Urban, Planning and Transport Research
An Open Access Journal

ISSN: (Print) 2165-0020 (Online) Journal homepage: https://www.tandfonline.com/loi/rupt20

Knowing the past, seeing the future - an exploratory study on the viability of retail patronage models based on revealed behaviour

Heli Marjanen, Meri Malmari, Janne Engblom & Anna-Maija Kohijoki

To cite this article: Heli Marjanen, Meri Malmari, Janne Engblom & Anna-Maija Kohijoki (2020) Knowing the past, seeing the future - an exploratory study on the viability of retail patronage models based on revealed behaviour, Urban, Planning and Transport Research, 8:1, 98-124, DOI: 10.1080/21650020.2020.1713209

To link to this article: https://doi.org/10.1080/21650020.2020.1713209

Published online: 28 Jan 2020.

URBAN, PLANNING AND TRANSPORT RESEARCH
2020, VOL. 8, NO. 1, 98–124
https://doi.org/10.1080/21650020.2020.1713209

Routledge
Taylor & Francis Group

ARTICLE

OPEN ACCESS Check for updates

Knowing the past, seeing the future - an exploratory study on the viability of retail patronage models based on revealed behaviour

Heli Marjanen, Meri Malmari, Janne Engblom and Anna-Maija Kohijoki

Turku School of Economics, University of Turku, Turku, Finland

ABSTRACT
Understanding the viability and impact of out-of-town shopping centres is of equal importance for property developers, retailers, and city planners. In the current study, we aim to model and predict behavioural loyalty (operationalised as relative shopping frequencies, RSF) for an existing out-of-town shopping centre. Two separate sets of quantile regression models, one consisting of demographic independents and the other behavioural independents, were constructed. Two datasets, sampled in 2006 and 2011, enabled evaluation of the predictive power of our models and independents. To compare the performance of our explanatory variables over time, models with interaction were used. The results indicate that in the short-run (5–15 years), forecasts based on the current retail provision and consumer demographics together with information on proposed retail agglomerations in the area are likely to give sufficient information about the future viability of a shopping centre.

ARTICLE HISTORY
Received 9 October 2019
Accepted 2 January 2020

KEYWORDS
Retail patronage; out-of-town shopping centres; urban planning; quantile regression; ratio method; impact studies

1. Introduction

The viability and impact of large-scale shopping agglomerations, especially at out-of-town/edge-of-town locations, constitute a crucial issue for developers, retailers, governments, and local planners alike. Although local governments and private sector investors may have partly different goals, the former trying to maximise the public welfare and balanced growth of the region, and the latter trying to maximise profit, they both face the same issues when striving to foresee the changes in consumer behaviour and, respectively, in the retail environment (Ozuduru & Guldmann, 2013). As long planning horizons (more often than not, from 5 to 15 years) are typical for large-scale retail developments like shopping centres, a clear understanding of the ongoing evolution of the retail environment in parallel with accompanying changes in patronage behaviour is vital although poorly embedded in the extant literature. From the retailer's point of view, the stricter planning legislation in Europe since the mid to late 1990s, especially in relation to the large stand-alone edge of town formats, has resulted in a growing demand for more sophisticated location planning tools to illustrate the impacts of the proposed agglomerations to the planners and other authorities. In turn, from the planner's point of view, the growing competition, the escalating number of retail channels, and the increasing size of the proposals have led to an urgent need for better tools to evaluate and

CONTACT Meri Malmari meri.malmari@utu.fi Turku School of Economics, University of Turku, Turku, Finland

understand their expected impact on city centres and existing service network. (Birkin, Clarke, & Clarke, 2017, p. 7.)

In the current study, we pursue to model and predict retail patronage, operationalized as relative shopping frequencies (RSF), for an out-of-town shopping centre using demographic and behavioural variables that have been found powerful in explaining store choice and/or visit frequencies in earlier studies (e.g. Clarke et al., 2006; El-Adly, 2007; Marjanen, Engblom, & Malmari, 2013; Moutinho & Hutcheson, 2007; Pan & Zinkhan, 2006; Teller, 2008; Wakefield & Baker, 1998). In addition to the store choice decisions, the concept of patronage includes the frequency of visits to shopping destinations available to a consumer (Pan & Zinkhan, 2006; Teller, Wood, & Floh, 2016). As consumers are not homogeneous in their choice of shopping destinations, the importance of shopping-centre characteristics to a consumer depends on several consumer-specific attributes such as age, stage of life, household composition, income, and access to a car (El-Adly, 2007).

In concert with developing theoretical knowledge, academic scholars should be more involved in applying this knowledge in practice (see, e.g. Parker, Ntounis, Millington, Quin, & Rey Castillo-Villar, 2017). Therefore, we aim to contribute by providing policy-makers, community planners, property developers, and retailers hands-on methods that would enable data-driven decision making at a reasonable cost. In the case of large-scale retail agglomeration, the consequences of these decisions are likely to have significant effects on their particular localities (for a thorough conceptual discussion about retail agglomerations, see, e.g. Teller, 2008; Teller et al., 2016).

Many of the earlier studies have applied some variant of the family of structural equation models or other methods based on least squares estimation, which provide convenient tools for estimating conditional mean models. More recently, a technique called quantile regression (Koenker & Bassett, 1978) has been advocated for estimating models for conditional quantile functions, like the heavy- or light-users of a shopping establishment (Koenker & Hallock, 2001). Koenker and Hallock (2001) illustrate the primary feature of quantile regression through a classical empirical application in economics, Engel's (1857) analysis of the relationship between household food expenditure and household income. Using Engel's household data, Koenker and Hallock (2001) demonstrate how estimated quantile regression line for median and the least-squares estimate of the conditional mean function have quite different fits if the distribution of the data is skewed. The non-robustness of least-squares fit provides some rather poor estimates while quantile regression lines perform better. As the distribution of RSFs in our data is heavily right-skewed, least squares regression techniques would lead to problems similar to those described by Koenker and Hallock (2001). Thus, the quantile regression approach was chosen. Due to our repeated cross-sectional measurement over two points of time, t_1 (2006) and t_2 (2011), we are able to predict the RSFs at t_2 based on the information available at t_1, and then apply the data collected at t_2 to evaluate the predictive power of the models.

The structure of the paper is as follows. After defining the research task in this introductory section, we present our research setting and the data. Next, the concept of relative shopping frequencies (RSF) is explained. We continue by describing our two sets of quantile regression models, the demographic and the behavioural, which were constructed to model the estimates for the medians (Md) and upper quartiles (Q_3) of the RSF for the shopping centre Mylly in 2006 and 2011. Next, the viability of the models is assessed by comparing the modelled and predicted values with observed values from the respective years. Out-of-sample estimates for 2011, based on 2006 data, are presented to demonstrate the predictive power of our models. Finally, to formally investigate the changes in the explaining power and behaviour of our selected independents during the five-year-long study period, demographic and behavioural models with interaction are constructed. We conclude by critically discussing the practical implications of our findings, and their contribution to the academic knowledge in the field of patronage studies, especially when considering the viability and impact of planned retail agglomerations.

2. The setting

During our study period, the number of shopping centres in Finland increased from 52 to 80. Simultaneously, their market share grew from around 13% to 14% (Finnish Shopping Centers, 2007, 2012). In line with the national development, the retail landscape in the study area, located on the south-west coast of Finland (Figure 1), has changed profoundly over the past two decades. The first out-of-town shopping centre in the area, Mylly, was opened in 2001. It was located about eight kilometres north-west of Turku centre, which has traditionally been the most prominent commercial agglomeration in the area. The next out-of-town shopping centre, Skanssi, located about five kilometres south-east of

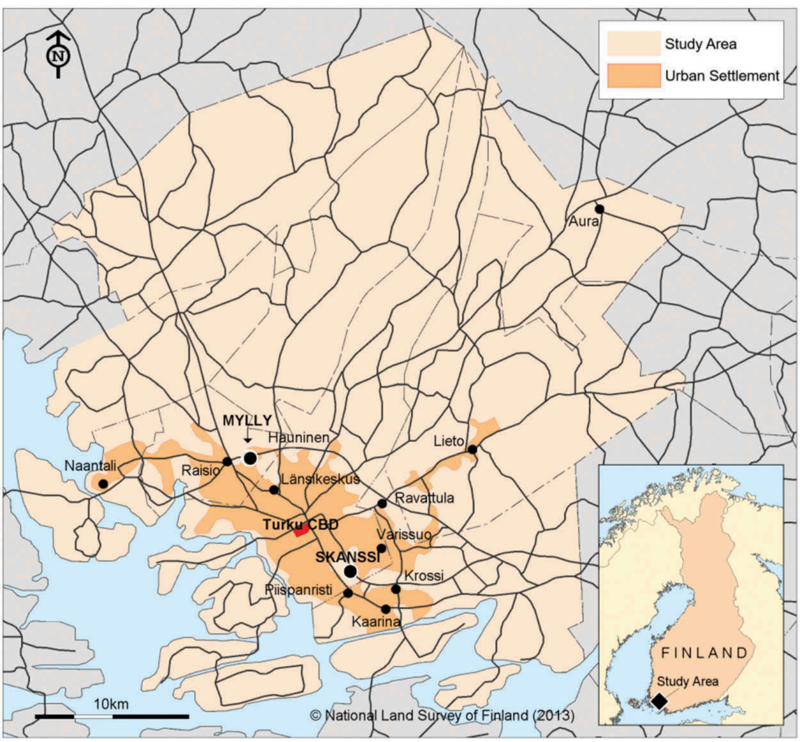

Figure 1. The study area, the road network, and retail agglomerations included in the questionnaires.

