual (presuming costs are shared). Such arrangements could still provide each individual with the same amount of car-borne shopping resources that they would have enjoyed had they not pooled together.

Nelson (1980) places an empirical value on the household economies of scale that were achieved in the US for food shopping. The results revealed with a strong degree of significance, that for households choosing to pool their resources 2 people can essentially live for the price of 1.19 people. The concept and existence of scale and scope economies would arise automatically with growing nuclear households, and would be a conscious decision within non-nuclear multiple person households. Following this reasoning, it is contended that the number of trips to a food store are negatively related to household size[1]. Further, larger households will also likely exhibit more diverse characteristics in their modal choices, i.e. as household size increases, the probability that one of the household members will prefer a non-car mode of travel to a food store also increases.

In relation to the variable 'SOCIOECON*' from equation (3), three alternative measures (average household size (AVHS), average large households (AVLH) and average households with children (AVHC)) are considered in the empirical phase, that give differing estimates for policy variables based on representations of the surrounding area demographic

---

[1] Furthermore, regardless of whether or not parking restrictions exist in a residential area (which in many, particularly urban, cases they do), there is typically finitely constrained space to store/park a car. As household size increases such spatial constraints may ultimately generate a stronger barrier to increasing car trips that could otherwise have been expected to emerge as a direct consequence of multiple car ownership.

decomposition. These are represented by a, b and c respectively in the results section.

## 3. Data

The data used in this study has been extracted from three sources, namely the TRICS database (Version 2004b), NOMIS 1986-2003 and the 2001 UK Census. The TRICS database[2] provides an unbalanced pseudo-panel (see Baltagi 2001 or Mckenzie 2001) of traffic counts and site characteristics for various land use types (including food superstores) primarily with a view to informing planning policy and development control at the local level. The database is typically consulted to provide a guide of the expected traffic flows associated with any given new development and thus inform planning decisions, junction improvements and transport network management.

The TRICS database provides in the main, cohorts for one time period only. The application of panel estimators may be feasible with a redefinition of the dataset (so as to produce cohorts over averages of groups over time[3]), however, this could lead to a loss of accuracy, due to the averaging which would be necessary (i.e. the imposed generality), and subsequently may introduce bias into the results. Consequently, all models are estimated by cross-sectional OLS.

To enhance the TRICS database, it is augmented with data from the government sources, namely the NOMIS archive and the 2001 Census of Population, thus culminating in a rich database. The NOMIS archive provides information pertaining to labour market statistics for local and national areas, combining data from the Labour Force Survey (LFS), Claimant Count, Annual Business Inquiry (ABI), New Earnings Survey (NES) and the 1991 UK Census of Population with information from 1970

[2] Further details and discussion of the TRICS database may be accessed via http://www.trics.org.uk

[3] The cohorts could be redefined, such that averages over groups (e.g. group 1 is the average values for all ‘mainstream’ superstores surveyed on a Friday) are created for sequential time periods. Multiple groups must be defined.

through to the present date, although only data from 1986 (onwards) is extracted for the purpose of this study.

The 2001 UK Census of Population provides further data on socio-demographic and economic characteristics for households, which is consistent with the data extracted from the NOMIS database. The data collection processes used in the 1991 and 2001 Censuses of Population are not perfectly consistent, resulting in a need to aggregate data together to create consistent variables, however no discernable difficulties are encountered in doing so. Variables from the Census (and NOMIS) data sources are considered as strong proxies (as opposed to more precise or exact values) due to the irregularity at which the data is collected.

The variables considered in the empirical phase of this analysis are described and defined in Table 1. The specific land use dummy variables introduced in this table represent food superstores in specific types of geographical zones or locations as indicated in Planning Policy Guidance Note 13 DETR (2001).

A dummy (*ECONOMY*) is used to separate the sites into the more 'mainstream' food superstores (e.g. Tesco or Sainsburys), and the less 'mainstream' food superstores which feature in much smaller chains and/or have a discount or low cost market orientation; taking the value 1 if the site is a 'less mainstream' store. *EMP* – is the average level of household employment in the surrounding area. Calculated as;

$$EMP_i = \left( \frac{(p_i - c_i)}{h_i} \right)$$

Where *EMP* = average employment per household, $p$ = population of working age, $c$ = claimant count rate and $h$ = number of households in the area, $i$ is the site identifier. This variable captures the average household's ability to consume/purchase goods and services, where it is assumed there is a direct and positive correlation between household employment and that household's income level.

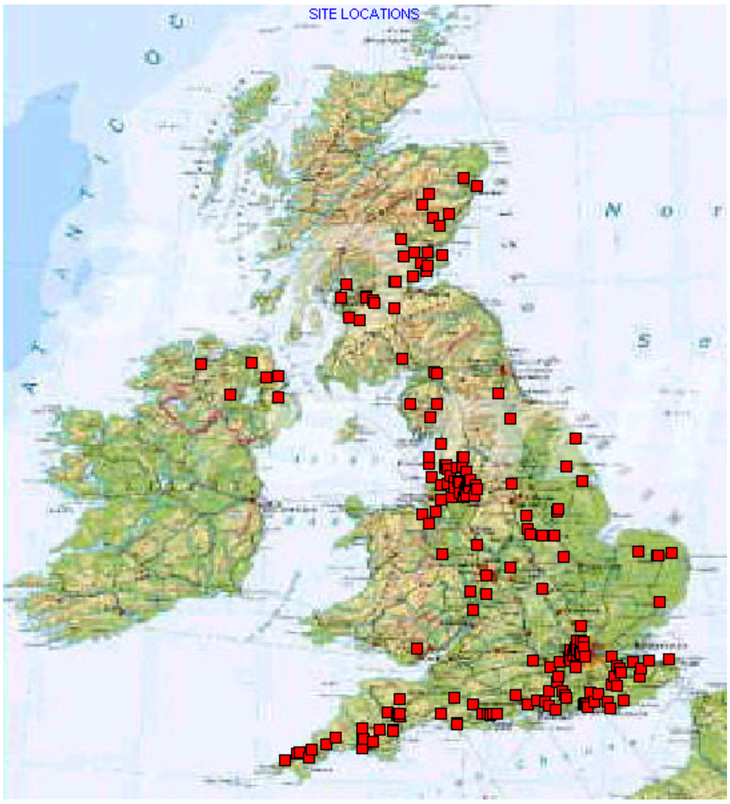
**Table 1: Source and Description for the Variables used in the Statistical Analysis**

