Journal of Retailing and Consumer Services 73 (2023) 103355

Contents lists available at ScienceDirect

Journal of Retailing and Consumer Services

journal homepage: www.elsevier.com/locate/jretconser

Evaluating store location and department composition based on spatial heterogeneity in sales potential

Auke Hunneman [a,*], Tammo H.A. Bijmolt [b], J. Paul Elhorst [c]

[a] Department of Marketing, BI Norwegian Business School, 0442 Oslo, UiS Business School, University of Stavanger, Norway
[b] Department of Marketing, Faculty of Economics & Business, University of Groningen, PO Box 800, 9700AV, Groningen, the Netherlands
[c] Department of Economics, Econometrics, and Finance, Faculty of Economics & Business, University of Groningen, PO Box 800, 9700AV, Groningen, the Netherlands

ARTICLE INFO

Keywords:
Retailing
Store location
Assortment composition
Store performance
Spatial dependence

ABSTRACT

In this paper, we extend a retail location evaluation model with the possibility to include the effect of department size adaptation at the store level. We relate department-level store sales to a store's competitive and demographic environment, thereby providing richer insights into the drivers of department sales than a model of just aggregate sales. Further, we accommodate heterogeneity in consumer characteristics over space by using zip code level data and unobserved spatial effects in department sales by including spatially autocorrelated error terms.

Using spatial panel data for 30 clothing stores belonging to one Dutch retail chain, we demonstrate how to use the modeling approach to analyze and predict sales performance of new and existing stores. We show that the predictive performance of our model is superior to that of a benchmark model that does not include spatial autocorrelation.

1. Introduction

Both consumers and retailers differ in their locations (Ma et al., 2011). Since consumers incur travel expenses, certain store locations are more attractive to them than others are (Chintagunta et al., 2012; Inman et al., 2004). In addition, they differ in their preferences for certain products, which determine the relative attractiveness of a store's department composition vis-à-vis that of competitors (Campo et al., 2000). From a firm perspective, location and relative department size decisions affect the degree of spatial competition. Retail firms can reduce the impact of competition by differentiating themselves either through locating a store farther away from competitors and/or by changing their marketing mix (Thomadsen, 2007). However, sometimes agglomerations of stores can attract customers from further afield who want to buy or compare several products during the same trip (Oppenwal and Holyoake, 2004). Thus, despite the growth of e-commerce in the last decades, store location and department composition remain important drivers of store choice, and key elements of a retailer's marketing mix (Blut et al., 2018; Rooderkerk et al., 2013).

The existing literature suggests that department composition and location decisions should be considered together (e.g., Hwang et al., 2010). Consumer preferences vary geographically and retailers who can adjust their relative department sizes to local preferences are able to increase their performance through store patronage decisions, sales, and gross margins. While other studies have considered location choice in isolation or in combination with other marketing mix variables (e.g., Albuquerque and Bronnenberg, 2012), we combine the evaluation of store location and department composition. For this purpose, we explore how store sales and relative department sizes are influenced by the demographic and competitive environment of a specific store.

In this paper, we extend a retail location evaluation model with the possibility to include the effect of department size adaptation at the store level. In particular, we allow for heterogeneity in consumer locations and demographics within the store's trade area rather than assuming one customer profile per store. We relate department-level store sales to a store's competitive and demographic environment, thereby providing richer insights into the drivers of department sales than a model of just aggregate sales. We also explicitly account for heterogeneity in consumer characteristics across space by using zip code level data and unobserved spatial effects in department sales by including spatially autocorrelated error terms. Finally, we not only model the impact of department composition on sales, but also its feedback effect (endogeneity), because more space may be allocated to well-performing departments.