Table 1. Key figures for Mylly and Skanssi in 2006 and 2011.

	Mylly 2006	Mylly 2011	Skanssi 2011
Total area (m^2)	64,000	64,000	52,000
GLA (m^2)	47,000	47,000	37,000
Grocery stores (m^2)	4,000	6,100	3,600
Visitors (million)	4.6	4.6	2.9
Sales (million €)	162	184	97
Number of tenants	91	97	87
Parking spaces	3,000	3,000	2,400
- of which covered	1,500	1,500	1,900

Source: Finnish Shopping Centers (2007, 2012)

Turku centre, was opened in 2009. Turku centre, likewise the minor town centres of Raisio and Kaarina, has been experiencing a steep decline in its relative importance as a retail destination since the 1990s when the first edge-of-town retail agglomerations emerged (Marjanen & Malmari, 2012).

By retail area, Mylly is the largest shopping centre in Southwest Finland (see Table 1). In 2011, it was the sixth-largest in Finland and second-largest outside the capital area. The anchor tenants include a hypermarket, a department store, an internationally renowned clothing store (H&M), and an athletic equipment & apparel dealer (Stadium). Due to the hypermarket Prisma, Mylly is an attractive grocery-shopping destination for households all around the study area.

The emergence of the shopping centre Skanssi was expected to have a significant effect on the retail landscape in the area. In 2011, Skanssi reported nearly 90 tenants and 2.9 million visitors per year (Table 1). The anchor tenants are somewhat similar to those of Mylly, i.e. a hypermarket, a department store, international fashion chains (H&M, KappAhl and Lindex), and an athletic equipment & apparel dealer (Intersport). At the time of our data collection, in addition to the food courts, there were no entertainment facilities in neither of the shopping centres.

3. The data

3.1 Data collection

For the purposes of the larger research project, two cross-sectional household surveys were conducted in 2006 and 2011, respectively. The region under study consists of a leading university town (Turku), together with adjoining smaller cities and townships, and rural areas. To ensure the statistical and spatial comparability of the data, both surveys were conducted following a similar procedure. The questionnaires contained items related to store choice criteria (separately for groceries and non-groceries), shopping frequencies in alternative destinations, behavioural patterns, and information on respondents' socio-demographic background.

In 2006, the questionnaires were mailed to 4,864 households, 1,370 of them being members of a household panel set up in 2001 (for details, see Marjanen et al., 2013). The respondents were selected using stratified random sampling based on the number of households living in each municipality under study. The questionnaires were addressed to the eldest female (if any) in the household. In the cover letter, however, they were asked to be filled in by the person who did most of the grocery shopping. Gift vouchers were raffled among those who returned a completed questionnaire. After one reminder, the 47% response rate was achieved. In 2011, 7,246 questionnaires were sent out, resulting in 2,010 usable responses (response rate 27.7%). As the study area was slightly modified in 2011, to ensure the spatial comparability, only those who lived within the 2006 study area were included in the analyses. Furthermore, respondents aged over 75 were excluded because of their small sample size. Thus, the effective sample sizes were 2,082 and 1,566, respectively. To account for the non-response bias, the demographics of the respondents were compared with the population statistics of the study area (place of residence, age, household size, educations, dwelling type). In both data sets, 85% of the respondents were female. This was expected as women often are the principal grocery shoppers, and as the questionnaires were initially addressed to the female heads of households. Regarding household size, income, and access to a car, the sample profiles in both years were highly similar, although in 2011 the respondents were slightly younger compared to 2006 (see Appendix). In both samples, the 55+ year olds were slightly overrepresented and the aged below 45 years underrepresented. Similarly, two-person households, those with higher education, and living in a house were overrepresented whereas single-family households, blue-collar workers and those living in flats were, in turn, underrepresented. However, considering that sample size and the magnitude of differences in socio-demographic profiles between the sample and the base area, the non-response bias was not considered as an issue (Armstrong & Overton, 1977; Marjanen & Malmari, 2012).

3.2 Relative shopping frequencies

Drawing on the well-known Huff model (Huff, 1964), we define relative shopping frequencies (RSF) as the number of shopping trips directed to a particular destination at a household level, in proportion to the total number of shopping trips conducted by that household. As the study is based on revealed choices, the RSF can be treated as a measure of behavioural loyalty (e.g. Dick & Basu, 1994; McGoldrick, 2002, p. 114; Pleshko & Al-Houti, 2012). Thus, instead of measuring either satisfaction or patronage intention, both being widely used dependent variables in earlier literature, we present the (environmentally constraint) realised outcomes of these two. Although the RSF reveals neither the absolute number of visits nor spending in a destination, in the absence of sales data it can be used as a proxy for market share (Howard, 1992; Marjanen, 1995; Teller et al., 2016).

In the questionnaires, the respondents were asked to report their typical visit frequencies (on household-level) in specified retail destinations[1] using a seven-point scale with anchors 'at least once a week' – 'never'. In addition to the selected brick-and-mortar destinations, e-shopping and mail-order were listed as shopping options. To cover all possible alternatives, the option 'other place' was added. As the frequency data was collected on the ordinal scale, we first transformed it into a quasi-rational numeric scale by replacing 'at least once a week' with 100 visits per year, '2–3 times in a month' with 30, 'once a month' with 12, 'once every two months' with 6, 'less than once every two months' with 3, and 'never' with 0 (for details, see Marjanen, 1995, p. 167). Based on this data, we were able to calculate the RSFs for each household and each destination as shown in equation 1, where c refers to the number of shopping trips by household (j) into each destination (i) per year, and K to the alternative shopping options (K = 13 in 2006, and K = 14 in 2011):

$$S_{ij} = \frac{c_{ij}}{\sum_{k=1}^{K} c_{kj}} \quad (1)$$

As the RSF includes both grocery and non-grocery shopping activities, it is heavily influenced by the more frequent grocery shopping. Despite the scarce population in the vicinity of the shopping centre, Mylly was a popular grocery shopping destination (especially at weekends) all over the study area. During the study period, the share of those heavy-users who reported Mylly as their most often used grocery-shopping destination grew substantially, from 13.8% to 20.7% on weekdays and from 30.7% to 36.4% at weekends.

4. Model building

4.1 Quantile regression approach to retail patronage

As parametric mean models and standard least-squares regression techniques provide summary point estimates for the average effect of the independent variables on the average customer, important features of the underlying relationship might remain unrevealed (Mosteller & Tukey, 1977, p. 266; Coad & Rao, 2008). Therefore, the quantile regression technique was applied as it was expected to provide a more comprehensive picture of the underlying relationship between our selected independents and visit frequency. Moreover, quantile regression works well for data with heavy-tailed or highly skewed distributions (Koenker & Bassett, 1978), which is more often than not the case in the context of retail patronage

The RSFs for the shopping centre Mylly in 2006 are presented in Figure 2. The figure poignantly reveals the heavy right-skewness of the distribution. The quartiles presented in the figure can be used to divide the respondents into four distinctive segments (Koenker & Hallock, 2001) according to their visit frequencies. Each quartile represents 25% of the studied

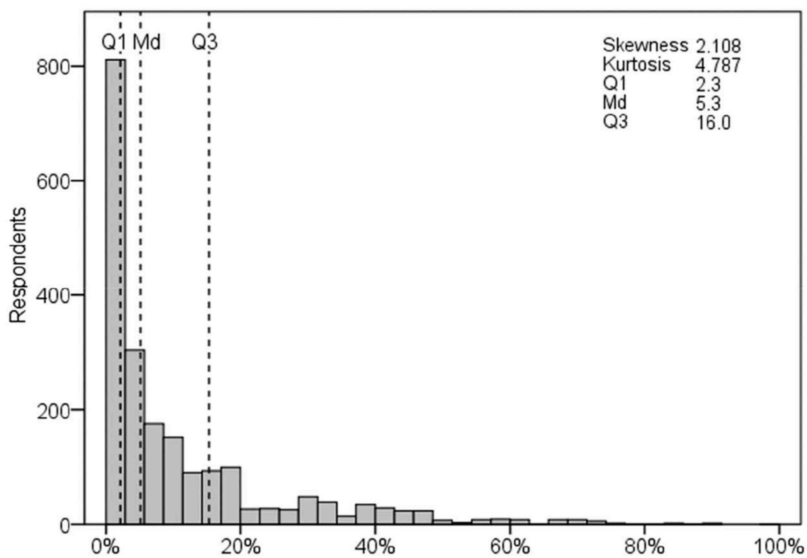

Figure 2. The distribution of RSF for Mylly in 2006.

population: those who visit the centre very seldom or never (RSF<Q_1); those with intermediate (Q_2, here referred to as Md) visit frequencies, and those who are the most loyal patrons of the shopping centre (RSF>Q_3). As the respondents in Q_1 directed 0–2.3% of their shopping trips to Mylly, it was not predicted. Even the median was as low as 5.3%, implying that half of the respondents channelled 95% of their shopping activities elsewhere. The heavy-users, in turn, directed 16–100% of their shopping trips to Mylly. Thus, they make a disproportionately large contribution to the clientele and turnover.

Stata 13 Qreg module was used to construct two separate sets of quantile regression models, the demographic and the behavioural, to produce estimates for Md and Q_3 of RSF in 2006 and 2011. Two separate models were applied to minimise the number of independents per model and thereby to enhance their statistical power and parsimony. Moreover, the character of information and the means to acquire the required data are very different regarding the behavioural and demographic variables, respectively.