| Variable | Description | Data Source |
|---|---|---|
| FLOW | *Average hourly flow of cars* | TRICS |
| CAR | *Average household car ownership for the area* | TRICS |
| ACCESSIBILITY | *A measure of public service accessibility (bus services) at the site* | TRICS |
| DATE | *Date variable* | TRICS |
| GFA | *Gross Floor Area* | TRICS |
| RFA | *Retail Floor area* | TRICS |
| RESISTANCE | *A measure of spatial resistance (via proximity to nearest similar competitor)* | TRICS |
| AVHS | *Average Household size* | CENSUS/NOMIS |
| AVLH | *Proportion of Large Households in the area* | CENSUS/NOMIS |
| AVHC | *Proportion of households with children in the area* | CENSUS/NOMIS |
| EMP | *Average household employment for the area* | CENSUS/NOMIS |
| PARKING | *Total parking provision for the site* | TRICS |
| PFS | *Does the site have a petrol station (1=yes)* | TRICS |
| LU1 | *Commercial zone* | TRICS |
| LU2 | *Edge of centre* | TRICS |
| LU3 | *Edge of town centre* | TRICS |
| LU4 | *Freestanding* | TRICS |
| LU5 | *Industrial zone* | TRICS |
| LU6 | *Neighbourhood zone* | TRICS |
| LU7 | *Suburban area* | TRICS |
| Base | *Town centre* | TRICS |
| ECONOMY | *Is the Store 'less mainstream' (1=yes)* | Author Specified |
| SAT | *Dummy variables identifying the day of week* | TRICS |
| SUN | | TRICS |
| MON-THURS | | TRICS |

*AVHS, AVLH and AVHC* as discussed above, are socio-economic variables reflecting population composition characteristics derived from Census and NOMIS data. *AVHS* measures the average household size, *AVLH* is a measure of the proportion of large households in the area (defined as households with three or more people) and *AVHC* is the proportion of households in the area with children.

*RESISTANCE* – measures the distance to the nearest competitor (in kilometres). This variable proxies the generalised cost of travelling between the surveyed site and the nearest similar site, consequently acting as an indirect attraction factor.

**Figure 1: Locations of Food Superstore Sites in the TRICS Database (1986-2003)**




The other variables included into the analysis are; PFS – which is a dummy indicating whether a site has a petrol station or not (i.e. an attraction factor for those who do drive cars, as they can buy their fuel also). CAR – which is a measure of the level of average household car ownership within the site area. PARKING, measures the number of parking spaces at a site, reflecting the ease at which a potential vehicle traveller can stop and shop.

A number of dummies are also included to identify the day of week of the traffic count at the site (MON-THURS, SAT and SUN). The base weekday is Friday, which is incorporated into the intercept term. Finally, the variable ACCESSIBILTY captures the level of public service provision at each site.

Figure 1 provides an illustration of the geographic extent of the observations around the UK. The coverage is dispersed around the country though largely concentrated, as is to be expected, around major urban conurbations. In particular the site data is concentrated within the North West of England, the Central Lowlands of Scotland and the Southern Coastal Belt of England.

**Table 2: Composite Dataset Descriptive Statistics**

| *Continuous variables* | | | | | |
|---|---|---|---|---|---|
| variable | Number | mean | Std. Dev. | Min | Max |
| CAR | 201 | -0.136 | 0.320 | -1.386 | 0.262 |
| ACCESSIBILITY | 201 | 3.882 | 0.633 | 2.303 | 4.500 |
| DATE | 201 | 1993.269 | 4.247 | 1986 | 2003 |
| GFA | 201 | 8.565 | 0.400 | 7.097 | 9.218 |
| RFA | 199 | 8.006 | 0.385 | 6.783 | 8.817 |
| RESISTANCE | 201 | 1.120 | 0.866 | -1.386 | 3.367 |
| AVHS | 201 | 0.861 | 0.527 | 0.759 | 0.977 |
| AVLH | 201 | -1.109 | 0.137 | -1.406 | -0.846 |
| AVHC | 201 | -1.322 | 0.142 | -1.639 | -1.003 |
| EMPLOYMENT | 201 | 0.190 | 0.205 | -0.185 | 0.595 |
| PARKING | 201 | 6.107 | 0.504 | 4.489 | 6.908 |

| *Dummy Variables* | | | | |
|---|---|---|---|---|
| Variable | Number | Mean | Std. Dev | Median |
| PFS | 201 | 0.507 | 0.501 | 1 |
| LU1 | 201 | 0.099 | 0.099 | 0 |
| LU2 | 201 | 0.358 | 0.481 | 0 |
| LU3 | 201 | 0.020 | 0.140 | 0 |
| LU4 | 201 | 0.054 | 0.228 | 0 |
| LU5 | 201 | 0.035 | 0.184 | 0 |
| LU6 | 201 | 0.129 | 0.336 | 0 |
| LU7 | 201 | 0.194 | 0.396 | 0 |
| ECONOMY | 201 | 0.094 | 0.293 | 0 |
| SATURDAY | 201 | 0.313 | 0.465 | 0 |
| SUNDAY | 201 | 0.144 | 0.352 | 0 |
| MON-THURS | 201 | 0.104 | 0.306 | 0 |

Table 2 provides the descriptive statistics for the data used in the empirical phase of the study.

390 cases of food superstore developments were originally extracted from the TRICS database but depending on the model specification, approximately 190 cases were removed (varying by model specification), as many of these sites featured multiple instances of missing values for some important explanatory variables. There are no *a priori* expectations that the distribution of missing values were associated with any systematic biases, furthermore Shapiro-Wilk normality tests (see Shapiro and Wilk, 1965) on the dataset used in the final analysis reveals that no bias was imposed.

## 4. Results

Given the two floor space variables and the three representations of the areas socio-economic characteristics, six versions of equation (2) were initially estimated by OLS (see Table 3).

