Centre for Transport Studies
STOCKHOLM

Modelling the attraction of shopping centres

Ida Kristoffersson[1]
Andrew Daly[2,3,4]
Staffan Algers[5]

CTS Working Paper 2017:1

Abstract
Development of major shopping centres continues even though online shopping is increasing. This has implications for mode and destination choice for shopping travel and therefore also for sustainability. In this paper, we estimate models for shopping travel using an unusually rich data set of shopping attractions. We find that shopping travel is best represented in three separate models: consumables in short and long-activity segments and durables. In all of these models, we show that representing nearby attractions outside the destination zone adds to the measured attraction. For both consumables segments, the addition of secondary attractions within 2 km of the main destination adds very significantly to the explanation given by the model, but attractions within 5 km do not add anything. For durables, both 2 km and 5 km added to the model, but 2 km was slightly better. Furthermore, we find significant within-zone correlation in all three models, but are unable to find significant between-zone correlation, indicating that zone boundaries have some behavioural meaning for shopping travellers, but larger areas are not viewed in this way. Shopping attractions with a specifically Swedish impact, Systembolaget (official alcohol outlet in Sweden) and IKEA, proved to be important in all the models. These attractors work better as part of the size than as part of the utility, indicating that they appear to be separate attractors of trips, rather than as adding to the utility of other attractors.

Keywords: Shopping Travel Behaviour; Large-Scale Modelling; Travel Attractions;

JEL Codes: R12, R14, R15, R41, R42, R48

Centre for Transport Studies
SE-100 44 Stockholm
Sweden
www.cts.kth.se

[1] VTI Swedish Road and Transport Research Institute, Stockholm, Sweden
[2] ITS, University of Leeds, Leeds, UK
[3] RAND Europe, Cambridge, UK
[4] CTS, Stockholm, Sweden
[5] KTH Royal Institute of Technology, Stockholm, Sweden

1. INTRODUCTION

The rapid development of on-line shopping has taken attention from the continuing importance of 'physical' shopping and travel to shops and shopping centres. But development in shopping facilities continues apace: the recent opening on the London 2012 Olympic site of Stratford Shopping City in east London (175,000 sq. m., completing the largest shopping centre in the EU) is evidence of the importance attached by retailers to 'footfall': attracting shoppers in person to their sites. Shopping centres aim at extending the conventional shopping trip with possibilities to combine it with visits to restaurants and cinemas and meetings with friends. In Sweden also large shopping malls have been established, like Mall of Scandinavia in Stockholm which opened in 2015 (101,000 sq. m.). IKEA is another example of an important shopping attractor, which not only attracts shoppers itself, but also attracts other shops (and their customers) to nearby locations.

The large volume of trips attracted to shops and shopping facilities means that transport planners need to understand the impact that these trips may have on congestion, emissions and energy consumption. Particularly the balance between central and peripheral locations and their very different impacts on mode choice mean that planners need to understand how planning decisions taken by developers and governments will influence shopper behaviour and hence the transport consequences of these decisions.

The most recent Swedish study (Algers and Jonsson 2013) has developed models for shopping using a richer source of shopping supply than had previously been available. A nationwide retail database was made available by the retail research institute HUI Research. The purpose of the model development was to achieve a better understanding of how supply affects shopping behaviour. These models offered improved insight into shopper behaviour, but were limited to model formulations including only information on attractions in the specific destination zone, without providing a broader context of the neighbouring areas. Important interaction phenomena are therefore largely neglected.

The present study aims to improve the modelling of shopping behaviour. In particular, models are developed that improve the representation of shopping attraction. Here, we benefit from the unusually rich HUI data set which represents shopping attractions in a number of dimensions, such as numbers of shops, floor space and retail employment. Additionally, we have identified the location of specific shops that has been shown to influence traveller behaviour, such as IKEA and Systembolaget (official alcohol outlets). This has allowed us to develop separate and richly detailed models for both consumables and durables shopping. A particular interest of the present study is the extent to which shopping travel behaviour is influenced by the presence of multiple attractions at the same site or nearby.

The following section of the paper reviews the literature on shopping models. Section 3 then describes the choice model set up. Results are presented in Section 4 and Section 5 concludes.

2. LITERATURE REVIEW

The first models of shopping trips were of the gravity model type, which is the most widely used type of spatial interaction model (Haynes and Fotheringham 1984). Gravity models assume that the probability to choose a certain destination for shopping decreases with distance to the destination and increases with the size (attraction) of the shopping centre. The most famous model of this type is the so-called Huff-model (Huff 1963). Several models of the gravity type have been developed following upon the Huff-model (Gonzalez-Benito 2005; Simmonds and Feldman 2011). They differ mainly in the specification of the attraction and the distance decay function. Spatial interaction models are deterministic, which means that running the model several times gives the same result for a given set of input data.

Criticism of spatial interaction models – that they were not based on standard economic theory – encouraged the development of another type of models, discrete choice models and also their application to shopping trips (McFadden 1974; Ben-Akiva and Lerman 1985). Discrete choice models consist of a deterministic part (the observed part of the utility function) and a stochastic error term. Assuming that the error terms are IID extreme value distributed led to a discrete choice model with a closed form – the multinomial logit model (MNL). The MNL model has been applied in several studies to explain and predict shopping behaviour, see e.g. Oppewal et al. (1997). The results of these studies suggest that travel distance, shopping centre characteristics (also called shopping centre image), selection of stores and parking facilities are the main variables explaining choice of shopping destination (Arentze and Timmermans 2001). Space is in general divided into zones in transportation models and in the destination choice the user chooses between these different zones. Size variables - such as number of shops - are therefore aggregated per zone. Estimation of models containing more than one size variable is not trivial. Daly (1982) presents an estimation algorithm that can handle multiple size variables without first combining the size variables into a composite variable and shows that this can prevent biases in some common situations.