4.2 Construction of the demographic models

Distance has been found as one of the key determinants in explaining consumer choice between shopping destinations (e.g. Gomes & Paula 2017; Jackson et al., 2006; Marjanen et al., 2013; Moutinho & Hutcheson, 2007; Pan & Zinkhan, 2006). Similarly, having a car has a direct impact on shopping and store choice as the possibility to use a car implies a wider selection of stores to choose from (Jackson et al., 2006; Kohijoki & Marjanen, 2013; Moutinho & Hutcheson, 2007). Also, age, income, and household size have been frequently found to have an effect on shopping frequencies (Nilsson, Gärling, Marell, & Nordvall, 2015; Roy, 1994). Thus, five independents were selected for the demographic models: *distance* (DIST) from respondent's home address to the shopping centre, *access to a car* (CAR), *household size* (HOU), *household income* (INC), and *age* (AGE) of the respondent. Gender was not included as the study was conducted on a household level, and as the majority of the respondents were female. The demographic model for quantiles $Q(p)$ of RSF ($\hat{Q}(p)$, $p = 0.50, 0.75$) is presented in equation 2, where the expected quantile $\hat{Q}(\mathrm{p})$ for household j is defined by

$$\hat{Q}(p)_j = \alpha + \sum_{i=1}^{2} \beta_{1i} DIST_{i(j)} + \sum_{i=1}^{2} \beta_{2i} CAR_{i(j)} + \sum_{i=1}^{2} \beta_{3i} HOU_{i(j)} + \sum_{i=1}^{2} \beta_{4i} INC_{i(j)} + \sum_{i=1}^{2} \beta_{5i} AGE_{i(j)} \quad (2)$$

Most of the data were collected on interval/rational level. However, to increase the statistical power and the parsimony of the models, and to make the results more decipherable, only categorical variables were used to build the models although. For the same reasons, the number of categories for each variable was reduced to three (see Appendix).

Distance was categorized using the buffering tool offered by MapInfo Professional. Ring buffers were used instead of street network-based isodistances as they are not affected by the modes of transport or alternative routes. After experimenting with several alternatives, we settled on the solution where the radius of the innermost zone is 3 km from the

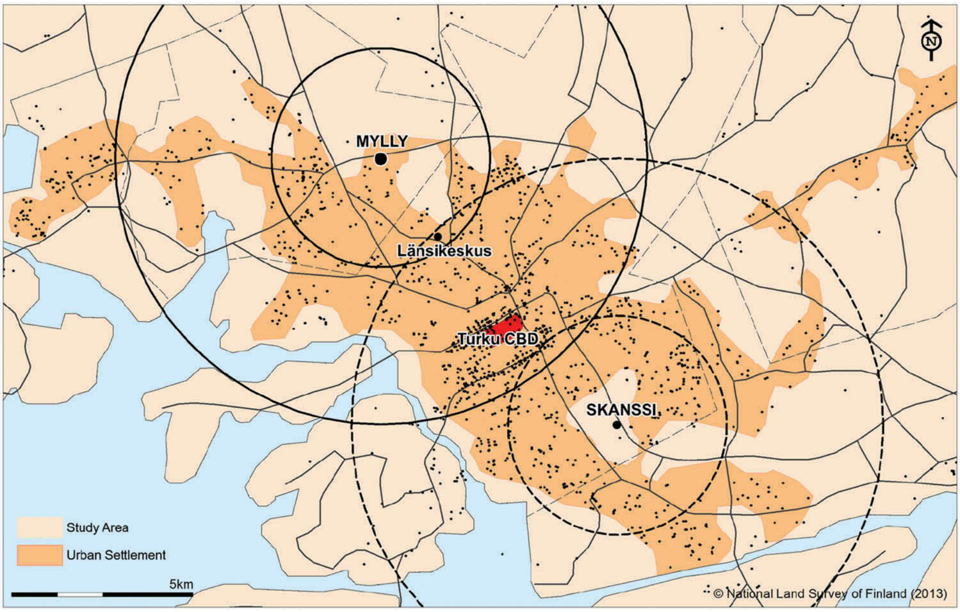

Figure 3. The distribution of the respondents with reference locations of Turku CBD, Mylly and Skanssi. The circles represent the 3 km and 7.3 km distance zones around Mylly and Skanssi.

shopping centre, the middlemost circle covers the area from 3 km to 7.3 km, and the outermost circle ranges from 7.3 km to 35 km (see Figure 3). The most important criterion in defining the zones was the distance separating Mylly from other commercial agglomerations and residential areas. Due to the sparse population in the vicinity of the shopping centre, only eight (2006) to nine per cent (2011) of the respondents lived within the three-kilometre radius while nearly half of them lived at the peripheral zone. To visualise the competitive positions (based on location) of Turku CBD and the two shopping centres, distance zones with similar radiuses for Skanssi were added into Figure 3.

A clear majority of those who reported their mode of transport if visiting Mylly went there by car whereas the travel choices for Turku centre and Skanssi were more versatile (see Table 2). *Access to a car* was categorised into 'no car', '1 car', and '2 or more cars'. One-fifth of the respondents fell in the category of 'no car', whereas 26% had access to two or more cars. Regarding *age*, the respondents were divided into young adults (18–34), middle-aged (35–54), and ageing (55–75). *Household size* was operationalised into single-person households, two-member households, and households with three or more members. Finally, the categories of *household income* were suppressed from nine to three (see Appendix).

Table 2. Travel modes if visiting Turku centre, Mylly and Skanssi.

	On foot (%)		Bus (%)		Car (%)		Other (%)	
	2006	2011	2006	2011	2006	2011	2006	2011
Turku centre (N = 1831/1360)	18.6	19.7	27.6	30.3	46.9	43.8	6.9	6.2
Mylly (N = 1614/1260)	0.9	0.6	7.9	9.4	85.6	83.2	5.6	6.8
Skanssi (N = 997)	-	2.2	-	10.7	-	78.5	-	8.6

4.3 Construction of the behavioural models

In numerous patronage studies, location, parking facilities, store mix, and appealing atmosphere have been identified as key drivers of competitiveness in the retail agglomeration context (e.g. Anselmsson, 2006; Teller et al., 2016). In the questionnaire, the respondents were asked to assess the importance of 47 store choice criteria that have been frequently used in studies attempting to explain either store choice or visit frequencies (e.g. Gomes & Paula, 2017; Moutinho & Hutcheson, 2007; Pan & Zinkhan, 2006). The evaluation task consisted of 23 items for grocery shopping trips (separately for weekdays and weekends), and 24 items for other shopping trips. The assessment was conducted using a 5-point Likert scale with anchors 'not at all important' (1) – 'very important to me' (5). When assessing the importance of each criterion, the respondents were asked to consider the relevance of a criterion in general instead of the actual choice between the shopping options available to them. Thus, the data reveals the shopping orientation expressed by the principal grocery shopper in each household (see, e.g. Laaksonen, 1993). The preferences for weekdays and weekends turned out to be very similar, the only exception being 'location' which was considered more pressing on weekdays. Moreover, there was a substantial amount of missing values for weekends, implying that the respondent either did not shop at weekends or saw no difference between the weekdays and weekends. Thus, the rankings for weekdays and weekends were either combined by calculating their arithmetic means or, in the case of a single ranking, one was selected to represent them both.

A principal component analysis with Varimax rotation, starting with all the 47 criteria from the 2006 data, was conducted to reduce the number of suggested behavioural independents. During the process, altogether 18 items were stepwise deleted because of having rotated factor loadings of less than 0.5 on a single factor. Finally, we settled with an eight-factor-solution presented in Table 3, which contained 29 items (eigenvalue>1) and explained 71% of the total variance (KMO = 0.857; The Bartlett's Test of Sphericity, $p < 0.001$). The literature suggests that items with cross-loadings greater than 0.40 should be eliminated to ensure that each factor would have only one dimension (Hair, Black, Babin, & Anderson, 2010). Based on this rule, the item 'skilled personnel' (groceries) should be eliminated. However, as the solution was logical and easy to interpret, it was kept in the analysis. All retained items had communalities greater than 0.40, indicating sufficient contribution to explaining the variance (Hair et al., 2010).

Two of the factors consisted of items drawn solely from the grocery shopping criteria (*quality & selection_G, convenience_G*), three from the non-grocery shopping criteria (*quality & selection_N-G, dining_N-G*, and *service_N-G*), and three included items from both choice sets (*parking, price*, and *location*). However, based on the purpose of the shopping trip and the magnitude of factor loadings, we decided to split the factor *location* into two separate independents, *LOC_G* for groceries and *LOC_N-G* for non-groceries, both consisting of two criteria from their respective sets. Notably, there is no separate factor for atmosphere-related items, which in several earlier studies have been noted as important patronage inducing variables (Gomes & Paula, 2017; Pan & Zinkhan, 2006; Teller et al., 2016). In the literature, the atmosphere has been found to motivate customers to stay longer, and to have a high impact on shopping centre satisfaction but little influence on visit frequency (Anselmsson, 2006). In our initial 47 item solution, there appeared a factor consisting of items related to atmospherics. However, when items were step by step dropped off, they spread on several dimensions and were subsequently eliminated except the item 'other

Table 3. The eight-factor solution for the behavioural model (2006 data).