**Table 3: Food Superstore Trip Generation Model Estimates (OLS)**

| | I | | | II | | |
|---|---|---|---|---|---|---|
| | **(a)** | **(b)** | **(c)** | **(a)** | **(b)** | **(c)** |
| Intercept | 1.572* | -0.64 | -0.956 | 1.228 | -0.7 | -0.912 |
| CAR | 0.202*** | 0.198*** | 0.165*** | 0.214*** | 0.207*** | 0.178*** |
| ACCESIBILITY | 0.138*** | 0.127** | 0.142*** | 0.141*** | 0.133*** | 0.147*** |
| GFA | 0.300** | 0.316** | 0.352*** | | | |
| RFA | | | | 0.342** | 0.348** | 0.374*** |
| RESISTANCE | 0.084*** | 0.083*** | 0.088*** | 0.103*** | 0.102*** | 0.107*** |
| AVHS | -1.757*** | | | -1.581*** | | |
| AVLH | | -0.623*** | | | -0.539*** | |
| AVHC | | | -0.567*** | | | -0.467*** |
| EMP | 0.497*** | 0.403*** | 0.454*** | 0.431*** | 0.344*** | 0.381*** |
| PARKING | 0.343*** | 0.333*** | 0.311*** | 0.345*** | 0.340*** | 0.324*** |
| PFS | 0.134*** | 0.129** | 0.140*** | 0.149*** | 0.144*** | 0.154*** |
| LU1 | 0.381** | 0.362** | 0.438** | 0.431** | 0.420** | 0.492** |
| LU2 | 0.100 | 0.107 | 0.103 | 0.095 | 0.103 | 0.102 |
| LU3 | -0.067 | -0.049 | -0.05 | -0.016 | 0.001 | 0.006 |
| LU4 | 0.281*** | 0.267*** | 0.261*** | 0.242*** | 0.231** | 0.227** |
| LU5 | 0.102*** | 0.073*** | 0.117*** | 0.087*** | 0.070** | 0.113** |
| LU6 | 0.152* | 0.158* | 0.160* | 0.159* | 0.167** | 0.171** |
| LU7 | 0.214*** | 0.207*** | 0.223*** | 0.173** | 0.168** | 0.180*** |
| ECONOMY | -0.336*** | -0.323*** | -0.313*** | -0.311*** | -0.302** | -0.294** |
| SAT | 0.029 | 0.032 | 0.029 | 0.032 | 0.034 | 0.032 |
| SUN | -0.657*** | -0.650*** | -0.651*** | -0.654*** | -0.649*** | -0.649*** |
| MON-THURS | -0.258*** | -0.256*** | -0.258*** | -0.251*** | -0.247*** | -0.247*** |
| Observations | 201 | 201 | 201 | 199 | 199 | 199 |
| RMSE | 0.290 | 0.291 | 0.292 | 0.291 | 0.292 | 0.293 |
| Adj R-squared | 0.6943 | 0.6927 | 0.6907 | 0.6895 | 0.687 | 0.6844 |

P-values; ***=1%, **=5%, *=10%

Specification tests on these initial estimates revealed non-normality and heteroskedasticity, thus making it problematic to make accurate statistical inference from the results given the possible biased standard errors and rendering standard t-tests ambiguous. As a result, a bootstrap algorithm was applied for estimating the models coefficients, thereby allowing the assumption of normality to be dropped. The bootstrap algorithm used, is derived from a combination of a non-parametric estimation of the sample distribution and standard (model based) OLS parametric inference, thus resulting in a semi-parametric regression model (see the Appendix A1). Based on this semi-parametric approach, the variable DATE was always not significantly different from zero and hence omitted. This gives the six estimated models summarised in Table 4. The significance levels reported in this Table are derived from one sample Achieved Significance Levels (ASL's), see Appendix A2 and are subsequently exact as opposed to the standard (approximate) p-values reported in Table 3.

**Table 4: Food Superstore Trip Generation Model Estimates (Bootstrap)**

| | I | | | II | | |
|---|---|---|---|---|---|---|
| | (a) | (b) | (c) | (a) | (b) | (c) |
| Intercept | 1.539* | -0.633 | -0.949 | 1.230 | -0.665 | -0.883 |
| CAR | 0.202*** | 0.199*** | 0.166** | 0.214*** | 0.208*** | 0.179*** |
| ACCESIBILITY | 0.141*** | 0.129** | 0.144*** | 0.143*** | 0.135*** | 0.149*** |
| GFA | 0.303** | 0.320** | 0.355*** | | | |
| RFA | | | | 0.339** | 0.345** | 0.373*** |
| RESISTANCE | 0.084*** | 0.083*** | 0.088*** | 0.102*** | 0.101*** | 0.107*** |
| AVHS | -1.724*** | | | -1.557*** | | |
| AVLH | | -0.613*** | | | -0.534*** | |
| AVHC | | | -0.555*** | | | -0.458*** |
| EMP | 0.494*** | 0.402*** | 0.454*** | 0.432*** | 0.347*** | 0.384*** |
| PARKING | 0.338*** | 0.327*** | 0.306*** | 0.343*** | 0.337*** | 0.321*** |
| PFS | 0.136*** | 0.130*** | 0.141*** | 0.150*** | 0.146*** | 0.155*** |
| LU1 | 0.334*** | 0.317*** | 0.382*** | 0.375*** | 0.366*** | 0.427*** |
| LU2 | 0.102 | 0.109 | 0.105 | 0.096 | 0.105 | 0.103 |
| LU3 | -0.067 | -0.050 | -0.052 | -0.019 | -0.003 | 0.001 |
| LU4 | 0.280*** | 0.266*** | 0.260** | 0.241** | 0.230** | 0.225** |
| LU5 | 0.109 | 0.080 | 0.124 | 0.096 | 0.077 | 0.120 |
| LU6 | 0.153* | 0.158* | 0.160* | 0.161* | 0.168** | 0.172** |
| LU7 | 0.214*** | 0.207*** | 0.222*** | 0.172** | 0.167** | 0.178** |
| ECONOMY | -0.341*** | -0.329*** | -0.318*** | -0.315*** | -0.307** | -0.298** |
| SAT | 0.029 | 0.032 | 0.030 | 0.032 | 0.034 | 0.032 |
| SUN | -0.658*** | -0.651*** | -0.652*** | -0.656*** | -0.650*** | -0.650*** |
| MON-THURS | -0.258*** | -0.256*** | -0.257*** | -0.251*** | -0.247*** | -0.246*** |
| Observations | 201 | 201 | 201 | 199 | 199 | 199 |
| RMSE | 0.276 | 0.276 | 0.277 | 0.276 | 0.277 | 0.278 |
| Adj R-squared | 0.694 | 0.693 | 0.691 | 0.689 | 0.687 | 0.684 |

Achieved Significance Level; ***=1%, **=5%, *=10%

The observed difference in the results presented in Tables 3 and 4 are indicative of the bias associated with using standard OLS in the presence of non-normality (see Table A1 in the appendix for the observed bias). Through the assumptions imposed by the application of the bootstrap process, the coefficients given in Table 4 can be considered to be the true population parameters (given that the population is restricted to food superstores), and therefore these parameters are the ones which are used for the subsequent inference purposes in the remainder of this paper.