The MNL model is convenient, but has some severe drawbacks. First, it relies on the principle of independence of irrelevant alternatives (the IIA-property), meaning that the alternatives that the user chooses between must be independent. Second, in MNL models users are treated as homogenous with common parameter values estimated for the entire population and thus ignoring heterogeneity among consumers. Third, all discrete choice models (as well as gravity models) are designed to handle single-purpose and single-stop trips. Recent research on shopping models has tried to overcome some of these drawbacks.

It is likely that some shopping destination alternatives appear more similar to the user than others. This means that the IIA property does not hold. One way to solve this problem is to group similar alternatives in nests and estimate a nested logit model. Evidence exists that a specific store in a shopping centre is often the reason for choosing that centre (Ahn and Ghosh 1989), which implies that all

shopping centres where this store is present are more attractive to the consumer than the other centres. Suárez et al. (2004) estimate a nested logit model with choice of hypermarket on the upper level and choice of shopping centre on the lower level. Their results show that all shopping centres do not compete with each other to the same extent; the opening of a new shopping centre will have greatest effect on other shopping centres containing the same hypermarket. Another way to relax the IIA property is by using a mother-logit model in which the utility of one alternative depends on the attribute of other alternatives. The mother-logit model has been applied to shopping destination choice in Timmermans et al. (1992).

Suárez et al. (2004) also tackle the question of preference heterogeneity. They estimate a random effects model in which the parameters are randomly distributed in the population and find that a segment of their consumers is less sensitive to travel time than the others. Preference heterogeneity has been accounted for also by other means than using a random effects model, i.e. by demand segmentation (Gupta and Chintagunta 1994), introducing a variable for revealed shopping behaviour of the consumer (Guadagni and Little 1983) and fixed effects models that assign a fixed parameter to the utility function of each consumer (Chamberlain 1979).

The third drawback of the conventional MNL model – that it cannot handle multi-purpose and multi-stop trips – has mainly been dealt with by replacing the discrete choice model with activity-based models. Activity-based models aim at predicting the daily schedule of a user including the performed activities, their destinations, transportation mode, time of day and travelled route, taking into account also constraints at the household level regarding e.g. access to a car. An activity-based model that has been specifically applied in shopping behaviour research is Albatross (Arentze and Timmermans 2004). Rasouli and Timmermans (2013) provide an uncertainty analysis of shopping forecasts using the Albatross model. However, advanced discrete choice models may also be used to analyse multi-purpose shopping, which is done e.g. in Leszczyc et al. (2004). Multi-purpose shopping has been proposed as an explanation for agglomeration effects, i.e. the clustering of shops into shopping districts.

Recently, models have been developed for comparison of in-store and online shopping. Schmid et al. (2016) show that travellers for different shopping purposes differ a lot in their preference for in-store or online shopping, with grocery shopping being focussed on in-store shopping, whereas durables shopping shows a larger share of online shopping.

In Sweden, models for shopping travel including destination choice have been in focus since the middle of the seventies (based on data from the towns of Västerås in 1975 and Jönköping in 1979). The Algers et al. (1996) household shopping model, which allows for multi-stop and multi-purpose journeys, was developed as a component in the Stockholm Integrated Model System (SIMS). Around the year 2000 a national transport model for Sweden was developed, including shopping travel (Beser and Algers 2001).

3. DEVELOPMENT OF THE CHOICE MODELS

The context in which the modelling described in this paper has been conducted is that of large-scale travel demand models which support transport planning in Sweden - the Sampers model system (Beser and Algers 2001). In the currently applied version of Sampers, shopping travel is modelled together with travel for 'other purposes' (that is, other than commuting, education, recreation, social and business). However, the modelling in this paper has been conducted in a research version of Sampers, based on newer data and where consumables shopping and durables shopping are two out of twelve modelled trip purposes. Sampers aims to give a complete picture of personal travel and traffic by all modes. Together with the Samgods system (Swedish Transport Administration 2016), representing freight traffic, and network models representing congestion, the system allows planners to test options for policies to improve accessibility and mitigate the negative impacts of transport.

The Sampers models operate at both regional and national level, which implies that the scale of the modelling is quite large, covering all of Sweden, which presented issues for the conduct of the work described here. The regional models include trips up to 100 km in length, while longer trips are represented in the national and international models.

3.1 General modelling considerations

In the present work, we focus on shopping trips and we restrict attention to shopping trips up to 100 km in length, i.e. travel covered by the regional models.

All the Sampers models of personal travel are based on the concept of individual utility maximisation, which has proved a reliable and flexible basis for modelling in many contexts. The shopping models developed here are also based on utility maximisation, not only for consistency but also because utility maximisation is an excellent basis for modelling.

Following Daly (1982), our models feature the separation of attractions and accessibility as separate components. For this work, which primarily focusses on attractions, we model accessibility using methods consistent with other components of Sampers. Thus, mode choice is included in the model to improve the representation of accessibility and to allow the prediction of shifts in modal choice that might follow from changes in the spatial distribution of shops. Attractions are modelled using the approach suggested by Daly (1982) and followed in numerous large-scale models since then, but here we extend the detail and sophistication of the modelling.

A central concern of this work is an important issue in modelling shopping attractions: the degree to which a shop benefits by the presence of neighbouring shops, so that shoppers can visit several shops in one trip. In modelling the choice of destination, the correlations among the utilities of each of the elementary alternatives within a zone (e.g. shops) can be estimated at the expense of admitting that the zone boundaries are not simply arbitrary constructs by the analyst but have meaning for the shopper. In this context, we must note that the

zone boundaries have been constructed with reference to groupings of developments that are generally known to the population, so that it is not unreasonable that behaviour might be related to specific zones.

A second aspect of the impact of utility correlation is the interaction between neighbouring zones. In our model, this correlation is represented by also including attraction variables defined over a coarser zone system, so that the possibility for shoppers to visit neighbouring shopping centres on one tour from home can increase the utility of each centre. Utility correlation can be represented through measured variables or through the correlation of unmeasured utility components; we investigate both of these possibilities.