Dimension	Groceries/ Non-Groceries	Components	Factor loading	Alpha
Parking	N-G	free parking	0.878	
(PAR)	G	good parking facilities	0.877	
	N-G	good parking facilities	0.875	
	G	free parking	0.866	0.926
Price	N-G	special offers	0.84	
(PRI)	G	special offers	0.837	
	N-G	low price level	0.823	
	G	low price level	0.737	0.87
Quality & Selection_G	G	service counters	0.787	
(Q&S_G)	G	local products in selection	0.742	
	G	high-quality products	0.697	
	G	wide selection	0.661	
	G	skilled personnel	0.622	0.823
Quality & Selection_N-G	N-G	plenty of good speciality stores	0.821	
(Q&S_N-G)	N-G	wide selection	0.817	
	N-G	unique products	0.73	
	N-G	high-quality products	0.671	0.816
Dining_N-G	N-G	variety of cafe and restaurant services	0.885	
(DIN_N-G)	N-G	value-for-money restaurant services	0.861	
	N-G	other customers	0.728	0.809
Convenience_G	G	quick shopping	0.84	
(CON_G)	G	products easy to find	0.793	
	G	everything from the same place	0.667	0.775
Location	G	close to home	0.771	
(LOC_G;	G	convenient location	0.738	0.697
LOC_N-G)	N-G	convenient location	0.604	
	N-G	convenient traffic connections	0.524	0.556
Service_N-G	N-G	skilled personnel	0.812	
(SER_N-G)	N-G	good customer service	0.781	0.805

customers' which is included in factor labelled *dining_N-G* in the eight-factor solution. Finally, a total of nine independents were created by calculating the means of the criteria that loaded highest on their assigned dimensions (see Table 3). Using the means instead of factor scores allowed us to keep a respondent in the analysis despite an occasional missing value. It also allowed the splitting of the factor *location*. To increase the parsimony of the models, the values were classified into categories of *little* (1 – 2.5), *some* (2.6 – 3.5), and *much* (3.6 – 5) effect on store choice (for the frequencies, see Appendix).

The behavioural model for quantiles $Q(p)$ of RSF ($\hat{Q}(p)$,$p = 0.50, 0.75$) is presented in equation 3, where the expected quantile $\widehat{Q}(p)$ for household j is defined by:

$$\hat{Q}(p)_j = \alpha + \sum_{i=1}^{2} \beta_{1i} PAR_{i(j)} + \sum_{i=1}^{2} \beta_{2i} PRI_{i(j)} + \sum_{i=1}^{2} \beta_{3i} Q\&S\_G_{i(j)} + \sum_{i=1}^{2} \beta_{4i} Q\&S\_N\text{-}G_{i(j)} + \sum_{i=1}^{2} \beta_{5i} DIN\_N\text{-}G_{i(j)} + \sum_{i=1}^{2} \beta_{6i} CON\_G_{i(j)} + \sum_{i=1}^{2} \beta_{7i} LOC\_G_{i(j)} + \sum_{i=1}^{2} \beta_{8i} LOC\_N\text{-}G_{i(j)} + \sum_{i=1}^{2} \beta_{9i} SER\_N\text{-}G_{i(j)} \quad (3)$$

4.4 Out-of-sample predictions

To produce out-of-sample predictions for the year 2011, the estimates based on the 2006 data needed to be adjusted for the changes in retail provision in the study area. The retail provision is measured as retail floor space. In the planning context, the absence of ready-to-use information about the amount of the existing retail floor space at a city or municipality level is a frequently encountered problem. Despite the considerable amount of human and economic resources devoted to solving this problem, the existing information remains highly unreliable (Ramboll Finland Oy, 2013). Therefore, both the current figures and future developments have to be estimated using various registers and data sources. In the current study, the ratio method (Rogers, 1992, 2005) was applied for this purpose as described below. The ratio method is frequently used in situations where no or very limited data is available

The ratio method is based on the assumption that the market share of a store (or an entire shopping centre) is equal to its share of the competing retail floor space in the area under study. Thus, as the RFS can be used as a proxy for market share in the absence of sales data (Howard, 1992). Based on the same logic, it may be used to produce an estimate for the amount of retail floor space. In the current study, the household survey yielded a mean observed RSF of 11.9% for Mylly in 2006, the GLA being 47,000 m^2, which produces an estimate of 395,000 m^2 for the amount of competitive retail floor space in the area ($\hat{\mathrm{A}}_{tot2006} = 47,000m^2/0.119$). To calculate the expected floor space in 2011 ($\hat{\mathrm{A}}_{tot2011}$), those retail development proposals that appeared likely to be realised by 2011[2] were added to this figure, together with a further 10% to cover the growing impact of e-commerce.[3] The global rise of online retailing since late 1990 inevitably changes the retail landscape and the way consumers shop (cf. Wrigley et al., 2015). At the time of our data collection, online retailing was gaining popularity among Finnish consumers. Still, its impact on brick-and-mortar stores remains unclear because of their complex relationship and multi-channel strategies (traditional retailers getting engaged with online retailing, pure online players opening physical stores, click-and-collect, etc.).

Based on above-described calculations, we got an expected total floor space of 601,700 m^2, which in turn gave us an estimated market share of 7,8% for Mylly in 2011. As the observed mean RSF in 2011 was 8.2%, the ratio method was accepted as a valid means for adjusting the models for the changes in the retail environment. Thus, we created a constant (r) as given in equation 4:

$$r = \frac{\mathrm{A}_{Mylly\ 2011}/\hat{\mathrm{A}}_{tot\ 2011}}{\mathrm{A}_{Mylly\ 2006}/\hat{\mathrm{A}}_{tot\ 2006}} \tag{4}$$

When assessing the accuracy of predictions for 2011, it should be noticed that whereas the mean RSF for Mylly decreased by one third during the study period, both Md and Q_3 decreased by half, Md from 5.3% (2006) to 2.3% (2011), and Q_3 from 16.0% to 8.1%. In situations like this, quantile regression is superior to traditional least square methods which assume that the changes in both tails of the frequency distribution are of similar magnitude (Koenker & Machado, 1999). However, as we adjusted the models using the expected total means, the changes in the quantiles of RSFs are not accounted for. Thus, the out-of-sample estimates for 2011 will be systematically slightly overestimated.

When attempting to model the future viability of a shopping destination, in addition to the evolution of the retail provision, the possible alterations in consumer characteristics and related behaviour are of uttermost importance. In actual planning situations, information on past and current behaviour are obtainable from various sources, including surveys, interviews, ethnographies, loyalty card data, big data etc. This data is then used to forecast future behaviour. These forecasts, of course, are highly subjective even if based on sophisticated research on past trends and alternative future scenarios. In the current study, we had the luxury of having access to the observed behavioural data in 2011, which was imported to the models as shown in equation 5.

$$\hat{Q}(p)_{\substack{adjusted \\ 2011}} = \hat{Q}(p)_{\substack{2006\,model \\ 2011\,freq}} \cdot r \tag{5}$$

5. Model performance

As the standard regression fit statistic R^2 measures the relative fit of a model of conditional mean function in terms of residual variance, it is not applicable for quantile regression models. As the independents may exert a significant effect on one tail of the conditional distribution of the response variable but might have no effect in the other tail, the variation explained is likely to vary between quantiles (Koenker & Machado, 1999, p. 1297). In our case, the higher the segment estimated, the higher the reported variation explained by the model. Thus, in the absence of reliable statistical methods, the model fit and predictive power of our models were assessed by visually comparing the observed quantiles of RSF with the respective in-sample and out-of-sample estimates.

5.1 The viability of the demographic models

The model fit and performance of the demographic models are presented and evaluated in Table 4. From the left, the table first shows the observed values (Md and Q_3) and the modelled in-sample estimates (= marginal means of predicted quantiles) for the year 2006 for the categories of independents. Next, the same information, supplemented with out-of-sample predictions based on 2006 data, is provided for the year 2011. Statistical significance (sig.) of the differences between the categories is reported for each independent in the in-sample column. Finally, the model fit was assessed by evaluating the differences between the modelled and observed values, and the rank order of categories. The results are interpreted using ‘minus’ and ‘plus’ (from one to three) signs, ‘+++’ indicating minor differences between the modelled and observed values, combined with identical rank orders of categories.

Table 4, in line with the earlier literature, depicts *distance* and *access to a car* as the most powerful independents in the demographic models. For these two, the differences in in-sample RSFs between the categories were highly significant ($p < 0.001$). According to the observed RSFs, the respondents living closest (<3 km) to the shopping centre were its most loyal patrons, the heavy-users (Q_3) directing one third or more of their shopping

Table 4. The demographic model: observed values, in-sample estimates, and out-of-sample estimates (2011) of medians (Md) and upper quartiles (Q_3) of RSF for Mylly.