Interpreting the results in the presence of dummies requires a degree of care; base interpretation should initially be done where all dummy variables are set equal to zero. Following this, base interpretation then identifies the level of average hourly passenger-vehicle trips on a Friday, attracted to a 'standard' food superstore without a petrol filling station, positioned in a town centre zone in the UK.

The results given in Table 4 (aswell as those in Table 3) are in accordance with *a-priori* expectations, though the coefficient on the public transport accessibility variable warrants further discussion. In terms of the subsequent discussion and analysis, focus remains upon model 'I(a)', as this model reflects the 'average' household in the UK, and the floor-space measure which results in a marginally better specified model (in terms of R-squared and Root Mean Squared Error (RMSE)). Due to the robustness of the results, across the alternative specifications, the following discussion and sensitivity analysis does not differentiate between the model specifications but rather focuses upon the qualitative implications of the results.

The public transport variable provides what may at first be considered a counterintuitive result. A positive coefficient on this variable is observed, implying that as public service provision increases, so do trips to that site by car. Of course, the positive value may simply identify a correlation between bus provision and large business centres. Large superstores may attract higher rates of public service provision as the providers 'realise' that the large superstores are likely to offer a larger (and more profitable) customer base than the smaller stores. If bus services were indeed being sucked towards more profitable food superstore sites and away from serving other more traditional shopping areas, then this can be considered as contributing to the development and reinforcement of 'food deserts'.

Although there exists a desire to encourage increased levels of public transport use (thereby reducing some notion of the average environmental cost per customer), the results do suggest that less provision to/from food superstores is associated with lower car traffic levels. A reasonable simple explanation could be that public transport service providers may not be offering adequately desirable service levels at the origin end of food-shopping journeys (whether to superstores or the town/city centre).

The rest of the coefficients follow standard micro-economic theory of demand. It is observed that floor space, distance to the nearest food superstore competition, household type and size, parking provision and the inclusion of a petrol filling station at a site all return positive coefficients. These variables generally reflect the ability to substitute the store for alternative shopping centres, and the costs involved when doing so.

A date variable was employed in initial specifications in an attempt to capture any exogenous growth effects over time not represented by the other variables in the model. However, they were insignificant in each run, suggesting that, although over recent years there has been increasing growth in the number of superstores being used by food retailers (as opposed to smaller store sizes), this growth in floor size has not created any shifts in the underlying behaviour of shoppers.

The results are observed to be extremely robust across the different specifications. The stability of the coefficients with changes in (qualitative) model specifications is encouraging and the robustness suggests that the general theoretical framework is a generally sound reflection of the observed relationships for vehicle trip-making behaviour at food retail sites. The only real notable change in the estimated parameters being those for the alternative measures of a site area's demography. These observed changes are intuitively consistent given the qualitative explanation of each; the measure AVHS is seen to have a coefficient of a greater magnitude than both AVLH and AVHC in absolute terms. The AVLH and AVHC measures essentially capture the nature of households that have greater restrictions imposed upon their time. The lifestyles may be typified by fewer trips to a food superstore for larger grocery needs (e.g. having a family monthly food run). This applies as much to large households as it does to households with children, as the family routines may be continued into young adult life within the family home. Alternatively, people in large shared houses may choose to car-share to reduce cost (and perhaps make shopping another semi-social event).

It should be noted that the results provide some response to the debate over which is the appropriate floor-space measure to use in the trip attraction models. The selection criteria discussed in the previous paragraph reveal that for food superstores, GFA is favoured. This suggests that, at least in terms of food shops, the attractiveness of a site is determined by their ability to meet demand, not just the goods they sell. Using RFA as a determinant reduces the explanatory power of the model by approximately 0.5 percentage points (i.e. R-squared for I(a) = 0.694, R-squared for II(a) = 0.689) in all cases. This change is not large enough to provide irrevocable evidence, but does imply that consumers consider the warehouse and administrative space in their trip making decision. This finding is consistent with that of Tan and Fan (2003) who find statistical evidence favouring the use of GFA at retail sites.

Bonsall *et al* (1977), undertake a sensitivity analysis on key significant variables in order to help identify the most effective policies, based on the reality of their application. This is based on the understanding that as the variables were found to be significant, the results of policy changes can be expected to have significant effects too. Two types of parameter are considered; firstly are those normally considered fixed in nature (e.g. average car ownership within an area), the second type of parameter represents more feasible (in the short to medium term) policy alternatives that can be enacted upon in a reasonable time frame, for example changes in the level of public transport provision. It is hard to conceive that any of the variables analysed in this study must be fixed in the long-run, though it is still prudent to consider that some of the variables included in this analysis will be 'more fixed' than others.

The sensitivity analysis in Figure 2 is based upon the fitted equation:

$$\ln \widehat{FLOW}_i = \hat{\beta}_0 + \hat{\beta}_\lambda X_{i\lambda} + \hat{\beta}_\eta \bar{X}_{i\eta}$$

With all independent variables fixed at the mean (i.e. all $X_\eta$, where $\lambda \neq \eta$), other than the policy variable under consideration (i.e. $X_\lambda$). From this the predicted average hourly trip value ($\hat{Y}_i$) is computed (for model I(a)). This is evaluated at regular intervals on the difference between the highest and lowest observed values for $X_\lambda$ within the dataset. i.e.

$$z\left(X_{\lambda,\max} - X_{\lambda,\min}\right)$$

Where $z$ is the evaluating parameter and ranges between zero and one at regular intervals (100 in this case, or 1% intervals of the difference).