* Corresponding author.
E-mail addresses: auke.hunneman@bi.no (A. Hunneman), t.h.a.bijmolt@rug.nl (T.H.A. Bijmolt), j.p.elhorst@rug.nl (J.P. Elhorst).

https://doi.org/10.1016/j.jretconser.2023.103355
Received 13 September 2022; Received in revised form 31 March 2023; Accepted 2 April 2023
Available online 7 April 2023

We structure this article as follows: We introduce our models to explain total chain sales (section 2.1), department sales shares (2.2), and the relative sizes of each department (2.3), and briefly discuss model estimation and prediction (2.4). Next, we apply our modeling framework in an empirical setting (section 3), the results of which we discuss in section 4. Sections 5 evaluates the performance of new store locations. It compares the predictive performance of the proposed model including spatial autocorrelation, an extension that received relatively little attention up to now, against a benchmark model excluding it. Finally, we present conclusions, managerial implications, and directions for future research.

2. Model specification

We define total chain sales as the sum of sales over all departments and zip codes in the area in which the chain operates, equivalent to the product of a department's share of sales and the total amount of sales generated by the chain in that zip code:

$$S_t = \sum_{m=1}^{M}\sum_{j=1}^{J} S_{mjt} = \sum_{m=1}^{M}\sum_{j=1}^{J} DSS_{mjt} \times S_{jt}, \tag{1}$$

where S_t is chain level sales at time t ($t = 1, \cdots, T$ and T is the number of time periods), S_{mjt} refers to sales of department m ($m = 1, \cdots, M$ and M is the number of departments) in zip code j ($j = 1, \cdots, J$ and J is the number of zip codes) at time t, DSS_{mjt} is the sales share of department m in zip code j at time t, and S_{jt} represents the chain-level sales in zip code j at time.

Since customers signing up for loyalty programs provide the retailer with their addresses, their purchases are registered. Furthermore, since these customers tend to be responsible for a large proportion of total sales (Singh et al., 2006; Van Heerde and Bijmolt, 2005), we focus on their purchases and use this information to determine how department-level store sales are distributed geographically.

We develop models for both variables on the right-hand side of Equation (1), DSS_{mjt} and S_{jt}, to capture the different ways in which store location and department composition affect store sales. We follow the extant literature on spatial interaction modelling (SIM) and retail forecasting (Bonfrer et al., 2022; Campo and Gijsbrechts, 2004; Campo et al., 2000; Fildes et al., 2022; Newing et al., 2015; Ye et al., 2021) by using both store and trade area characteristics for explaining store performance. The trade area characteristics can be split up further into variables related to distance, i.e. consumers exhibit a greater likelihood to shop at stores nearby, the local population (e.g., the number of households), and competition (e.g., the number of competitors for each department). This literature also shows that similar variables can explain aggregate sales, as well as the performance at the product category level. If store departments differ in their attractiveness locally, we expect to find parameters that vary across the equations that explain each department's sales share.

For example, whereas the effect of an increase in consumers' local buying power might lead to higher store sales but not necessarily to changes in departments' sales shares, location characteristics, such as the number of children living in a particular area, might differentially affect the attractiveness and sales shares of individual departments (e.g., children's, women's clothes). Overall, a store's total assortment drives its attractiveness to certain consumer groups and thus store choice, generally leading to higher overall sales (Briesch et al., 2009; Chernev and Hamilton, 2009), while changes in its composition may lead to higher sales shares for departments that constitute a larger proportion of a store's assortment.

We recognize that if retailers allocate more floor space to departments that perform well in a particular store, the reverse effect may also emerge. A department's (past) sales may determine its (relative) size in the store, because of which department sizes should be considered endogenous. We therefore also develop a model to explain the (relative) amounts of floor space attributed to each department. In total, we model three variables: chain-level sales and departments' sales shares at the zip code level, and departments' relative floor space sizes at the store level.

2.1. Chain-level sales

To account for non-negativity and zero observations when analyzing the amount of sales generated per zip code, we adopt a Tobit model:

$$S_{jt}^{*} = \alpha_0 + \sum_{q=1}^{Q}\alpha_q X_{jtq} + \sum_{\ell=1}^{L}\beta_\ell Z_{jt\ell} + \varepsilon_{jt} \tag{2a}$$

and

$$S_{jt} = \begin{cases} S_{jt}^{*} \text{ if } S_{jt}^{*} > 0 \\ 0 \text{ if } S_{jt}^{*} \leq 0 \end{cases} \tag{2b}$$

where S_{jt}^{*} is a latent variable measuring the amount of sales generated in a particular zip code. The set of explanatory variables includes variables observed at the store (X) and zip code (Z) levels. The store-specific explanatory variables are further measured such that they refer to the nearest store.