These considerations form the basis for the detailed modelling work described in the following sections.

3.2 Availability of data

To model shopping travel behaviour, we need data on observed travel behaviour, data on shopping activity supply and level of service for the different modes connecting the shopper with the different shopping supply locations.

Data on observed shopping behaviour was provided by the 2005/06 Swedish national travel survey. From the one day travel diary, choice of shopping type, choice of destination (geocoded to the 10 000 zones used in the Sampers system) and choice of mode (car driver, car passenger, public transport, cycle and walk; other modes were not modelled) were available. For shopping type, the respondent was asked to classify the shopping activity into “dagligvaror” or other shopping (“övriga inköp”). Here, we will use the terms “consumables” for “dagligvaror”, which would cover groceries and other frequently purchased items, and “durables” for other shopping. It would be expected that consumables would be bought on shorter tours than durables, with more possible locations for consumables than for durables, and that the mode split might well be different because of the tour length but also because of the different types of goods to be transported and different accessibilities of the shopping locations. After simplification of trip chains into home based tours, about 2900 observations of consumables shopping tours and about 1800 observations of durables shopping tours were obtained.

In the Sampers system, data on activity supply of different kinds has generally been in the form of numbers of employees in different economic branches, in some cases supplemented by other activity data. For shopping, the number of supermarkets was also used in the initial Sampers model estimation in 1999. Such data has not been part of the Sampers maintenance procedures, and has not been properly updated over time, in contrast to data on the number of employees that has been updated together with other land use data.

Policy issues related to shopping, specifically the impact of external shopping malls, have become more and more important over time. Further development of shopping travel demand models was therefore initiated (Algers and Jonsson 2013), and a much richer data base on shopping supply was established in

cooperation with HUI Research (the Swedish Institute for Retail Research). In addition to the conventional data on numbers of employees (corresponding to the 2-digit level EU NACE classification) supplied by Statistics Sweden, HUI Research supplied two databases including retail supply data. One of them is the Delfi data base containing all shops with consumables supply in Sweden 2005. The Delfi data base contains several variables related to the specific consumables shops, including the type of shop (warehouse, supermarket, etc.) and the sales area (floor space).

The other data base is a business place register containing information on each business place used by every firm or organisation in Sweden building on information from Statistics Sweden (a firm can have several business places). From this source, including durables supply as well as consumables supply, 11 types of consumer sales activities active in 2005 were chosen, as shown in Table 1.

Table 1: Consumer sales activities in the business place database

Branch code
SNI50 – cars
SNI505 – fuel stations
SNI521 – warehouses, supermarkets, large diversity
Consumables stores
SNI522 – smaller consumables stores, tobacco shops
SNI52250 – Systembolaget (alcohol shops)
SNI523 – Drug stores, medical items, cosmetics
Durables stores
SNI524a – Clothes, shoes
SNI524b – Furniture, radio and television
SNI524c – very mixed…
SNI525 – Antiques, second hand
SNI527 – Repair

For business places, the number of employees in each business place was also supplied (coded in intervals).

Data on transport level of service was provided by the Sampers system using the EMME network analysis tool, and travel cost matrices were supplied by the Swedish Transport Administration.

3.3 Specification of the model

For this study, we have adopted the basic outline of the model used in recent updates of the Sampers models (Algers and Jonsson 2013), keeping many details of that specification, particularly those relating to the modelling of mode choice. In that modelling, the interaction of mode and destination choice was found to be best represented in a tree-nested logit model, giving the joint probability of choosing mode k and destination j in a structure nesting mode 'above' destination by

$$p_{jk} = Pr\{choose\ k\}.Pr\{choose\ j\ given\ k\}$$

$$= \frac{\exp V_k}{\sum_{k'} \exp V_{k'}} \cdot \frac{\exp V_{kj}}{\sum_{j'} \exp V_{kj'}} \quad (1)$$

where $V_k = \theta \sum_j \exp V_{kj}$ is the 'composite' utility for mode k;
V_{kj} is the utility for mode k and destination j, which is discussed in more detail below;
$0 < \theta \leq 1$ is a structural coefficient indicating the ratio of relative sensitivity of mode and destination choice.

Note that this specification follows the 'bottom-normalised' specification of the tree logit model as indicated in Daly (1987), which is consistent with the ALOGIT software used for this work. This model structure with modes 'above' destinations is more common in European work and is justified in the present context by the finding that θ is significantly less than 1, confirming the previous results in the Sampers work (Beser and Algers 2001). The discussion of destination choice issues in the present paper would however not be greatly changed in principle for the alternative model structure with destinations 'above' modes, but this model structure is not consistent with the data. The structure of the model does not imply any conditionality or order in the overall choice but relates purely to the distribution of errors in the utility.

Following Daly (1982) we define the mode-destination utility V_{kj} as having a component dependent on the separation between origin and destination and a component dependent only on the destination; the latter would apply to travellers from any origin. This separation can be described as somewhat arbitrary, in that it assumes that access to a zone and the attraction of the zone are independent, but it seems a reasonable approximation. Thus, we formulate

$$V_{kj} = \sum_{r \in L} \beta_r x_{rjk} + \sum_{r \in D} \beta_r y_{rj} + \phi \log S_j \quad (2)$$

where x_{rjk} gives component r of the level of service offered by mode k to zone j;
y_{rj} gives component r of the *quality* of destination j;
β are coefficients to be estimated, divided into a set L relating to level of service and a set D relating to destination quality;
S_j is a measure of the *quantity* or size of the attraction of zone j;
ϕ, the log size multiplier, is described in detail below.

The definition of the level of service measures in this work is entirely conventional and we follow the previous Sampers specification. However, our main focus is on the attraction variables. Here, the first issue is whether a zonal attribute should be considered to relate to the quality or quantity of the attraction offered by the zone.

- An attribute that can add to the utility of all trips to the zone needs to be included in the quality component. Here we think primarily of measures of the shopping environment.