**Figure 2: Effects of Key Parameter Changes upon Food Superstore Traffic Flow**

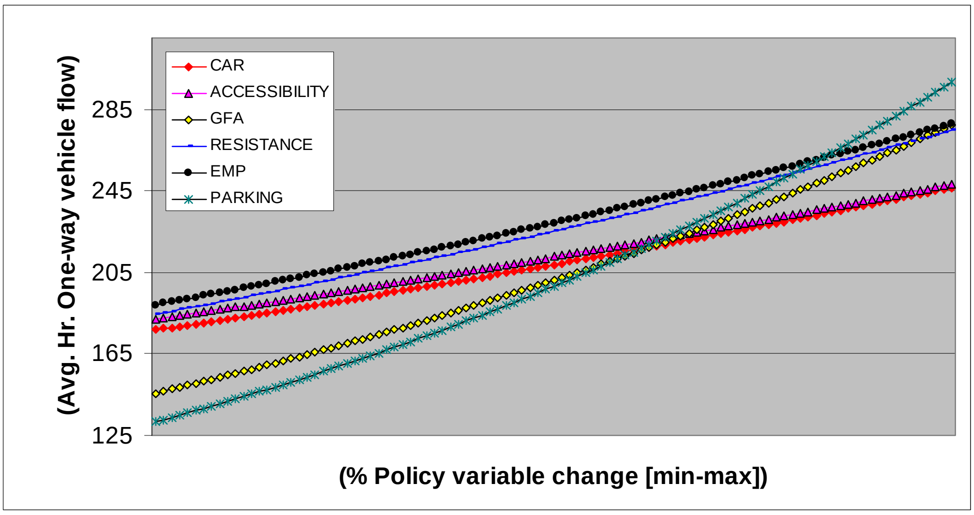



Figure 2 identifies several important features with respect to the implementation of policies (for the average household) based on model

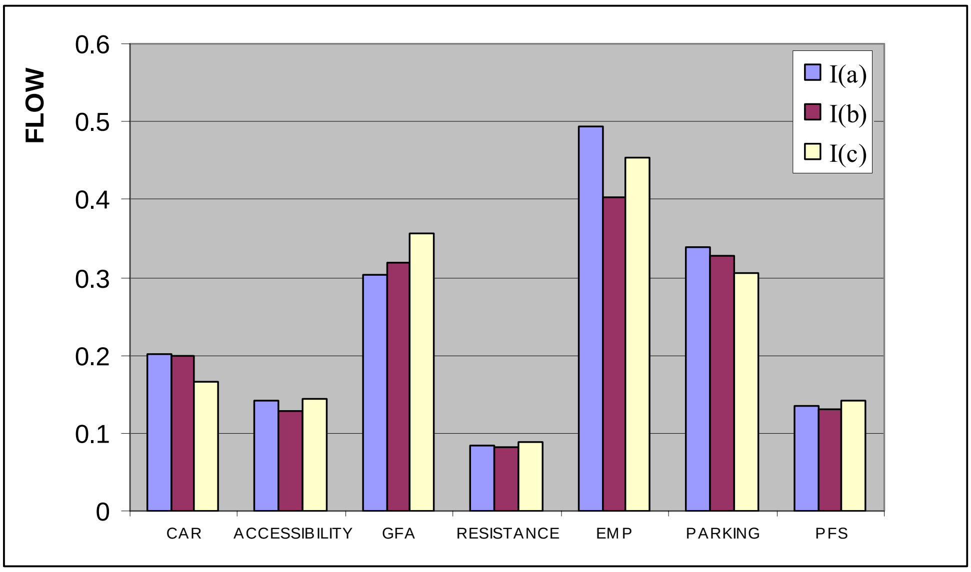
I(a)[4]. All relationships in this figure are non-linear and convex (as a result of the log-linear model) implying that a) as the actual value of the policy variables depicted in this graph decrease, so do the level of vehicle trips, b) the marginal effects of these policies decreases as actual traffic flows move towards the origin. This is true for all policy variables given in Figure 2, but most prominent for the PARKING variable. Analysing the results in this manner adds a new light to the estimated model, suggesting that in terms of feasible application, the most effective way of reducing vehicle trips to a supermarket is to decrease parking provision. The estimated model would suggest (via initial inspection of the coefficients) that the variable GFA is almost equally as effective as a policy variable as PARKING. However, the variability in the dataset (assuming this represents acceptable bounds from which policy makers can influence changes), implies that regardless of its equally large elasticity, imposing restrictions on floor-space) may not be the most feasible route to take in trying to reduce vehicle trip levels. A further example would be the comparison between EMP and CAR. The sensitivity analysis reveals them to be almost equally effective though the estimated coefficients would imply that changes in car ownership levels are markedly less effective in influencing car trip rates (see Table 3).

Figure 3 extends the sensitivity analysis to include the effects of policy variables on each of the qualitatively different models I(a), I(b) and I(c) (though subsequently only considering the constant elasticities). The horizontal axis on this graph is the natural log of the average hourly flow, whilst the vertical axis identifies the policy measure considered. Consistent relationship directions are observed, with only very subtle changes in magnitude relative to the 'base' model I(a). Changes in elasticities across the three models provide insights into the relative importance of policy measures on large households and households with children (compared to the average household). Policy makers and authorities could potentially use this information to focus policy implementation at specific demographic groups. This graph exemplifies the robustness of the results in Table 4. It

[4] Even though the focus is on model I(a), it is worth noting that the qualitative results are the same for I(b), I(c), II(a), II(b) and II(c).

reveals that in terms of policy based discussion, there is not much to justify a preference for models I(b) or I(c) over I(a), or indeed to think that, in the context of food superstores, that policy decisions can really be effectively focussed on specific socio-economic groups.

**Figure 3: Estimated Elasticities on Policy Variables, Models I(a), I(b) and I(c).**




The three variables with the strongest impact upon traffic flow were identified as household size/composition, parking provision and floor-space. The results indicate that decreasing the number of parking spaces will reduce the trips to a food superstore by car. The graph also highlights the relationship observed that smaller households seem to make more trips to 'smaller' food superstores (or at least a 1% change in GFA attracts more trips from large households and households with children than it does from the average household).

A further key advantage of the TRICS database is the ability to explore in greater depth the role of land-zone type, that is not commonly presented in the literature. Banister *et al* (1990) identify, among other things, that further empirical research is required on this issue. This study addresses that identified concern to some extent. Figure 4 shows the effects of land-zone placement upon trip rates to a mainstream food superstore (on a Friday) without a petrol filling station, when all parameters are held at their means.