We extend the Tobit model, and the attraction models in the next two subsections, with spatially autocorrelated error terms (Bradlow et al., 2005; Bronnenberg and Mahajan 2001; Elhorst, 2017), because zip codes in close proximity tend to share unobserved characteristics (e.g., history, resources, infrastructure), and consumer spending levels in neighboring zip codes cannot be considered fully independent. This issue did not receive any attention in the SIM of Newing et al. (2015) and the review on retail forecasting of Fildes et al. (2022). By stacking observations ($j = 1, \ldots, J$) in a vector for each cross-section of zip codes at time t, we account for spatial error autocorrelation by

$$\varepsilon_t = \lambda W \varepsilon_t + \varsigma_t \tag{3}$$

where $E(\varsigma_t) = 0$, $r(\varsigma_t) = \sigma^2 I_J$, W is a row-standardized first-order binary contiguity matrix of size ($J \times J$) describing the spatial arrangement of zip codes, and λ is the spatial autocorrelation coefficient.

2.2. Department sales shares

We use an attraction model (Cooper and Nakanishi, 1988) to explain a department's sales share in a particular zip code, based on its relative size, other store attributes, and competitor and consumer characteristics observed at the zip code level:

$$DSS_{mjt} = \frac{A_{mjt}}{\sum_{c=1}^{M} A_{cjt}} \tag{4a}$$

where

$$A_{mjt} = exp\left(\beta_{1m} + \varepsilon_{mjt}\right) SS_{mjt}^{\beta_{2m}} \prod_{k=1}^{K} exp\left(\gamma_{km} X_{tk}\right) \prod_{n=1}^{N} exp\left(\lambda_{nm} Z_{jtn}\right) \tag{4b}$$

A_{mjt} is the attraction of department m in zip code j at time t, and SS_{mjt} is the fraction of store space devoted to department m in the store closest to zip code j at time t. The set of explanatory variables includes variables observed at the store (X) and at the zip code level (Z). Campo et al. (2000) and Campo and Gijsbrechts (2004) have used attraction models for similar purposes.

2.3. Department sizes

Over time, a retailer may decide to allocate more floor space to departments that are selling well in a particular store (Van Dijk et al., 2004; Van Nierop et al., 2008). If the amount of space attributed to a department becomes a function of its past performance and this endogeneity is ignored, the impact of changes in a department's (relative) size on its share of sales might be overestimated. To correct for this, we specify a model for assortment composition in which department sizes are a function of store, competitor, and (aggregated) consumer characteristics. We again adopt an attraction model specification because relative department sizes also satisfy the logical consistency requirements of this model type. Hence, we model SS_{mit} using[1]

$$SS_{mit} = \frac{Att_{mit}}{\sum_{c=1}^{M} Att_{cit}} \tag{5a}$$

where

$$Att_{mit} = exp(\varrho_m + \upsilon_m) \prod_{g=1}^{G} exp\left(\delta_{gm} X_{itg}\right) \prod_{r=1}^{R} exp(\varphi_{rm} \overline{Z}_{itr}) \tag{5b}$$

and the zip code–level variables $\overline{Z}_{itr}$ are averages over all zip codes for which store i is the nearest store. In this model, the store-specific explanatory variables are augmented with a variable measuring the one-period lag of department m's share of sales, while spatial autocorrelation is modelled between the trade areas of stores rather than zip codes.

2.4. Model estimation and prediction

Estimating the parameters of attraction models is not straightforward (Fok et al., 2002), the more so since we also extend these models to include spatially correlated error terms. In web appendix A, we detail the applied estimation procedure. Another important element of the proposed modeling approach involves predicting the sales impact of changes in a store's retail environment. In web appendix B, we elaborate on the methodology to obtain predictions based on our models.