	Md 2006		Q3 2006		Md 2011			Q3 2011		
	obs. value	in-sample	obs. value	in-sample	obs. value	in-sample	out-of-sample	obs. value	in-sample	out-of-sample
Distance (DIST)										
< 3 km	0.17	0.16	0.33	0.31	0.12	0.13	0.11	0.29	0.26	0.20
3–7.3 km	0.05	0.06	0.13	0.15	0.02	0.03	0.04	0.05	0.08	0.09
7.3–35 km	0.05	0.06	0.14	0.14	0.03	0.03	0.04	0.07	0.08	0.10
sig.		0.000		0.000		0.000			0.000	
Model fit[a]		+++		++		+++	+++		++	++
Access to car (CAR)										
no car	0.03	0.03	0.05	0.06	0.01	0.02	0.02	0.02	0.04	0.04
1 car	0.06	0.07	0.15	0.15	0.03	0.04	0.04	0.08	0.08	0.10
2 or more cars	0.11	0.11	0.27	0.24	0.05	0.06	0.07	0.17	0.16	0.16
sig.		0.000		0.000		0.000			0.000	
Model fit[a]		+++		+++		+++	+++		+++	+++
Household size (HOU)										
1 person	0.03	0.07	0.10	0.15	0.02	0.04	0.03	0.04	0.09	0.07
2 persons	0.05	0.06	0.16	0.14	0.03	0.04	0.04	0.09	0.09	0.10
3 or more persons	0.09	0.09	0.24	0.20	0.05	0.05	0.07	0.15	0.11	0.16
sig.		0.003		0.001		0.442			0.508	
Model fit[a]		+		+		++	++		+	++
Income (INC)										
INC1	0.03	0.07	0.10	0.16	0.02	0.04	0.03	0.05	0.09	0.07
INC2	0.05	0.07	0.15	0.16	0.03	0.04	0.05	0.09	0.10	0.11
INC3	0.08	0.07	0.19	0.15	0.03	0.04	0.06	0.12	0.09	0.12
sig.		0.497		0.742		0.901			0.918	
Model fit[a]		-		-		+++	++		-	++
Age (AGE)										
18–34	0.08	0.09	0.17	0.18	0.03	0.05	0.05	0.08	0.10	0.10
35–54	0.07	0.07	0.18	0.16	0.04	0.04	0.06	0.11	0.10	0.12
55–75	0.04	0.06	0.11	0.14	0.02	0.03	0.04	0.06	0.09	0.09
sig.		0.000		0.070		0.084			0.392	
Model fit[a]		++		+		+	++		-	++

aNote
'+++' minor differences between modelled and observed values; rank order of categories identical.
'++' some differences between modelled and observed values; rank order of categories identical.
'+' differences between modelled and observed values; differences in rank order of categories.
'-' differences between modelled and observed values; reverse rank order of categories.

activities to Mylly in 2006. The power of proximity was even more eminent in the 2011 data. In turn, the differences between the second and third category of distance were modest to negligible, the low RSFs in the 3–7.3 km category being explained by the multitude of intervening opportunities in the densest populated parts of this distance zone (see Figure 3).

Regarding the independent *access to a car*, the differences between the categories were of considerable size and statistically significant at 0.001 level. The model fit was very good, the importance of having access to at least one car and high RSF of two-car households being well detected by the models. This independent also produced the most accurate out-of-sample estimates for 2011.

The differences between the categories of *household size, income*, and *age* were not statistically significant ($p < 0.05$) except for *household size* in 2006. Moreover, their in-sample model fit for Q_3 was modest to poor. The observed values suggest that the heavy-users consisted of middle-aged households with higher than average income, but the models were unable to capture the effects of these independents. Despite the poor in-sample model fit, they produced fairly accurate out-of-sample estimations for 2011. A potential explanation for this anomaly is collinearity between these independents. Furthermore, Wakefield and Baker (1998) have suggested that age might be a moderating factor for consumer response to retail environments. Following the same logic, this could be stretched to cover household size and income as well.

5.2 The viability of the behavioural models

Following the procedure described above, the model fit and performance of the behavioural models are presented and evaluated in Table 5. As the table shows, only for the independent *parking*, and *dining_N-G* (2006), the differences between categories were statistically significant ($p < 0.05$). *Parking* also produced very accurate in-sample estimates and, accordingly, very accurate out-of-sample predictions for 2011, indicating that those to whom parking facilities were important were the most loyal patrons of Mylly. The independent *dining_N-G* refers to the social dimension of shopping (other customers, dining as a social activity) but it also includes aspects of convenience (value-for-money restaurant services; no need to go somewhere else for catering services or to cook at home). Both the academic studies and evidence from the industry highlight the trend of shopping centres investing increasingly in dining facilities and food-courts (Sit, Merrilees, & Birch, 2003; Finnish Shopping Centers, 2012; 2013; Teller et al., 2016). The data in the current study, however, indicate only a modest increase in the importance of dining as a shopping destination choice criteria (see Appendix). In both samples, the highest observed RSFs and modelled estimates for RSFs were found in the category of ‘much’ of *dining_N-G*. However, its performance as an explanatory variable decreased in 2011 as the differences between categories decreased.

The observed values for the independent *price* show that those respondents to whom low prices and/or good special offers were of importance were more likely to visit Mylly. As this was not reflected in the models, the model fit in both years was deemed poor. During the study period, price-consciousness in general increased as the number of respondents who fell in the category of ‘little’ decreased from 12% to 7% and, respectively, in the category of ‘much’ increased from 56% to 69%.

Table 5. The behavioural model: observed values, in-sample estimates, and out-of-sample estimates (2011) of medians and upper quartiles of RSF for Mylly.

	Md 2006		Q3 2006		Md 2011			Q3 2011		
	obs. value	in-sample	obs. value	in-sample	obs. value	in-sample	out-of-sample	obs. value	in-sample	out-of-sample
Parking (PAR)										
Little	0.03	0.03	0.06	0.08	0.01	0.02	0.02	0.02	0.03	0.05
Some	0.05	0.05	0.14	0.13	0.03	0.03	0.03	0.08	0.06	0.10
Much	0.08	0.08	0.19	0.20	0.04	0.04	0.06	0.11	0.11	0.14
sig.		0.000		0.000		0.003			0.000	
Model fit[a]		+++		+++		+++	+++		+++	++
Price (PRI)										
Little	0.05	0.08	0.13	0.17	0.02	0.03	0.04	0.05	0.08	0.11
Some	0.05	0.07	0.14	0.16	0.02	0.03	0.04	0.07	0.08	0.11
Much	0.06	0.06	0.17	0.16	0.03	0.03	0.05	0.09	0.09	0.12
sig.		0.272		0.871		0.771			0.945	
Model fit[a]		+		+		++	++		+	+
Quality & Selection_G (Q&S_G)										
Little	0.04	0.06	0.09	0.15	0.02	0.03	0.04	0.07	0.09	0.09
Some	0.05	0.07	0.16	0.17	0.03	0.03	0.05	0.08	0.09	0.12
Much	0.06	0.06	0.17	0.16	0.03	0.03	0.05	0.08	0.08	0.12
sig.		0.675		0.761		0.739			0.940	
Model fit[a]		+		+		+++	++		++	++
Quality & Selection_N-G (Q&S_N-G)										
Little	0.03	0.06	0.10	0.16	0.01	0.01	0.03	0.03	0.06	0.10
Some	0.04	0.06	0.13	0.15	0.02	0.02	0.04	0.07	0.07	0.11
Much	0.06	0.07	0.17	0.17	0.03	0.03	0.05	0.09	0.09	0.12
sig.		0.139		0.665		0.082			0.487	
Model fit[a]		++		+		+++	++		++	+ +
Dining_N-G (DIN_N-G)										
Little	0.05	0.05	0.12	0.14	0.02	0.03	0.04	0.07	0.08	0.10
Some	0.08	0.08	0.18	0.17	0.03	0.03	0.05	0.09	0.09	0.12
Much	0.09	0.09	0.19	0.21	0.03	0.04	0.06	0.11	0.10	0.14
sig.		0.000		0.009		0.326			0.542	
Model fit[a]		+++		++		+++	++		++	++
Convenience_G (CON_G)										
Little	0.03	0.06	0.11	0.15	0.02	0.03	0.03	0.05	0.08	0.08
Some	0.05	0.07	0.15	0.16	0.03	0.03	0.04	0.07	0.09	0.11
Much	0.06	0.07	0.17	0.17	0.03	0.03	0.05	0.09	0.09	0.12

(Continued)

Table 5. (Continued).

	Md 2006		Q3 2006		Md 2011			Q3 2011		
	obs. value	in-sample	obs. value	in-sample	obs. value	in-sample	out-of-sample	obs. value	in-sample	out-of-sample
sig.		0.415		0.731		0.898			0.978	
Model fit[a]		+		++		+++	++		+	++
Location_G (LOG_G)										
Little	0.05	0.07	0.17	0.18	0.03	0.04	0.04	0.11	0.11	0.13
Some	0.07	0.07	0.18	0.19	0.03	0.03	0.05	0.10	0.10	0.13
Much	0.05	0.06	0.14	0.15	0.02	0.03	0.04	0.07	0.08	0.11
sig.		0.038		0.068		0.261			0.232	
Model fit[a]		++		+++		+++	++		+++	++
Location_N-G (LOG_N-G)										
Little	0.03	0.07	0.17	0.23	0.03	0.03	0.05	0.08	0.10	0.15
Some	0.06	0.07	0.15	0.17	0.03	0.03	0.05	0.09	0.09	0.12
Much	0.05	0.06	0.17	0.15	0.02	0.03	0.04	0.08	0.08	0.10
sig.		0.168		0.050		0.372			0.727	
Model fit[a]		++		+		+++	+++		-	-
Service_N-G (SER_N-G)										
Little	0.05	0.07	0.11	0.17	0.03	0.04	0.04	0.07	0.09	0.10
Some	0.07	0.07	0.18	0.18	0.03	0.03	0.05	0.08	0.09	0.12
Much	0.05	0.06	0.16	0.15	0.02	0.03	0.04	0.09	0.08	0.11
sig.		0.060		0.270		0.384			0.862	
Model fit[a]		++		+		+++	++		-	+

aNote:
'+++' minor differences between modelled and observed values; rank order of categories identical.
'++' some differences between modelled and observed values; rank order of categories identical.
'+' differences between modelled and observed values; differences in rank order of categories.
'-' differences between modelled and observed values; reverse rank order of categories.