**Figure 4: Estimated Elasticities on Land Zone Accessibility, Models I(a), I(b) and I(c)**

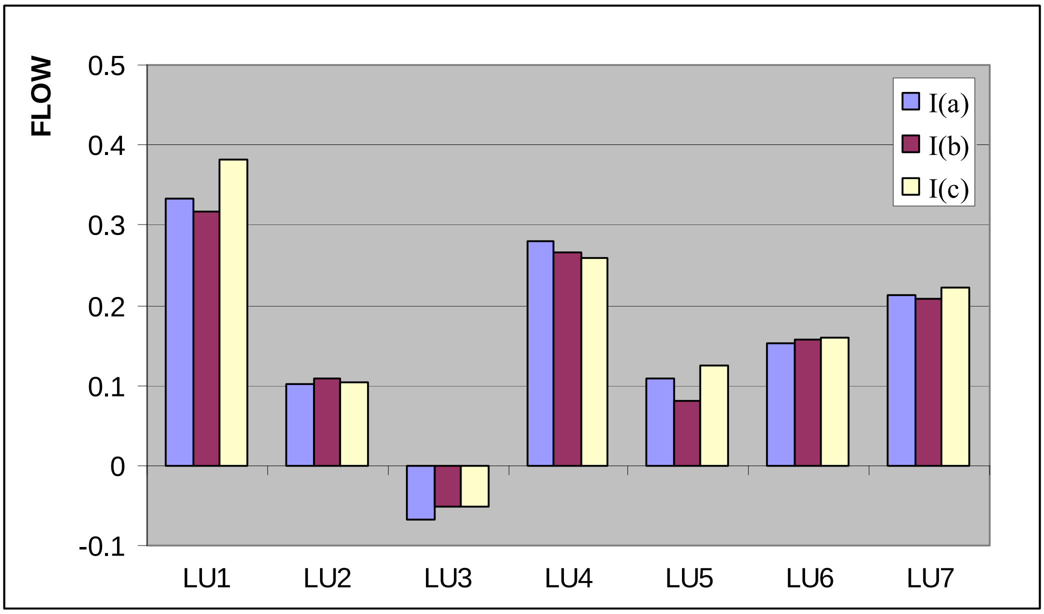



Figure 4 therefore identifies the perceived level of accessibility for each of the demographic groups considered in this study. As with the policy variables, the elasticities on accessibility remain stable across each of the models (in terms of direction and magnitude), suggesting that fundamental relationships hold true across all representations of the population considered. Households with children have a preference towards food superstores in industrial and commercial zones, whilst large households in general (including shared houses) prefer them less, at least in relation to trips by car, when compared to the average household. It should be noted that land uses 2 and 3 were found to be statistically insignificant and consequently take values of zero.

## 5. Concluding Remarks.

In the face of increasing political focus on the external effects arising from traffic growth, there is overlaid additional controversy in the role that food superstores play in creating and maintaining 'food deserts' in the UK. This paper offers insights on this phenomenon via a different route than previous related studies. These insights are made though the application of a semi-parametric regression-based trip generation model. This is applied to UK traffic count and site specific data (the TRICS database). This data source has not previously been used for econometric model estimation. Some sensitivity analysis and basic policy modelling has been undertaken on the final preferred (Ia) model to show the effects of changes to key parameters.

It is found that traffic to a given food superstore, *ceteris paribus*, increases with car ownership, parking provision, retail floor space, distance to the nearest food superstore competitor and, perhaps surprisingly, increased public transport provision. The latter effect is discussed in the light of a possible explanation linked to the 'food deserts' debate. Trips by car to a food superstore are also seen to decrease as average household size increases.

That said, shopping in a large food superstore is generally a time consuming experience, therefore the decision to engage in such an activity will be consciously influenced and weighed against the size of the shopping baskets that are being filled (i.e. the extent of the grocery needs). The utility derived from the ‘bundle of goods’ purchased at each visit will be weighed against the costs involved in obtaining those goods. Even in the face of current changes to the convenience store market, as a general rule, goods are cheaper in the food superstores than in other outlets for food (due to market power being exerted along the vertical supply chain, economies of scale and specialisation). Households with large shopping requirements (e.g. established ‘large’ households or households with children) are shown to be particularly attracted towards food superstores. Furthermore, due to their larger shopping load requirements, the relative ease in which the trip can be conducted may actually be enhanced by the use of a car (due to increased comfort and security, door-to-door service etc).

Food superstores may therefore seek to justify their extensive requirements for land in terms of customer parking provision. However, given city space limitations (and subsequently land price constraints too) it may be less feasible for superstores to locate at inner urban locations in a way that meets the objectives of their business model. That is to say, on the general assumption that these food retailers act largely as profit maximisers, they would rationally aim to facilitate greater customer access. As these empirical results suggest, the level of parking provision significantly influences this ability. Thus, food superstore developers may rationally voice a preference towards outer urban areas, where it is easier to satisfy their parking ratio requirements and thereby contribute to the genesis and maintenance of food deserts.

The scale and scope household economies identified through the SOCIOECON* variable(s), in connection with food superstore trips, suggests that communities with established larger household sizes will naturally be inclined towards large food superstores. Given the earlier arguments, these stores are increasingly less likely to be placed in inner urban locations. Thus, of specific relevance to communities characterised by car-owning, growing families, there is a statistically significant burden of evidence to suggest that with further development of food superstores aiming to serve them, food desert concerns will inevitably be perpetuated and accentuated for some other sections of the community. Such effects are intimately tied up with wider concerns about facilitating urban sprawl and achieving poor levels of progress in the development of more generally sustainable communities.

**References**


Badoe, D. and Miller, E. (2000), Transportation-land-use Interaction: Empirical Findings in North America, and their Implications for Modelling, *Transportation Research Part D* **5**, 235-263.

Baltagi, B. (2001) *Econometric Analysis of Panel Data*, John Wiley and Sons, UK.

Banister, D., Cullen, I. and Mackett. R. (1990), *The Impact of Land-Use on Travel Demand*, Working Paper 3, Planning and Development Research Centre, University College London.

Berry, D. (1960), Traffic Engineering, Section 2 in *The McGraw-Hill Engineering Reference Guide Series: Highway Design Reference Guide*, ed. Ross. S, (1988), McGraw-Hill, USA.

Bonsall, P., Champernowne, A., Mason, A. and Wilson, A. (1977), Transport Modelling: Sensitivity Analysis and Policy Testing, *Progress in Planning*, **7**(3).

Brady, C. and Betz, M. (1971), An Evaluation of Regression Analysis and the Gravity Model, *Journal of Transport Economics and Policy* **5**(1), 76-90.

Clarke, I., Horita, M. and Mackaness, W. (2000), The Spatial Knowledge of Retail Decision Makers: Capturing and Interpreting Group Insight Using a Composite Cognitive Map, Chain, *International Review of Retail, Distribution and Consumer Research* **10**(3), 265-285.

Cummins, S. and Macintyre, S., (2002), Food Deserts – Evidence and assumption in health policy making, *British Medical Journal*, **325**, 436-438.