3. Data

We use data from a Dutch retail chain with 30 clothing stores to apply our model to an empirical setting. The chain's positioning is targeted at middle-class families, as reflected in the stores' average price levels and medium-quality assortments. All stores offer clothes for men, women, and children, but the relative amount of floor space devoted to each department differs for each store. Store size ranges from approximately 500 m^2–2500 m^2, of which an average of 44 and 21 percent contains women's and children's clothes, respectively.

From the chain's customer database, we obtained yearly data on department-level sales for all zip codes in the Netherlands in four successive years (2003–2006[2]); we use the first three years for estimation and the last year for validation. These data are supplemented with survey data on the retail environment in which each store operates and commercially available geodemographic information. It is important to note that the data is relatively unique for our setting (e.g., the apparel market in our case) and might thus differ for other settings. The explanatory variables are selected based on their ability to predict store sales, the extant literature (see for example, Bonfrer et al., 2022; Campo and Gijsbrechts, 2004; Campo et al., 2000; Fildes et al., 2022; Newing et al., 2015; Ye et al., 2021), and data availability. The type of data that we use are readily available from census statistics or commercial parties that offer similar data. This makes our modelling approach easily applicable in practice.

We identify the number of competitors for each store using information obtained from a survey among store managers. Other store attributes include store size (in 10,000 m^2), the relative sizes of the various departments, the number of months of the year that a store is open (in case it was opened in a year that is part of the sample period), and store age (1932 = 0).

We use a variety of socio-demographic variables observed at the zip code level to evaluate the impact of consumer characteristics on overall and department-level sales. These variables can influence store performance in several ways. First, variables such as the number of households living in a particular zip code and their socio-economic status determine local buying power and affect overall spending levels. Second, other consumer variables, such as the percentage of households with children, may differentially affect the sales level of each department.

To check for multicollinearity, we examine the correlations between the explanatory variables. These correlations are mostly small; the highest correlation among two explanatory variables is 0.73 across all equations. The VIF for the explanatory variables ranges from 1.36 to 4.87 and the mean VIF equals 2.65. Hence, we conclude that multicollinearity is not an issue.

4. Estimation results

We use the store level model to predict the relative amount of floor space allocated to a department in every store. Since relative department sizes depend on a department's past sales, we use the predicted department sizes as explanatory variables in the total sales and department sales shares equations observed at the zip code level. Table 1 reports the estimation results of the store level model and Table 2 of the zip code level models. The number of observations for the store level models are 84 observations (three years times 28 stores). For the zip code level model, we have three years of data of 4008 zip codes, which amounts to 12,024 observations. However, since the trade areas of the stores do not cover the whole country, the number of observations used in the first two columns of Table 2 is smaller than 12,024.

4.1. Store level department sizes

Table 1 provides empirical evidence in favor of the effect that store managers tend to allocate larger (smaller) amounts of store space to better (worse) performing departments; the coefficients of lagged department sales on the amount of floor space allocated to this department are positive and significant. This finding is consistent with previous findings of Van Dijk et al. (2004) and Van Nierop et al. (2008).

Larger and newer stores are likely to have a larger children's department. The size of a store and the year in which a store is established maintain a negative relationship with the relative sizes of the men's and women's departments. By contrast, store managers tend to enlarge the men's department if the proportion of households with low and high socio-economic status grows. The competitor characteristics and the spatial autocorrelation coefficients for both the women's and the men's department turn out to be insignificant, perhaps due to of the limited number of observations at this level. Note that the coefficients of the number of competitors for children's department in columns (1) and (2) of Table 1, as well as Table 2, are the same due to estimating the model in relative sizes (see also web appendix A). Unobserved factors captured by the error terms might be uncorrelated across space as trade areas of stores are considerably larger than zip codes.