- An attribute that simply measures the number of attractions of the same quality needs to be included in the size component S. Here we think of measures of the number or size of shops.

For example, consider the presence in the zone of one of the state alcohol monopoly outlets, Systembolaget. If this is seen as an additional facility that could add to the usefulness of shopping at other shops, then it might appear as a utility component; but if it is seen as an attractor in its own right, not necessarily in combination with other shopping, then it should appear as a size component. Of course, tests of the appropriateness of each formulation can be made by testing different specifications in the model.

In formulating the size component S we follow the approach of Cochrane (1975) and Daly (1982) in viewing the destinations included in a zone as the aggregation of elementary alternatives. This aggregation, necessary to incorporate destination choice into large-scale transport models, represents the number of elementary alternatives as proportional to the number of shops, retail employees or retail area in the zone. Through the logarithmic formulation ($\log S$) in equation (2) it can be ensured that the probability of choosing the zone is proportional to its size S and therefore that the number of trips attracted to an area does not depend on the zone boundaries. This proportionality however requires that $\phi = 1$.

The log size multiplier ϕ represents the inverse correlation of the utility of the elementary alternatives in a zone. If $\phi = 1$, the utilities are uncorrelated. However, for $\phi < 1$ the model implies that the utilities within each zone are correlated, so that the zone boundaries appear to have a behavioural meaning. The parameter ϕ is exactly equivalent to a structural parameter in a tree-nested logit model (McFadden 1978; Daly 1987) and plays exactly the same role in this model, specifically, giving the ratio of the standard deviations of the zonal utility differences to the standard deviations of the elementary alternative utility differences.

Within S, there may be several measures of elementary attractions, including different measures of size, such as retail area or retail employment, and measures of different types of retail activity. The importance of each of these measures will be different, of course, so that coefficients need to be estimated to correspond to these differences and also to adjust for the units of measurement. Thus, the overall attraction measure is formulated

$$S_j = \sum_r \tau_r s_{rj} \qquad (3)$$

where s_{rj} is the r^{th} size measure for zone j;
τ are coefficients to be estimated.

It follows from the nature of the size variable that it must be positive and all of the components must be non-negative. To maintain this constraint and to avoid problems in estimation the coefficients τ are parametrised as exponentials. Moreover, because of the way in which the choice probabilities are calculated (see equations 1 and 2) it is not possible to identify coefficients for all of the size

components and to deal with this issue the constraint $\tau_1 = 1$ is applied. Therefore, the size measure is reformulated as

$$S_j = s_{1j} + \sum_{r>1} \exp(\gamma_r)\, s_{rj} \qquad (4)$$

The γ coefficients are estimated along with β, θ and ϕ. It should be noted that $\gamma = 0$ does not mean that the corresponding size component does not influence choice; it means rather that $\tau = 1$ and the influence of that component is the same as that of s_1.

In the modelling work described below, s_1 was set to the population of each possible destination zone. In this context, using population in this way covers the possibility of purchases from private houses but also allows for gaps in the data base where shops are not recorded.

3.4 Estimation of the model

The model framework is described by equations (1), (2) and (4). Maximum likelihood estimation is used, as is standard for models of this type. This form of estimation allows extensive testing of the results of estimation and also allows use of standard efficient software.

Maximum likelihood estimation also allows calculation of error measures. In the case of size variables, an estimation of error needs to take account of the logarithmic formulation. The standard error of the coefficient τ can be calculated, following Daly et al. (2012) as

$$se(\tau) = se(\gamma).\frac{\partial \tau}{\partial \gamma} = se(\gamma).\exp\gamma \qquad (5)$$

where $se(x)$ gives the standard error of x, i.e. $se(x) = \sqrt{var(x)}$.

This gives a more realistic assessment of whether the coefficients of the size component are accurately estimated.

However, because of the formulation of the model and the fact that such errors are asymptotic, a test based on the standard error of the exponentiated coefficient can only be approximate. The best test that can be made on whether a size component is contributing significantly to the explanation of behaviour is to estimate models with and without the component and to make a χ^2 test on the likelihood difference.

A key issue for this study is to investigate the measures of attraction that are available. First, we have a richer description of the potential attractions than has been available to previous studies; these include specifically Swedish features such as the alcohol monopoly Systembolaget and IKEA. Second, we measure the intra-zone interaction of attractions through the log-size multiplier ϕ. Third, we investigate the potential interaction of attraction between zones, in two ways:

a) One possibility is that the geographic scale given by the zones does not adequately cover unmeasured interactions between attractions. We investigate this by modelling the possibility that the utilities of neighbouring zones are correlated. We model this correlation within larger zones, in our case the geographical zones of the national model. The regional models contain in total about 10000 zones, differing in size depending on population density, with smaller zones in more dense areas. However, the regional models model only trips to destinations less than 100km from the origin. A maximum of 1705 regional model zones lie within a 100km radius of any origin. The national model on the other hand has about 600 zones which are of about equal geographic size. Therefore, the number of small zones per large zone differs from one to about 100. The model is then extended from equation (1) to

$$p_{jk} = Pr\{choose\ k\}.Pr\{choose\ j^*\ given\ k\}.Pr\{choose\ j\ given\ kj^*\}$$

$$= \frac{\exp V_k}{\sum_{k\prime} \exp V_{k\prime}} \cdot \frac{\exp V_{kj^*}}{\sum_{j\prime^*} \exp V_{kj\prime^*}} \cdot \frac{\exp V_{kj}}{\sum_{j\prime} \exp V_{kj\prime}} \qquad (6)$$

where j^* indexes the larger zones to which j may belong and

$$V_{kj^*} = \psi \sum_{j \in j^*} \exp V_{kj}$$

Here the significance of correlation across zones within larger zones can be tested by testing whether the additional structural parameter $\psi < 1$. If $\psi = 1$ there is no significant correlation and the additional level of the modelling can be omitted.