Dasgupta, M., Raha, N. and Sharman, K. (1996), *Review of Trip Generation Studies*, TRL Report 190, Transportation Research Laboratory, UK.

Efron, B. and Tibshirani, R. (1993), *Monographs on Statistics and Applied Probability 57: An Introduction to the Bootstrap*, Chapman and Hall/CRC, USA.

Ettema, D. and Timmermans, H. (1997), *Activity Based Approaches to Travel Analysis*, Pergamon Press/Elsevier Science, UK.

Goldner, L. and Portugal, L. (2002), Trip Generation by Brazilian and Spanish Shopping Centres, *International Planning Studies* **7**(3), 227-241.

Guy, C. and David, G. (2004), Measuring physical access to 'healthy foods' in areas in areas of social deprivation: A case study in Cardiff, *International Journal of Consumer Studies*, **28**(3), 222-234.

Hensher, D., and Button, K. (2000) *Handbook of Transport Modelling*, Pergamon, Oxford.

Hensher, D and Dalvi, Q. (1978), *Determinants of Travel Choice*, Saxon House, UK.

Hibbs. J. (2003) *Transport Economics and Policy: A Practical Analysis of Performance, Efficiency and Marketing Objectives*, Kogan Page Limited, UK.

Lazear, E. and Michael, R. (1980), Family Size and the Distribution of Real Per Capita Income, *The American Economic Review*, **70**(1), 91-107.

MacKinnon, J., (2002), Bootstrap inference in econometrics, *Canadian Journal of Economics,* **35**, 615-645.

Mackett, R. (1993), Structure of Linkages Between Transport and Land Use, *Transportation Research Part B* **27**, 189-206.

Mckenzie, D. (2001), Estimation of (AR1) Models with Unequally Spaced Pseudo-Panels, *Econometrics Journal* **4**, 89-108

Meurs, H. and Haaijer, R. (2001), Spatial Structure and Mobility, *Transportation Research Part D* **6**, 429-446.

Meyer, M. (1997), *A Toolbox for Alleviating Traffic Congestion and Enhancing Mobility*, Institute of Transportation Engineers, Washington D.C.

Meyer, M. (1999), Demand management as an element of transportation policy: using carrots and sticks to influence travel behaviour, *Transportation Research Part A* **33**, 575-599.

Mokhtarian, P., Salomon, I., Redmond, L. (2001) Understanding the Demand for Travel: Its Not Purely 'Derived', *Innovation* **14**(4), 355-380.

Mooney, C. and Duval, R. (1993) *Bootstrapping: A Nonparametric Approach to Statistical Inference*, A SAGE University Paper, Series: Quantitative Applications in the Social Sciences, SAGE Publications, USA.

Nelson, J. (1988), Household Economies of Scale in Consumption: Theory and Evidence, *Econometrica*, **56**(6), pp1301-1314.

NOMIS – Official Labour Market Statistics, Accessed via 'www.nomisweb.co.uk' (2005).

Ortuzar, J., and Willumsen, L. (1990), *Modelling Transport*, John Wiley & Sons, UK.

Planning Policy Guidance 13: Transport (PPG13) (2001), Office of the Deputy Prime Minister.

Tan, Y. and Fan, H. (2003), Traffic Generated by Office and Retail developments in Singapore, *Proceeding of the Institute of Civil Engineers: Transport* **156**(4), 189-99.

TRICS 2004(b) Build 9.43 (2004), JMP Consultants Ltd.

Shapiro, S. and Wilk, M. (1965) An analysis of variance test for normality, *Biometrika*, **52**(3), 591-599.

Smith, L. and Sanchez, S. (2003), Assessment of Business Potential at Retail Sites: Empirical Findings from a US Supermarket Chain, *International Review of Retail, Distribution and Consumer Research* **13**(1), 37-58.

STATA /SE 8.2 for Windows, (2004), StataCorp LP, USA.

UK Census of Population (2001), Accessed via 'www.statistics.gov.uk', (2005).

Vickerman, R. and Barmby, T. (1984), The Structure of Shopping Travel: Some developments of the Trip Generation Model, *Jounal of Tranport Economics and Policy* **18**, 109-121.

Washington, S. (2000), Iteratively specified Tree-Based Regression: Theory and Trip Generation Example, *Journal of Transportation Engineering*, **126**(6), 482-491.

Wrigley, N. (2002), Food deserts in British cities: policy contexts and research priorities, *Urban Studies* **39**, 2029-2040.

Wrigley, N., Warm, D.L., Margetts, B.M. and Whelan A. (2002), Assessing the impact of improved retail access on diet in a 'food desert': A preliminary report, *Urban Studies* **39**, 2061-2082.

Yim, Y. (1992), The Effects of Transportation Services on the Scale of Food Retailing, *Working Paper, No 112*, University of California Transportation Centre (UCTC).

**Appendix: The Bootstrap regression model.**

**A1. The bootstrap model**

The results presented in Table 4 are derived from the following model;

$$\hat{Y} = (B^{-1}\sum_{b=1}^{B}\beta_0^*) + \sum[B^{-1}\sum_{b=1}^{B}(\beta_1^*,...,\beta_{20}^*)](X_1,...,X_{20}) + \varepsilon$$

Thus a standard log-linear (given that $Y$ and $X$ are the logarithmic transformations of the original variables) regression model is specified, where the coefficients are equal to the mean parameter estimate from $B$ bootstrap replications of the regression model (where $B =$ 100,000). Each replication of the model is conducted upon a unique 'bootstrap' data sample, which is a data set with the same dimensions as the original dataset (i.e. the same number of observations and variables), however each cell is uniquely drawn with replacement from the original dataset, with each individual observation in the original dataset having an equal probability of being drawn. On the assumption that the original dataset is truly representative of the entire population, then as the number of bootstrap draws increases, the more accurately we can observe the true distribution of the population data, and subsequently the distribution of the coefficients. For further explanation of the tenets of bootstrapping processes, see for example Efron and Tibshirani (1993), or for a brief overview of bootstrapping in econometrics see the survey article of Mackinnon (2002). A conventional ordinary least squares estimator is used to evaluate the coefficients for each bootstrap sample.

The resulting model offers a theoretical advantage over the standard OLS regression as the bootstrap procedure reveals the full empirical distributions for the estimated coefficients. Thus means that the *assumption* of normality can be dropped and accurate inference can be conducted regardless of the shape of the distribution.

Table A1 indicates the bias of the OLS parameter estimates relative to the bootstrap parameter estimates. The values in this table are calculated as the OLS coefficient minus the bootstrapped coefficient, thereby creating a value which, when positive, indicates that the OLS estimate is overestimating the true parameter[5].