The independent *quality & selection_G* was drawn from the choice criteria for grocery shopping destinations. The observed values for 2006 reveal that those who valued high quality, wide selection and personal service while shopping for groceries were more likely to frequent Mylly compared to those to whom they were of minor importance. The model was not able to identify this effect which, however, was no more to be detected in the 2011 data. Despite the poor in-sample model fit, *quality & selection_G* produced fairly accurate out-of-sample predictions for 2011. When assessing the model fit, the high share of respondents in the category of 'much' and, respectively, low in the category of 'little' should be noticed. The corresponding independent for the non-grocery shopping context, *quality & selection_N-G*, performed very similar to its grocery-counterpart. Also, the independent *service_N-G*, which was drawn from the non-grocery choice criteria, performed very similarly to *quality & selection_G*, although the respondents were more evenly distributed in the categories of *service_N-G*.

The importance of convenience in the grocery-shopping context increased substantially during the study period as 82% of the respondents fell to category 'much' in 2011, compared to 63% in 2006. In both samples, the observed values show that those to whom *convenience_G* (referring to one-stop shopping, and time-saving in general) was important were more likely to fall into the heavy-user category of Mylly. Although the models were not able to detect this, and thus, the model fit was rather poor, the predictions for 2011 were reasonably accurate.

In previous studies, proximity/location has appeared as the most important determinant of shopping centre visit frequency (e.g. Anselmsson, 2006; Gomes & Paula, 2017; Marjanen, 1995; Marjanen et al., 2013). This notion was only partly supported by the current study. The observed values show that, in line with the previous research, those who appreciated the grocery store being close to home and/or conveniently located were less likely to choose Mylly over other shopping options. Although the model fit for *location_G* was good to excellent, the differences between categories were modest and not statistically significant. Moreover, the independent *location_N-G* had a very modest impact on observed RSFs. A likely explanation for the negligible observed importance of *location_N-G* is that 'convenient traffic connections' is associated with public transport. In contrast, the overwhelming majority of the customers in Mylly are car-borne. However, in 2006 the in-sample estimates for Q_3 suggested a stronger effect on RSF, which resulted in poor model fit and, consequently, in poor out-of-sample predictions for 2011. According to the ratings given to the items forming these independents, *location_N-G* became less important, whereas *location_G* slightly gained importance during the study period (see Appendix).

6. Models with interaction

To statistically test the differences between the two points of time (YEAR = 2006, YEAR = 2011), interaction terms for each independent and each year were added to the models. The models analyse whether there are statistically significant differences in the way the independents explain the RSFs. The demographic model with interaction for quantiles $Q(p)$ of RSF ($\widehat{Q}(p), p = 0.50, 0.75$) is presented in equation 6, where the expected quantile $\widehat{Q}(p)$ for household j is defined by:

$$\begin{aligned}\hat{Q}(p)_j &= \alpha + \sum_{i=1}^{2}\beta_{1i}DIST_{i(j)} + \sum_{i=1}^{2}\beta_{2i}CAR_{i(j)} + \sum_{i=1}^{2}\beta_{3i}HOU_{i(j)} + \sum_{i=1}^{2}\beta_{4i}INC_{i(j)} \\ &+ \sum_{i=1}^{2}\beta_{5i}AGE_{i(j)} + \sum_{i=1}^{2}\beta_{1\mathrm{Y}i}DIST_{i(j)}*YEAR_{(j)} + \sum_{i=1}^{2}\beta_{2\mathrm{Y}i}CAR_{i(j)}*YEAR_{(j)} \\ &+ \sum_{i=1}^{2}\beta_{3\mathrm{Y}i}HOU_{i(j)}*YEAR_{(j)} + \sum_{i=1}^{2}\beta_{4\mathrm{Y}i}INC_{i(j)}*YEAR_{(j)} \\ &+ \sum_{i=1}^{2}\beta_{4\mathrm{Y}i}AGE_{i(j)}*YEAR_{(j)} \end{aligned} \tag{6}$$

The behavioural model with interaction was constructed accordingly. The analysis revealed statistically significant differences between the estimated quantile profiles for the Md and Q_3 of *distance, household size*, and *parking* ($p < 0.01$). Besides, there were statistically significant differences in the Md profiles of *dining_N-G* ($p < 0.01$), and *access to a car* ($p < 0.05$) between the years 2006 and 2011. The in-sample estimates and observed values for the categories (Q_3) of these independents are visualised in Figure 4. It should be noted that as the analyses are based on in-sample estimates, the results are meaningful only regarding independents with good model fit. In turn, if the model is unable to detect the observed variation between the categories of an independent, it would not be noticed here either.

As Figure 4 shows, the decrease in RSFs was modest regarding those respondents who lived in the vicinity of the shopping centre (DIST<3 km) and grew more substantial the further away from it they lived. However, both in 2006 and 2011, over 40% of the heavy-users lived within the 7.3–35 km distance zone. It should also be noted that only 3% of the respondents directed more than half of their shopping activities to Mylly. The overall decrease in RSFs is attributable to the substantial change in the retail environment (including the emergence of the new competitor, Skanssi), whereas the more modest decline in <3 km distance zone is, arguably, explained by grocery shopping from the nearby areas.

In both samples, households with two or more cars were more likely to choose Mylly compared to those with access to no/one car. Although their RSF in Mylly decreased most dramatically during the study period, the share of multi-car households grew among the heavy-users of Mylly. However, the segmenting power of *parking* slightly diminished during the study period although the items forming this independent (free parking, good parking facilities) gained importance as choice criteria, and performance of Mylly regarding these criteria was rated higher in 2011 compared to 2006. Thus, despite the stated importance, other store choice criteria seem to have been more pressing when the actual patronage decisions were made. Moreover, since 2009, also the shopping centre Skanssi offers free and spacious parking in a location which is more convenient to many of those living within the 7.3–35 km distance zone.

The most notable decrease in the segmenting power is to be seen when the respective quantile profiles (Q_3) of *dining_N-G* in 2006 and 2011 are compared. As the ratings for Mylly on criteria forming this independent lowered during the study period, we might suggest that those to whom catering facilities and fellow customers are of importance found more attractive offerings at other places.

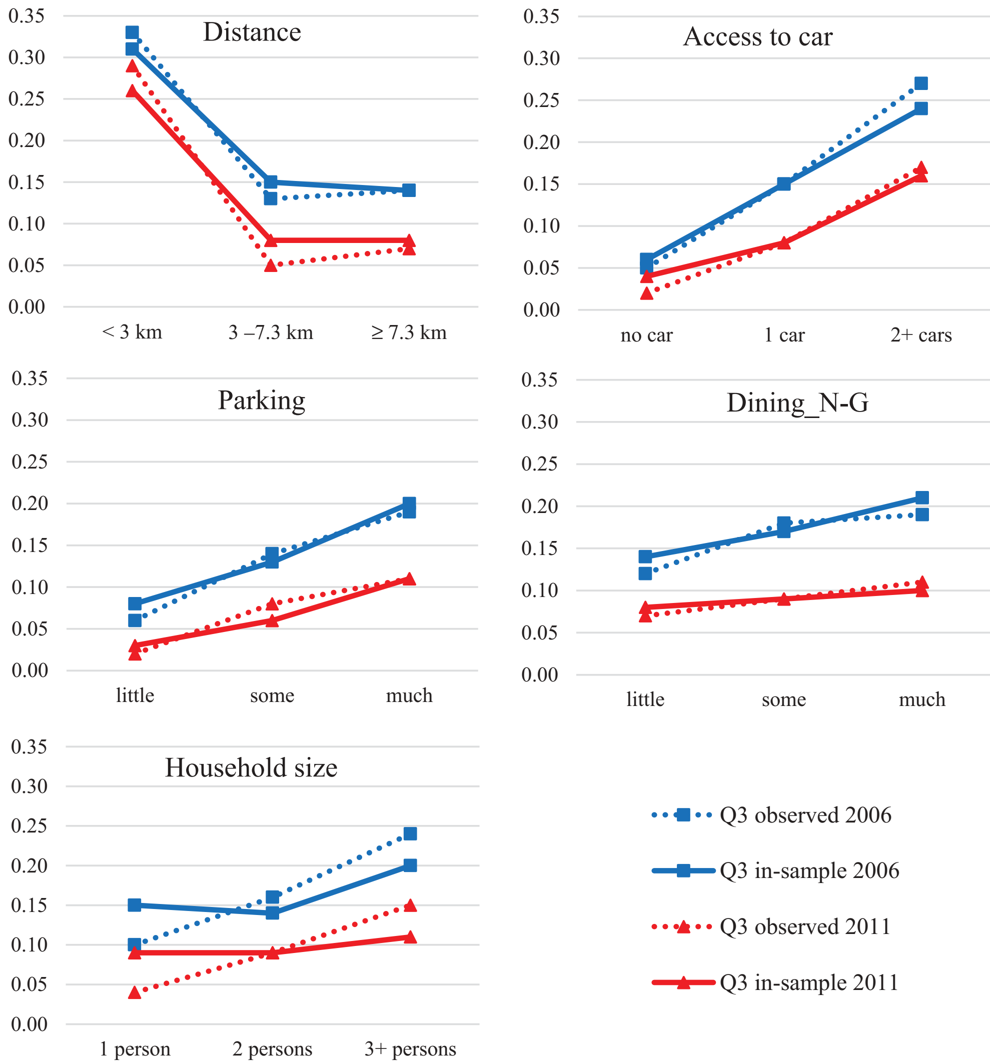

Figure 4. The modelled and observed quantile profiles (Q_3) of distance, access to a car, parking, dining, and household size in 2006 and 2011.

The model fit for *household size* was modest at best as the model failed to detect the dependence of household size and the RSF. Thus, the relevance of the differences found by models with interaction is low despite the reasonably accurate out-of-sample predictions. However, the profiles of the observed RSFs and modelled estimates, respectively, were highly similar in both years.