Because the number of small zones within each of the large zones varies substantially across Sweden, the file sizes needed for model estimation, and hence the run times, can become very large. By careful processing, substantial reductions could be made, but the runs remained lengthy.

b) We also study the measured addition to the attraction of a zone offered by neighbouring zones. This is done simply by calculating an additional attraction added to the size:

$$S_j = s_{1j} + \sum_{r>1} \exp(\gamma_r)\, s_{rj} + \exp(\gamma_a)\, s_{aj} \qquad (7)$$

where the additional (secondary) attraction s_{aj} is given by

$$s_{aj} = \sum_{l \in A(j)} S'_j = \sum_{l \in A(j)} \{s_{1l} + \sum_{r>1} \exp(\gamma'_r)\, s_{rl}\} \qquad (8)$$

and $A(j)$ is the set of zones within a given radius of zone j, excluding j;
γ' are approximate estimates of γ, e.g. derived from a previous model estimation.

Tests can be made for different radii: below we report results for 2 km and 5 km. If the model including s_{aj} is significantly better (e.g. on a χ^2 test)

than the model without it, we conclude that the attractions in neighbouring zones do indeed add to the attraction of a zone.

4. RESULTS

4.1 Preliminary findings

The rich shopping data available for this work includes information for each zone on specific attractions such as Systembolaget and IKEA. We tested model specifications where these were included either in the utility or in the size functions and found that the model fit was better when they were included in the size function. We concluded that they are seen as separate attractors and further model development includes them in this way.

Using the nested structure with small (regional model) zones as part of larger (national model) zones described above, we tested whether the utilities of neighbouring zones within the larger zones are correlated. We found however that the additional structural parameter ψ was not significantly different from 1, and concluded that there is no significant correlation between unmeasured components of utility of neighbouring zones in this model structure. We therefore omit the intermediate choice level of larger zones and continue investigating between-zone interaction through the application of secondary attractions.

4.2 Consumables

The data base for consumables covers quite a wide range of shopping activities, from buying a newspaper to extensive family shopping in a supermarket. To investigate whether these types of activities could usefully be separated, we tested splitting the consumables data into two segments. After some experimentation, we determined that a useful split could be obtained by separating the tours into those with activity time more or less than 30 minutes, presumably relating to the volume of shopping. These two segments are identified as 'long' and 'short' activity times in the report of the modelling work.

In addition to the split on activity time, for consumables shopping, we estimate models with secondary attractions within a radius of 2km and 5km respectively. These results are set out in Table 2 below, along with the results for the complete consumables data set and for models without secondary attractions, giving nine models in total.

Table 2: Comparison of log-likelihood for nine consumables models

	All activity times (2897 tours)	Short activity (1766 tours)	Long activity (1131 tours)	Total after split	Gain from activity time split	Significance level
No SA[6]	-9304.9	-4960.0	-4212.8	-9172.8	132.1	5.0E-32
SA <2	-9294.4	-4952.5	-4205.6	-9158.1	136.3	3.7E-33

[6] SA stands for secondary attractions.

SA <5	-9304.9	-4960.0	-4212.8	-9172.8	132.1	1.2E-31
SA <2 gain	10.5	7.5	7.2	14.7	-	-
Significance level	4.6E-06	1.1E-04	1.5E-04	4.1E-07	-	-
SA <5 gain	0.0	0.0	0.0	0.0	-	-

It can be seen from Table 2 that adding secondary attractions within 2 km of the destination zone improves the model significantly (based on a χ^2 test) for all activity times and for long and short activity times separately and together. However, adding secondary attractions within a 5 km radius does not improve the model noticeably. We conclude that for consumables it is necessary to include the secondary attractions within 2 km.

It can also be seen in Table 2 that the split into short and long activity time segments very significantly improves the model, both with and without secondary attractions included. We conclude that this split is essential and that the best model for consumables is therefore the split model with secondary attractions within 2 km.

Table 3 shows these best consumables models. Results show a large difference between short and long-activity consumables shopping for θ, which is much smaller (0.23 compared to 0.72) in the case of short-activity consumables shopping, indicating a lower relative sensitivity of destination choice compared to mode choice in this case, possibly indicating that the attractiveness of specific shops is relatively less important for the short-activity tours. The log-size-multiplier ϕ differs also between short- and long-activity consumables shopping with an estimated value of 0.62 for short-activity and 0.90 for long-activity, indicating a reduced variance of utilities within each zone for short-activity tours. The findings for θ and ϕ seem to be consistent with the idea that preference between shops is less clear for short-activity than for long-activity consumables shopping.

Many of the parameters in Table 3 are level of service parameters and part of the mode choice utility specification. In the explanation, which mode utility function the parameter belongs to is marked by *C* (Car), *CP* (Car as passenger), *PT* (Public Transport), *W* (Walk) and *BC* (Bicycle). Parameters that are part of the size function are marked with *Size*. In the data set few consumables tours have public transport as the chosen mode; this is especially true for short-activity consumables tours. For this reason, some of the PT parameters are not well estimated. Among level of service variables (Time, transfer, cost and distance), only two (log of costs and bicycle distance) were significantly different between short and long activities, and only the first one was significantly different from the final model in Algers and Jonsson (2013). The estimated level of service parameters and the values of time they imply are reasonable. However, for a practical model, it would be necessary to adjust the specification to obtain a better representation of the PT accessibility, but here the interest is more to maintain comparability across the models shown in Tables 2 and 3.

It should be noted that the size parameter presented in Table 3 is τ together with its corresponding t-statistic (not γ, see equations 3-5 in Section 3.3). Similarly, the θ and ϕ parameters are tested for difference from 1, not 0.

We were, after introducing secondary attractions into the size component, unable to make satisfactory estimates for the parameters Dag_55 (multiplying the number of employees in hotel and restaurant activities) and TotB_ant (multiplying the total number of shops) because of high correlations and the form of the size variables. χ^2 tests showed that it was best to drop the parameter Dag_55 and keep the parameter TotB_ant in the model, but constrain it to a value found in the estimation of a simpler model, as in Table 3. These changes to the variables included in the size component, as well as the subdivision into short and long activities makes it difficult to compare the size parameters to earlier models.