4.2. Total sales

The total sales model results in the last column of Table 2 indicate that the floor space allocated to each of the three store's departments

[1] The notation for department sizes in Equation (5) differs from that in Equation (4). SS_{mit} in Equation (5) refers to the relative size of department m in a particular store i at time t, whereas SS_{mjt} in Equation (4) measures the relative size of department m in the store closest to zip code j at time t.

[2] The data originates from a project conducted a few years ago, in close collaboration with the retailer. As the contribution of our paper is in the modelling approach and not in the specific empirical findings, this is not a major problem.

Table 1
Parameter estimates of attraction model explaining relative department sizes.

Explanatory Variable	Dependent Variable			
	Women's Department Size		Men's Department Size	
	Coeff	t-value	Coeff	t-value
Constant	3.837	2.92	2.151	2.20*
Store characteristics				
Size (in 10,000 m^2)	−7.429	−4.31**	−2.207	−1.78
Lagged sales female department	0.961	3.83**		
Lagged sales male department			0.960	5.41**
Lagged sales children's department	−0.745	−1.47	−0.745	−1.47
Period open (months per year)	0.753	−1.78	−0.492	−1.50
Year of establishment (/100)	−0.737	−3.28**	−0.356	−2.24*
Competitor characteristics				
# competitors female department (/100)	−0.295	−1.18		
# competitors male department (/100)			0.563	1.63
# competitors children's department (/100)	−0.272	−0.25	−0.272	−0.25
Consumer characteristics				
% households with high SES	−2.074	−1.43	2.960	2.70**
% households with low SES	−0.059	−0.04	3.577	2.99**
% of foreigners	2.752	2.81**	1.265	1.59
% couples	−5.243	−2.95**	−6.214	−4.39**
% households with children	−0.925	−0.70	0.182	0.18
D2004	−0.020	−0.49	0.007	0.15
D2005	−0.037	−0.82	−0.023	−0.47
Spatial autocorrelation (λ)	−0.530	0.03	0.055	0.00
R^2	0.73		0.60	
# observations	84		84	

Notes: **$p < 0.01$ * $p < 0.05$.

does not significantly affect the overall sales level, which is a counter-intuitive finding. Increasing the absolute size of a department apparently does not lead to higher store-level sales, possibly because this increase comes at the expense of other departments. The first two columns of Table 2 show that the relative size of a department positively correlates with its total sales share, supporting this interpretation. This finding is also consistent with research by Campo et al. (2000), who show that the space share of a product category affects both its relative attractiveness and its total sales share.

The last column of Table 2 shows that the overall sales level at a particular zip code is negatively related to the distance to the nearest store. This finding is consistent with previous findings that consumers perceive disutility from traveling, and therefore are more likely to visit stores closer to their residence (Bawa and Ghosh, 1999; Bhatnagar and Ratchford, 2004). In addition to distance, we find a positive and significant spatial autocorrelation coefficient, which indicates that zip codes in proximity have similar sales levels due to shared unobserved characteristics. A similar result is found for the two department sales shares in the next section. In section 5, we demonstrate that utilizing this information improves the prediction performance of the model significantly.

Regarding the location characteristics, we further find a significant positive effect of the number of households. Regions with large populations constitute potentially larger markets, causing retail stores to generate more sales from these regions (Kumar and Karande, 2000; Reinartz and Kumar 1999; Zhu et al., 2009). Moreover, households with children and couples spend significantly more on clothes than do single-person households at this retailer, likely because larger families have more diverse needs and buy a wider variety of products, which on average produces higher sales for these consumer groups (Bawa and Ghosh, 1999).

We finally find a positive and significant effect of the number of competitors on the chain's overall sales level, which means that more rival stores enhance target store performance. Stores in proximity may earn more business than those located far apart, because consumers visiting multiple stores have a lower risk of product unavailability (González-Benito and González-Benito, 2005). Moreover, since shopping for clothes is sometimes regarded as a recreational activity, clothing stores may benefit from the presence of competitors, who offer the promise of comparison and fun shopping (Dholakia, 1999).