7. Conclusions

7.1 Discussion of study results

The current study contributes to the existing research on patronage behaviour in the context of retail agglomerations in several unique ways. Most importantly, we managed to formally investigate and forecast the effects of the evolution of a particular retail environment by combining them with simultaneous changes in consumer demographics and behaviour. As the most efficient predictors for RSF in our study were the distance, access to a car, and parking, despite the claims arguing 'the death of distance' (Cairncross, 1997), our findings further underline the overwhelming importance of distance and accessibility in patronage behaviour. The results also indicate that in the short-run (5–15 years), forecasts based on the current retail provision and consumer demographics in area, together with information about the size, location and level of specialisation of the proposed retail agglomerations, are likely to give sufficient information about the impacts and/or viability of those proposals.

Our results suggest that despite the considerable changes in the retail environment (e.g. the emergence of a major competitor), and the more modest changes in consumer demographics in the area, the models based on data collected in 2006 produced fairly accurate predictions for 2011 when fitted using the ratio method (Rogers, 2005). In line with our findings, in a recent study on consumer experiences in physical stores, Bäckström and Johansson (2017) reported, based on data collected in 2006 and 2016, that although the retail sector changed dramatically during their ten-year-long study period, the in-store experiences were largely created by the same aspects in both studies. Even in today's multi-channel shopping environment where retail offerings are becoming more and more diversified, traditional values such as personnel, layout, and display of products persisted although they were understood in new ways, and complemented with technological solutions (Bäckström & Johansson, 2017, p. 255–256).

In line with the existing theory, we found that the most efficient predictors for RSF were the distance separating respondents and the shopping centre, access to a car, and parking facilities. As the RSF includes both components of patronage, i.e. the store choice and visit frequencies (Pan & Zinkhan, 2006), it can be regarded as a series of repeated choices where the frequency of visits into each destination is mitigated by the relational nature of the concept. Therefore, the independents found to be more powerful in explaining store choice might be expected to perform better than those having more influence on visit frequencies. Gomes and Paula (2017), in their meta-analyses, identify a 'location/access' dimension of mall image that includes distance, the convenience of travel, and accessibility. In the current study, these were covered by two access-related independents in the behavioural models and by distance in the demographic models. In earlier studies, in addition to location, the retail tenant mix and selection of merchandises have been found significant drivers of patronage, especially store choice (Anselmsson, 2006; Pan & Zinkhan, 2006; Teller et al., 2016). In the context of the current study, neither of them turned out to perform well as segmenting variables. As our dependent variable consists of revealed choices, compared to study designs with stated preferences where measures like 'satisfaction' or 'patronage intention' (including more wishful thinking and wanna-be consumption) are applied, it is more influenced by the constrained everyday choices.

In contrast to much of the existing literature based on non-representative samples of a population (students, internet users, city centre shoppers, etc.), our research is based on a representative sample of households with at least one of the household members aged 18–75. The availability and use of representative samples are equally essential for practitioners and academics, but increasingly challenging. For example, if the use of the internet or the digital services enabled by the internet is studied using only web-based surveys, the voice of those less active or less competent in the digital environment would be missing. Consequently, their needs would not be covered. In the context of shopping centres, convenience (e.g. parking and way-finding) has been found to be of greater importance for the ageing customers whereas the younger put more emphasis on selection and dining (Anselmsson, 2006). In the planning context, future orientation is inherently included. Thus, insights into the habits and behaviours of the young are essential. However, from the public service point of view, it is also vital to secure a sufficient level of services for the growing 55+ consumer segment. From the retailer's point of view, these ever more active and healthy senior citizens form an increasingly lucrative market potential.

To our knowledge, this is the first study to combine the store choice criteria for non-grocery and grocery shopping trips to formally test the power of a set of demographic and behavioural determinants of behavioural loyalty, operationalised as RSFs. Such a holistic approach is especially required in cases where shopping centres are anchored by large-scale food retailers. The repeated cross-sectional samples over time enabled us to formally evaluate the predictive power of the models that were created. Similar data often exists, gathered and stored by various retail consultant companies (e.g. *Turku today* -research series by Taloustutkimus (2018)), but it is only seldom available either to academic researches or public sector planners.

7.2 Practical implications

Our results indicate that the demographic variables outperform the behavioural determinants in identifying the heavy-users and explaining RSF for a regional shopping centre. Very detailed information on the demographic composition of the population is readily available for planners and retailers. If this information can be used to derive information on the store choice criteria (which is only seldom available), this can be used to make more informed decisions. For example, whereas our results underline the importance of parking facilities (Gomes & Paula, 2017; Teller et al., 2016), they also suggest that knowing the number of cars in the household (access to a car) gives the same information in a much easily accessible form. From the planner's point of view, such findings are utterly valuable when evaluating the viability and impact of planning proposals. In practice, the empirical data supporting the decision-making process is usually acquired by hiring an experienced consultant.

7.3 Study limitations

Forecasting models reduce the degree of subjectivity but do not remove it. Moreover, all models are highly dependent on the quality of data and expertise employed in their development and application (Rogers, 2005). The current study was based on two extensive household sample, making it potentially suspect for non-response bias. This was accounted for by carefully comparing the sample profiles with the demographic information available from the base area. Because of our data collection (the questionnaires addressed to the eldest female in the household), majority of the respondents were female. However, this was not considered as a problem because a similar procedure was followed in both samples. The same applies to the concerns that our data might be dated. As our final goal was to develop and test models that might be helpful in reducing the uncertainty inevitably inherent when forecasting the future, our two datasets were suitable for that purpose.

In the planning context, if the aim was to provide the decision-makers with practical tools, the decision to apply the frequencies of categories of independents from the 2011 data to adjust the models for predictions may sound odd at first sight. However, in real life, very similar data is available from several sources.

Although the empirical study was limited to one particular shopping centre in Finland, the literature referred to in the study shows that the results can be seen as reflecting more general tendencies in store choice behaviour. Thus, they are of interest also for a wider audience of researchers and practitioners.

7.4 Future research

Although new shopping centre development has slowed down in several European counties, the total amount of shopping centre floorspace continues to increase. Finland has been one of the most active European countries regarding the shopping centre development, supported by growing urbanisation and tourism. (European Shopping Centres: The development story, 2018.) Since 2011, over twenty new shopping centres have emerged, topped with the extensions of the existing ones. Their current market share is around 17% of the total retail turnover (Finnish Shopping Centres 2012; 2019). Thus, their viability and impact will be a relevant research issue also in the future.

In addition to repeating the current study, a possible extension would be to forecast the market share of a planned new centre by using the information available from an existing one. In further studies, replacing the shopping frequencies with the money spent in alternative shopping destinations would be worth testing. That would also allow the empirical investigation of the relationship between shopping frequencies and market share (Howard, 1992).

From the retail patronage point of view, it would be interesting to apply quantile regression models in a setting where the heavy-users direct a higher proportion of their spending to a single outlet. That might be a hypermarket or a grocery superstore. In our current data, the variation explained by the model increased the higher the segment estimated was. Thus, data with more committed heavy-users might reveal significant insights.

Notes

1. In 2006, the alternatives were Turku centre, five town/municipality centres, Mylly-area (the shopping centre with adjacent big boxes), Länsikeskus (a retail park), Ravattula (agglomeration of big boxes), Varissuo (a sub-regional commercial centre), post order, e-retailing, and ‘other place’. In 2011, the alternatives included Turku centre, four town centres, the shopping centres Mylly and Skanssi, Länsikeskus, two retail agglomerations, own residential area, post order, e-retailing, and ‘other frequently visited shopping destination’.
2. The data used to estimate the to-be floor space was gathered by an experienced consulting company on behalf of the Regional Council of Southwest Finland in co-operation with the

local retailers (Entrecon, 2006). Our floor space estimates are well in line with those presented by Ramboll Oy (2013) in a report commissioned by the Regional Council of Southwest Finland.

3. A 10% decrease in the demand of retail floor space is frequently used in Finland by retail planners and consultants to cover for the share of e-commerce.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Heli Marjanen http://orcid.org/0000-0002-8015-1217
Meri Malmari http://orcid.org/0000-0001-8461-5392
Janne Englblom http://orcid.org/0000-0002-5680-0993
Anna-Maija Kohijoki http://orcid.org/0000-0002-2011-7009

References

Anselmsson, J. (2006). Sources of customer satisfaction with shopping malls: A comparative study of different customer segments. *International Review of Retail, Distribution and Consumer Research*, *16*(1), 115–138.

Armstrong, J. S., & Overton, T. S. (1977). Estimating nonresponse bias in mail surveys. *Journal of Marketing Research*, *14*(3), 396–402.

Bäckström, K., & Johansson, U. (2017). An exploration of consumers' experiences in physical stores: Comparing consumers' and retailers' perspectives in past and present time. *The International Review of Retail, Distribution and Consumer Research*, *27*(3), 241–259.

Birkin, M., Clarke, G., & Clarke, M. (2017). *Retail location planning in an era of multi-channel growth*. New York: Routledge.

Cairncross, F. (1997). *The death of distance: How the communications revolution will change our lives*. Boston, Massachusetts: Harvard Business School Press.

Clarke, I., Hallsworth, A., Jackson, P., de Kervenoael, R., Perez-del-Aquila, R., & Kirkup, M. (2006). Retail restructuring and consumer choice 1. Longterm local changes in consumer behaviour: Portsmouth, 1980–2002. *Environment and Planning A*, *38*(1), 25–46.