Table 3: Best models for short and long-activity consumables shopping including secondary attractions within a radius of 2 km. Figures in brackets indicate the most relevant t statistics.

	Explanation	Short-activity	Long-activity
Observations		1766	1131
Final log (L)		-4952.5	-4205.6
Parameters		48	48
$Rho^2(0)$		0.61	0.50
$Rho^2(c)$		-2.15	-3.25
Cent_l	PT/W/BC – County centre zone	1.052 (2.8)	0.723 (3.1)
Cent_Si	C/CP – Stockholm inner city zone	-1.008 (-2.2)	-0.381 (-1.5)
PT_dwltime	PT – Activity time	-0.199 (-1.3)	0.008 (2.1)
BC_dwltime	BC – Activity time	-0.157 (-2.3)	-0.003 (-0.8)
W_dwltime	W – Activity time	-0.274 (-3.3)	-0.068 (-4.0)
CPPt_costl	C/CP/PT – Log(tour cost)	-0.497 (-4.2)	-0.130 (-2.1)
CPPT_cost	C/CP/PT – Tour cost	-0.012 (-1.8)	-0.005 (-1.3)
Car_t_li	C – Tour travel time	-0.113 (-8.7)	-0.098 (-15.6)
CarP_t_li	CP – Tour travel time	-0.075 (-6.4)	-0.083 (-13.3)
BC_75plus	BC – Over age 75	-2.055 (-1.3)	-1.330 (-1.8)
BC_15minus	BC – Under age 15	4.930 (2.5)	-0.860 (-0.5)
BC_alone	BC – Travelling alone	4.528 (2.6)	2.454 (2.9)
BC_Skane	BC – Region is Skåne	3.001 (2.2)	1.977 (2.9)
BC_Sydost	BC – Region is Sydost	4.858 (3.1)	1.130 (1.8)
BC_OeqD	BC – Origin and Destination is within the same zone	0.367 (1.3)	0.707 (1.6)
BC_distLi	BC – Tour distance	-0.351 (-9.6)	-0.222 (-5.1)
W_Samm	W – Region is Samm	1.445 (1.8)	0.429 (1.4)
W_OeqD	W – Origin and Destination is within the same zone	1.281 (9.8)	0.461 (1.7)
W_distLi	W – Tour distance	-0.566 (-19.1)	-0.545 (-9.7)
PT_Transf	PT – Number of transfers	-0.868 (-2.1)	-0.340 (-2.2)

PT_FW_lo	PT – Log(first waiting time)	-1.411 (-1.6)	-0.592 (-2.0)
PT_AcEg	PT – Access and egress time	-0.079 (-1.8)	-0.017 (-1.6)
PT_Inv_li	PT – In-vehicle travel time	-0.126 (-1.9)	-0.019 (-1.6)
PT_Inv_lo	PT – Log(in-vehicle travel time)	1.171 (1.0)	-0.725 (-2.5)
CP_y15	CP – Under age 15	3.856 (2.3)	3.966 (3.1)
Wom_CP	CP – Woman	3.916 (2.6)	2.942 (4.6)
Car_HH_CP	CP – Access to car in household	9.210 (2.9)	3.022 (4.4)
CarP_OeqD	CP – Origin and Destination is within the same zone	-2.883 (-3.8)	-0.828 (-1.8)
Car_OeqD	C – Origin and Destination is within the same zone	-3.473 (-4.5)	-1.343 (-3.0)
CComp_C_2	C – Car competition (#cars/#licences) when not travelling alone	-2.932 (-3.0)	-1.140 (-3.8)
Car_Palt	C – Region is Palt	1.746 (1.9)	0.690 (1.9)
Car_HHStlk	C – Household size	1.568 (3.3)	0.656 (4.6)
Wom_C	C – Woman	-2.376 (-2.7)	-0.920 (-2.9)
CComp_C_1	C – Car competition (#cars/#licences) when travelling alone	-3.771 (-3.1)	-1.425 (-4.3)
Car_HH_C	C – Access to car in household	10.590 (3.3)	4.069 (5.3)
PT_Sunday	PT – Day is Sunday	-4.428 (-0.9)	-0.543 (-0.7)
Bike_Const	BC – Alternative specific constant	-4.271 (-1.9)	-3.657 (-3.0)
W_Const	W – Alternative specific constant	9.934 (3.0)	4.603 (4.4)
PT_Const	PT – Alternative specific constant	-3.915 (-1.0)	0.226 (0.2)
Car_P_Cons	CP – Alternative specific constant	-7.408 (-2.5)	-3.054 (-3.5)
ϕ	Log size multiplier	0.624 (10.8)	0.899 (2.5)
Population	Size – Zone population	1 (*)	1(*)
Systemet	Size – Zone contains at least one state alcohol monopoly outlet	178082 (1.3)	17258 (2.5)
TotB_ant	Size – Total number of shops in each zone	422 (*)	422 (*)
VHall_yta	Size – Floor space for consumables shops 400-799 m2	967 (1.2)	10.5 (1.2)
ODVB_yta	Size – Floor space for consumables shops 1-399 m2	158 (1.5)	1.1 (1.0)
DVH_yta	Size – Floor space for consumables shops larger than 800 m2	99.6 (1.4)	4.9 (2.5)
Dag52	Size – Number of employees in consumables branches (Table 1)	1755 (1.6)	199 (3.4)
SecAttr	Size – Secondary attractions within a radius of 2km	0.5 (1.5)	0.05 (2.1)
θ	Tree parameter	0.230 (12.4)	0.716 (2.7)

4.3 Durables shopping

Also for durables shopping, we estimate models with secondary attractions within a radius of 2 km and 5 km respectively. Adding secondary attractions

improves the durables model both for secondary attractions within a radius of 2 km and 5 km (Table 4), where it can be seen that the 2 km model has a slightly higher log-likelihood value. It is notable that attractions up to 5 km away are significant in influencing these longer-distance tours, in contrast to the situation found for consumables, even though the 2-km limit gave a slightly better explanation.