**Table A1: Bias in OLS estimates.**

| | I | | | II | | |
|---|---|---|---|---|---|---|
| | (a) | (b) | (c) | (a) | (b) | (c) |
| CAR | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | 0.001 |
| ACCESIBILITY | 0.003 | 0.002 | 0.002 | 0.003 | 0.002 | 0.002 |
| GFA | 0.003 | 0.004 | 0.004 | | | |
| RFA | | | | -0.003 | -0.002 | -0.001 |
| RESISTANCE | 0.000 | 0.000 | 0.000 | -0.001 | -0.001 | 0.000 |
| AVHS | 0.034 | | | 0.024 | | |
| AVLH | | 0.010 | | | 0.005 | |
| AVHC | | | 0.012 | | | 0.009 |
| EMP | -0.003 | -0.001 | 0.000 | 0.001 | 0.003 | 0.003 |
| PARKING | -0.005 | -0.006 | -0.006 | -0.002 | -0.003 | -0.003 |
| PFS | 0.002 | 0.002 | 0.001 | 0.001 | 0.002 | 0.001 |
| LU1 | -0.047 | -0.045 | -0.057 | -0.055 | -0.054 | -0.065 |
| LU2 | 0.002 | 0.002 | 0.001 | 0.002 | 0.001 | 0.001 |
| LU3 | 0.000 | -0.001 | -0.002 | -0.003 | -0.004 | -0.004 |
| LU4 | -0.001 | -0.001 | -0.001 | -0.001 | -0.001 | -0.002 |
| LU5 | 0.008 | 0.007 | 0.007 | 0.008 | 0.007 | 0.006 |
| LU6 | 0.001 | 0.000 | 0.000 | 0.002 | 0.002 | 0.001 |
| LU7 | -0.001 | 0.000 | -0.001 | -0.001 | -0.001 | -0.002 |
| ECONOMY | -0.005 | -0.006 | -0.005 | -0.004 | -0.005 | -0.003 |
| SAT | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| SUN | -0.001 | -0.001 | -0.001 | -0.001 | -0.001 | -0.001 |
| MON-THURS | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | 0.001 |
| INTERCEPT | -0.034 | 0.007 | 0.007 | 0.002 | 0.033 | 0.029 |

The observed bias in the results from table A1 is low across all specifications, implying that only marginal gains were made by the application of the semiparametric model in relation to the inferences drawn.

## A2. One sample Achieved Significance levels

The following defines a significance test procedure for a single sample of data (i.e. one variable), which bears strong resemblance to standard hypotheses tests (t-tests) used for Gaussian normal variables. The test statistic is defined as;

[5] This process is done using real parameter values, therefore this argument holds true so long as the estimated parameters take a positive sign. If however they are negative, the effects are exactly reversed (i.e. for a negative coefficient a positive bias would indicate that the OLS parameter is underestimating the true parameter effect in real terms).

$$t(z^*) = \frac{\tilde{z}^* - \hat{z}}{\sigma^2 / \sqrt{n}}$$

where $n$ is the original sample size (i.e. 201), and $z^*$ are the bootstrap replications of the value for $z$. $\hat{z}$ is the value of $z$ that is being hypothesised/tested against and $\sigma^2$ is the standard deviation of the observed bootstrap coefficients $z^*$.

Applying the conventional null hypothesis for standard two-tailed significance tests the following null hypothesis is offered;

$$H_0 : \hat{z} = 0$$

Thereby reducing the hypothesis test to;

$$t(z^*) = \frac{\tilde{z}^* - \hat{z}}{\sigma^2 / \sqrt{n}} = \frac{\tilde{z}^* - 0}{\sigma^2 / \sqrt{n}}$$

This is the `critical' t-value for testing the assumption that the estimated coefficient is equal to zero. Once this has been computed, the empirical distribution of the coefficients must then be translated about zero (i.e. the mean is forced to be zero), which is the hypothesised value of the observed coefficient. Following this, each bootstrap replication is then tested against this critical value using the following decision criteria;

$$H0 : t_{calc} < t_{crit}$$

thus indicating that there is no evidence that parameter of interest (i.e. the individual bootstrap replication) is significantly different from zero.

The `translation' of the empirical distribution of $z$ (so as to create a new `null distribution' with mean equal to the null hypothesis) is done using the following formula;

$$\tilde{z}_i = z_i - \bar{z}_i + z_{H0}$$

thereby centering the distribution first about zero (by subtracting the observed mean $\bar{z}_i$), and then redistributing it about the hypothesised mean $z_{H0}$. These values $\tilde{z}_i$ are subsequently used in calculating the t-values for hypothesis tests on the now known empirical null distribution, noting that the null distribution does not need to be normal, using the formula;

$$t(z^*) = \frac{\tilde{z}^* - 0}{\sigma^2 / \sqrt{n}}$$

thus t-values are created for each bootstrap replication, where $\bar{z}^* = t_{calc}(z^*)$. These values are then compared to the previously calculated critical value in order to reveal the number of bootstrap replications which violate the null hypothesis, i.e. When $t(\tilde{z}^*) > t_{crit}$ it is not possible to reject the null hypothesis that the (empirical or untranslated) distribution of $z$ is centered around zero.

Defining;

$$\gamma = \# t(\tilde{z}^*) > t_{crit}$$

i.e. the number of times that the null hypothesis cannot be rejected. Then the achieved significance level (ASL) is found to be

$$ASL = \frac{\gamma}{B}$$

where $B$ is the number of bootstrap calculations. This is then interpreted as the probability that the (untranslated) empirical distribution of the bootstrap coefficients is centred around a zero mean, and can consequently be considered statistically insignificant.

The key difference between this test and standard significance test used in mainstream econometric applications is in the assumption that the estimated coefficient is derived from a standard normal distribution is now flexible. Hypothesis testing becomes feasible irrespective of the actual distribution that the data follows, where this evidently includes a normal distribution. Therefore the bootstrap framework coupled with the ASL approach to interpreting the significance of the coefficients provides a theoretical advantage over the standard OLS t-test procedure.

**Surrey Energy Economics Centre (SEEC)**

**Department of Economics**

**University of Surrey**

**Guildford**

**Surrey GU2 7XH**

SURREY

ENERGY

ECONOMICS

DISCUSSION PAPER

SERIES

**For further information about SEEC please go to:**

**www.seec.surrey.ac.uk**