Coad, A., & Rao, R. (2008). Innovation and firm growth in high-tech sectors: A quantile regression approach. *Research Policy*, *37*, 633–648.

Dick, A. S., & Basu, K. (1994). Customer loyalty. Toward an integrated conceptual framework. *Journal of Academy of Marketing Science*, *22*(2), 99–113.

El-Adly, M. I. (2007). Shopping malls attractiveness: A segmentation approach. *International Journal of Retail & Distribution Management*, *35*(11), 936–950.

Engel, E. (1857). Die Produktions- und Konsumptionverhaltnisse des Konigreichs Sachsen. Reprinted in Die Lebenskosten Belgischer Arbeiter-Familien Früher und Jetzt. *International Statistical Institute Bulletin*, *9*, 1–125.

Entrecon. (2006). *Turun seudun kaupan palveluverkkoselvitys 2006* [Retail trade service network report of the Turku area 2006]. Turku: Regional Council of Southwest Finland.

European Shopping Centres: The development story. (2018). *Research & insights. Cushman & Wakefield*. Retrieved from http://www.cushmanwakefield.co.uk/en-gb/research-and-insight/2019/european-shopping-centres-the-development-story

Finnish Shopping Centers. (2007). Helsinki: Finnish council of shopping centers.

Finnish Shopping Centers. (2012). Helsinki: Finnish council of shopping centers. .

Finnish Shopping Centers. (2013). Helsinki: Finnish council of shopping centers.

Finnish Shopping Centres. (2019). Helsinki: Finnish council of shopping centers.

Gomes, R. M., & Paula, F. (2017). Shopping mall image: Systematic review of 40 years of research. *International Review of Retail, Distribution and Consumer Research*, *27*(1), 1–27.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis*. Upper Saddle River, NJ: Prentice-Hall.

Howard, E. (1992). Evaluating the success of out-of-town regional shopping centres. *International Review of Retail, Distribution and Consumer Research*, *2*(1), 59–80.

Huff, D. L. (1964). Defining and Estimating A Trading Area. *Journal of Marketing*, *28*(3), 34–38.

Jackson, P., Perez-del-Aquila, R., Clarke, I., Hallsworth, A., de Kervenoael, R., & Kirkup, M. (2006). Retail restructuring and consumer choice 2. Understanding consumer choice at the household level. *Environment and Planning A*, *38*(1), 47–67.

Koenker, R., & Bassett, G. (1978). Regression Quantiles. *Econometrica*, *46*(1), 33–50.

Koenker, R., & Hallock, K. F. (2001). Quantile Regression. *Journal of Economic Perspectives*, *15*(4), 143–156.

Koenker, R., & Machado, J. A. F. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, *94*(448), 1296–1310.

Kohijoki, A.-M., & Marjanen, H. (2013). The effect of age on shopping orientation. Choice orientation types of the ageing shoppers. *Journal of Retailing and Consumer Services*, *20*(2), 165–172.

Laaksonen, M. (1993). Retail patronage dynamics: learning about daily hopping behavior in contexts of changing retail structures. *Journal of Business Research*, *28*, 1–2,3–174.

Marjanen, H. (1995). Longitudinal study on consumer spatial shopping behaviour with special reference to out-of-town shopping: Experiences from Turku, Finland. *Journal of Retailing and Consumer Services*, *2*(3), 163–174.

Marjanen, H., Engblom, J., & Malmari, M. (2013). Viability of demographic and behavioural independents in quantile regression models in predicting retail patronage. *International Review of Retail, Distribution and Consumer Research*, *23*(5), 523–536.

Marjanen, H., & Malmari, M. (2012). *Ostoksia ja elämyksiä*. [The changing consumer behaviour in the Turku Area: An overview of the findings of the KUMU2011-research project]. Turku: Publications of the Turku School of Economics, Series KR-2:2012.

McGoldrick, P. (2002). *Retail marketing*. London: McGraw-Hill.

Mosteller, F., & Tukey, J. (1977). *Data analysis and regression*. Reading, MA: Addison-Wesley.

Moutinho, L., & Hutcheson, G. D. (2007). Store choice and patronage: A predictive modelling approach. *International Journal of Business Innovation and Research*, *1*(3), 233–252.

Nilsson, E., Gärling, T., Marell, A., & Nordvall, A. C. (2015). Who shops groceries where and how? The relationship between choice of store format and type of grocery shopping. *International Review of Retail, Distribution and Consumer Research*, *25*(1), 1–19.

Ozuduru, B. H., & Guldmann, M.-J. (2013). Retail location and urban resilience: Towards a new framework for retail policy. *S.A.P.I.E.N.S*, *6*(1), 1–13.

Pan, Y., & Zinkhan, G. M. (2006). Determinants of retail patronage: A meta-analytical perspective. *Journal of Retailing*, *82*(3), 229–243.

Parker, C., Ntounis, N., Millington, S., Quin, S., & Rey Castillo-Villar, F. (2017). Improving the vitality and viability of the UK High Street by 2020. Identifying priorities and a framework for action. *Journal of Place Management and Development*, *10*(4), 310–348.

Pleshko, L. P., & Al-Houti, S. (2012). Heavy versus light users: A preliminary study of behaviour patterns in retail services. *Academy of Marketing Studies Journal*, *16*, 61–73.

Ramboll Finland Oy. (2013). *Varsinais-Suomen kaupan palveluverkkoselvitys 2013*. Turku: Regional Council of Southwest Finland.

Rogers, D. (1992). *Retail location analysis in practice*. Harlow: Longman.

Rogers, D. (2005). Developing a location research methodology. *Journal of Targeting, Measurement and Analysis for Marketing*, *13*(3), 201–208.

Roy, A. (1994). Correlates of mall visit frequency. *Journal of Retailing*, *70*(2), 139–161.

Sit, J., Merrilees, B., & Birch, D. (2003). Entertainment-seeking shopping centre patrons: The missing segments. *International Journal of Retail & Distribution Management*, *31*(2), 80–94.

Taloustutkimus. (2018). *Turku Tänään -tutkimus 30 vuotta 9. 10.2018*. Retrieved from https://www.taloustutkimus.fi/ajankohtaista/uutisia/turku-tanaan-tutkimus-30-vuotta.html

Teller, C. (2008). Shopping streets versus shopping malls. Determinants of agglomeration format attractiveness from the consumers' point of view. *International Review of Retail, Distribution & Consumer Research*, *18*(4), 381–403.

Teller, C., Wood, S., & Floh, A. (2016). Adaptive resilience and the competition between retail and service agglomeration formats: An international perspective. *Journal of Marketing Management*, *32*, 17–18,1537–1561.

Wakefield, K. L., & Baker, J. (1998). Excitement at the mall: Determinants and effects on shopping response. *Journal of Retailing*, *74*(4), 515–539.

Wrigley, N., Lambiri, D., Astbury, G., Dolega, L., Hart, C., Reeves, C., & Wood, S. (2015). *British high streets: From crisis to recovery? A comprehensive review of the evidence*.Southampton, GB: Economic and Social Research Council (ESRC).

Appendix

Frequencies in the categories of demographic and behavioural independents.

	2006		2011	
	Frequency	%	Frequency	%
Distance (DIST)	N = 2049		N = 1565	
< 3 km	173	9	133	9
3–7.3 km	886	43	659	42
7.3–35 km	990	48	773	49
Access to car (CAR)	N = 2082		N = 1514	
No car	436	21	308	20
1 car	1098	53	782	52
2 or more cars	548	26	424	28
Household size (HOU)	N = 2070		N = 1550	
1 person	618	30	515	33
2 persons	928	45	667	43
3 or more persons	524	25	368	24
Income[a] (INC)	N = 1982		N = 1511	
INC1	559	28	444	30
INC2	668	34	518	34
INC3	755	38	549	36
Age (AGE)	N = 2082		N = 1566	
18–34	411	20	387	25
35–54	794	38	509	32
55–75	877	42	670	43
Parking (PAR)	N = 2044		N = 1555	
Little	451	22	302	19
Some	298	15	210	14
Much	1295	63	1043	67
Price (PRI)	N = 2062		N = 1563	
Little	241	12	114	7
Some	677	33	377	24
Much	1144	55	1072	69
Quality & Selection_G (Q&S_G)	N = 2044		N = 1560	
Little	181	9	146	9
Some	607	30	587	38
Much	1256	61	827	53
Quality & Selection_N-G (Q&S_N-G)	N = 2032		N = 1545	
Little	74	4	67	4
Some	399	20	363	24
Much	1559	76	1115	72
Dining_N-G (DIN_N-G)	N = 2014		N = 1538	
Little	1072	53	753	49
Some	559	28	487	32
Much	383	19	298	19
Convenience_G (CON_G)	N = 2049		N = 1560	
Little	253	12	44	3
Some	507	25	242	15
Much	1289	63	1274	82

(*Continued*)

(Continued).

	2006		2011	
	Frequency	%	Frequency	%
Location_G (LOC_G)	N = 2038		N = 1556	
Little	258	13	145	9
Some	582	28	462	30
Much	1198	59	949	61
Location_N-G (LOC_N-G)	N = 2024		N = 1544	
Little	161	8	224	15
Some	508	25	482	31
Much	1355	67	838	54
Service_N-G (SER_N-G)	N = 2023		N = 1542	
Little	198	10	196	13
Some	555	27	479	31
Much	1270	63	867	56

[a]In 2006: INC1 = <€1700; INC2 = €1700—€3199, INC3 = ≥€3200
In 2011: INC1 = <€2000, INC2 = €2000—€3999, INC3 = ≥€4000.