Table 4: Comparison of models with and without secondary attractions for durables shopping

	No SA	SA <2	SA <5
Final log-likelihood	-8167.3	-8130.8	-8131.2
Gain from SA	-	36.5	36.1
Significance level	-	1.3 E-17	1.9 E-17

The best model is shown in Table 5. The log-size multiplier is around 0.9, indicating that zone boundaries have some meaning. Among level of service variables (time, transfer, costs and distance), there are no significant differences compared to the final specification in (Algers and Jonsson 2013). The size parameters differ considerably, but a comparison is made more difficult by the fact that the earlier model contained an additional size variable (number of employees in hotel- and restaurant activities), for which we were unable to make satisfactory estimates after introducing secondary attractions into the model and for which a χ^2 test showed that it was best to drop this variable.

Table 5: Best model for durables shopping including secondary attractions within a radius of 2km. Figures in brackets indicate the most relevant t statistics.

	Explanation	**Durables model**
Observations		1812
Final log (L)		-8130.8
Paremeters		35
$Rho^2(0)$		0.399
$Rho^2(c)$		-3.6492
Cent_k	C/CP/PT/W/BC – Municipality centre zone	-0.384 (-4.8)
Cent_Si	C/CP – Stockholm inner city zone	-0.321 (-2.1)
WBC_dwl_30	W/BC – Activity time less than 30 minutes	3.036 (7.0)
WBC_dwl_60	W/BC – Activity time 30-60 minutes	1.542 (4.6)
CPPt_costl	C/CP/PT – Log(tour cost)	-0.216 (-4.8)
CPPT_cost	C/CP/PT – Tour cost	-0.009 (-4.0)
Car_t_li	C – Tour travel time	-0.049 (-12.5)
CarP_t_li	CP – Tour travel time	-0.040 (-12.0)
BC_OeqD	BC – Origin and Destination is within the same zone	-0.577 (-1.7)
BC_distLi	BC – Tour distance	-0.357 (-9.1)
W_Samm	W – Region is Samm	0.674 (2.6)
W_Wkend	W – Day is Saturday or Sunday	0.956 (3.7)

W_distLi	W – Tour distance	-0.494 (-17.0)
PT_Transf	PT – Number of transfers	-0.393 (-4.9)
PT_FW_lo	PT – Log(first waiting time)	-0.459 (-2.8)
PT_AcEg	PT – Access and egress time	-0.049 (-5.3)
PT_Inv_li	PT – In-vehicle travel time	-0.035 (-8.2)
Wom_CP	CP – Woman	1.406 (4.2)
Wom_C	C – Woman	-1.578 (-5.4)
Car_HH_CP	CP – Access to car in household	2.310 (4.5)
Car_HH_C	C – Access to car in household	4.635 (6.7)
CarP_OeqD	CP – Origin and Destination is within the same zone	-1.754 (-4.0)
Car_OeqD	C – Origin and Destination is within the same zone	-1.523 (-4.3)
CComp_C	C – Car competition (#cars/#licences)	-0.866 (-3.9)
PT_Sunday	PT – Day is Sunday	-1.357 (-2.5)
Bike_Const	BC – Alternative specific constant	-2.245 (-2.9)
W_Const	W – Alternative specific constant	-0.587 (-0.9)
PT_Const	PT – Alternative specific constant	1.557 (2.3)
Car_P_Cons	CP – Alternative specific constant	-2.291 (-3.3)
ϕ	Log size multiplier	0.885 (4.1)
Population	Size – Zone population	1 (*)
Ikea	Size – Zone contains Ikea	1201 (3.8)
Tot_arst	Size – Sum of number of employees in branches of Table 1.	37.7 (4.4)
Dag52	Size – Number of employees in consumables branches (Table 1)	111 (6.0)
SecAttr	Size – Secondary attractions within a radius of 2km	0.0008 (1.0)
θ	Tree parameter	0.58 (6.8)

5. CONCLUSIONS

The intention of the work described in this paper was to investigate choices made for shopping travel in Sweden in more detail than has been attempted previously. In particular, we studied the representation of attractions in terms of their complementarity and the detail in which they are modelled. We were helped by the availability of a detailed data set describing shopping attractions.

We find that shopping travel is best represented in three separate models: consumables in short and long-activity segments and durables. The evidence for this division is very clear.

Unmeasured aspects of attraction are represented in the structural parameters of the model. In all three of the models recommended, we find significant within-zone correlation of the elementary utilities, represented by the finding that $\phi < 1$. In contrast, we were unable to find significant between-zone correlation. We

conclude that the zone boundaries have some behavioural connotation for shopping travellers, but larger areas are not viewed in that way.

The structural parameter reflecting the utility variance between modes and destinations, θ, shows a high value for long-activity consumables, a low value for short-activity consumables and an intermediate value for durables. The suggestion is that the variance at destination level is highest for long-activity consumables and lowest for short-activity consumables, which is plausible.

We were able to show that including nearby attractions outside the destination zone adds to the measured attraction in all of the models. For both consumables segments, the addition of secondary attractions within 2 km of the main destination adds very significantly to the explanation given by the model, but attractions within 5 km do not add anything. For durables, both 2 km and 5 km added to the model, but 2 km was slightly better. The spatial scales suggested by these results probably relate to the tour lengths involved in the separate models and to the scale of the level-of-service coefficients estimated.

The shopping attractions with a specifically Swedish impact, Systembolaget and IKEA, proved to be important in all the models. Variables indicating these attractors work better as part of the size than as part of the utility, indicating that they appear to be separate attractors of trips, rather than as adding to the utility of other attractors. Otherwise, these models largely confirm the attraction variables found by Algers and Jonsson (2013), with differences corresponding to the changes of model specification that we have described.

The models developed in this paper can be used either as sophisticated models in a larger modelling package (such as Sampers) together with other trip purposes such as work and recreation models, or for evaluation of specific shopping policies, such as extension of external shopping malls versus city-centre development.

Future work that should be carried out in this area would be to apply these models to test specific scenarios for shopping-related transport policy. In particular, the impacts of new shopping developments, whether in city centres or in out-of-town malls, or changes to mode costs should be investigated.

6. REFERENCES

Ahn, K. H., and A. Ghosh. 1989. “Hierarchical Models of Store Choice.” *International Journal of Retail & Distribution Management* 17 (5).

Algers, S., A.J. Daly, P. Kjellman, and S. Widlert. 1996. *Stockholm Model System (SIMS): Application. In: Seventh World Conference on Transport Research, Edited by D Hensher, J. King, T. Oum, Pergamon.*

Algers, S., and D. Jonsson. 2013. “Utveckling Av En Modell För Inköpsresor Med Förbättrad Målpunktsbeskrivning. Energimyndighetens Forskningsprogram Avseende Energieffektivisering I Transportsektorn.” Technical Report 34050–1. KTH Royal Institute of Technology.

Arentze, T. A., and H. J. P. Timmermans. 2001. "Deriving Performance Indicators from Models of Multipurpose Shopping Behavior." *Journal of Retailing and Consumer Services* 8 (6): 325–334.

———. 2004. "A Learning-Based Transportation Oriented Simulation System." *Transportation Research Part B: Methodological* 38 (7): 613–633.

Ben-Akiva, M., and S. Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. Vol. 9. The MIT Press.

Beser, M., and S. Algers. 2001. "SAMPERS — The New Swedish National Travel Demand Forecasting Tool." In *National Transport Models: Recent Developments and Prospects*, 101–18.

Chamberlain, G. 1979. *Analysis of Covariance with Qualitative Data*. National Bureau of Economic Research Cambridge, Mass., USA. http://www.nber.org/papers/w0325.

Cochrane, R. A. 1975. "A Possible Economic Basis for the Gravity Model." *Journal of Transport Economics and Policy*, 34–49.

Daly, A. 1987. "Estimating 'tree' Logit Models." *Transportation Research Part B: Methodological* 21 (4): 251–267.

Daly, A., S. Hess, and G. de Jong. 2012. "Calculating Errors for Measures Derived from Choice Modelling Estimates." *Transportation Research Part B: Methodological* 46 (2): 333–341.

Daly, A.J. 1982. "Estimating Choice Models Containing Attraction Variables." *Transportation Research Part B: Methodological* 16 (1): 5–15.

Gonzalez-Benito, O. 2005. "Spatial Competitive Interaction of Retail Store Formats: Modeling Proposal and Empirical Results." *Journal of Business Research* 58 (4): 457–466.

Guadagni, P. M., and J. D. C. Little. 1983. "A Logit Model of Brand Choice Calibrated on Scanner Data." *Marketing Science* 2 (3): 203–238.

Gupta, S., and P. K. Chintagunta. 1994. "On Using Demographic Variables to Determine Segment Membership in Logit Mixture Models." *Journal of Marketing Research*, 128–136.

Haynes, K. E., and A. S. Fotheringham. 1984. *Gravity and Spatial Interaction Models*. Vol. 2. Sage publications Beverly Hills. https://www.researchgate.net/profile/Alexander_Fotheringham/publication/44411121_Gravity_and_spatial_interaction_models__Kingsley_E._Haynes_A._Stewart_Fotheringham/links/02e7e52a62f20407e3000000.pdf.

Huff, D. L. 1963. "A Probabilistic Analysis of Shopping Center Trade Areas." *Land Economics* 39 (1): 81–90.

Leszczyc, P. T.L., A. Sinha, and A. Sahgal. 2004. "The Effect of Multi-Purpose Shopping on Pricing and Location Strategy for Grocery Stores." *Journal of Retailing* 80 (2): 85–99.

McFadden, D. 1974. *Analysis of Qualitative Choice Behavior. Zarembka, P.(ed.): Frontiers in Econometrics*. Academic Press. New York, NY.

———. 1978. "Modeling the Choice of Residential Location." *Transportation Research Record*, no. 673. https://trid.trb.org/view.aspx?id=87722.

Oppewal, H., H. J. P. Timmermans, and J. J. Louviere. 1997. "Modelling the Effects of Shopping Centre Size and Store Variety on Consumer Choice Behaviour." *Environment and Planning A* 29 (6): 1073–1090.

Rasouli, S., and H. Timmermans. 2013. "Assessment of Model Uncertainty in Destinations and Travel Forecasts of Models of Complex Spatial Shopping Behaviour." *Journal of Retailing and Consumer Services* 20 (2): 139–146.

Schmid, B., S. Schmutz, and K. W. Axhausen. 2016. "Exploring the Choice between in-Store and Online Shopping." https://www.ethz.ch/content/dam/ethz/special-interest/baug/ivt/ivt-dam/vpl/reports/ab1144.pdf.

Simmonds, D., and O. Feldman. 2011. "Alternative Approaches to Spatial Modelling." *Research in Transportation Economics* 31 (1): 2–11.

Suárez, A., I. R. del Bosque, J. M. Rodríguez-Poo, and I. Moral. 2004. "Accounting for Heterogeneity in Shopping Centre Choice Models." *Journal of Retailing and Consumer Services* 11 (2): 119–129.

Swedish Transport Administration. 2016. "Samgods User Manual V1.0." 2014/26249. Accessed November 21. http://www.trafikverket.se/contentassets/93a250fc470d426998f0a496fea3787a/samgods_user_manual_v1.0.pdf.

Timmermans, H., A. Borgers, and P. van der Waerden. 1992. "Mother Logit Analysis of Substitution Effects in Consumer Shopping Destination Choice." *Journal of Business Research* 24 (2): 177–189.