4.3. Department shares

The second and third columns of Table 2 contain the parameter estimates for the model that explains sales shares in the women's and men's departments, using the children's department as reference category. The results show that more floor space allocated to a particular department increases the sales share for this department relative to the children's department. This finding confirms the results of Campo et al. (2000), Desmet and Renaudin (1998), and Van Nierop et al. (2008). One explanation is that products in departments with a large share of the store's total floor space receive more attention from consumers and are more likely to appear in their shopping baskets (Desmet and Renaudin, 1998). If more floor space is allocated to a particular department, the retailer also can display more items. This gives consumers more products to choose from, so that they are more likely to find what they want, which increases sales in this department (Hoch et al., 1995). Similarly, more space allows the retailer to hold extra inventory and lower the risk of out-of-stocks, which may have a positive effect on sales (Campo et al., 2003; Desmet and Renaudin, 1998).

The sales shares of the men's and women's department are positively related to the number of department-specific competitors. Holding total sales constant, the men's and women's departments thus benefit more from the presence of competitors than does the children's department. Travel distance also positively affects the sales shares of men's and women's departments. The combination of these findings suggests that men's and women's departments benefit most from the spatial concentration of apparel stores.

5. Store location evaluation and prediction

One of the most important elements of store location evaluation involves predicting sales. According to Fildes et al. (2022, p.1288), the forecasting of store sales can be divided into two categories. First, the forecasting performance of existing store sales, in our case at the zip code level and especially whether these predictions are better than a benchmark model that does not include spatial autocorrelation. Second, the forecasting models can be used for new store potential sales for site selection, in our case for the two newly opened stores in 2006.

The parameter estimates for the models without spatial autocorrelation are reported in Appendix C, Table C2, for reasons of comparison. Table 3 shows the R-squares of predictions of total sales based on the Tobit model with and without spatial autocorrelation and the relative contributions of the store, competitor and consumer characteristics, for both the estimation sample (2003–2005) and the holdout sample (2006). We first observe, as expected, that, for each model, the R-squares for the estimation period are higher than those for the holdout sample. More importantly, we also find that models that include spatial autocorrelation have a better R square than those that do not, both for the estimation and the holdout sample. In other words, the results show that the standard predictor for the Tobit model (see Equation B.4 in Appendix B) can be improved by borrowing information about unobserved characteristics covered by the residuals of neighboring zip codes, as set

Table 2
Parameter estimates of models explaining department sales shares and total sales.

Explanatory Variable	Dependent Variable					
	Women's Department Sales Share		Men's Department Sales Share		Total Sales	
	Coeff	t-value	Coeff	t-value	Coeff	t-value
Constant	1.386	4.11**	1.617	4.73**	−8.191	−1.95
Store characteristics						
m^2 female department	0.274	3.02**			16.147	0.41
m^2 male department			0.501	6.29**	−19.330	−0.45
m^2 children's department	−0.276	−3.98**	−0.276	−3.98**	15.046	0.33
Period open (months per year)					6.076	1.86
Year of establishment (/100)					1.096	0.40
Competitor characteristics						
# competitors female department (/100)	0.829	2.56*				
# competitors male department (/100)			2.414	4.50**		
# competitors children's department (/100)	−5.522	−6.01**	−5.522	−6.01**		
# competitors in total (/100)					7.343	4.03**
Consumer characteristics						
Distance to nearest store (miles)	0.018	13.27**	0.012	8.24**	−0.264	−17.90**
# households (/1000)					2.989	27.53**
% households with high SES	−0.199	−1.20	−0.369	−2.05*	−1.178	−1.06
% households with low SES	0.884	5.55**	0.671	3.91**	2.604	2.33*
% of foreigners	0.315	1.56	1.222	5.60**	−12.357	−7.43**
% couples	1.357	4.51**	1.095	3.33**	8.585	4.65**
% households with children	−2.057	−10.84**	−1.614	−7.86**	3.462	2.54*
D2004	0.056	1.14	0.131	2.43*	−0.341	−0.65
D2005	0.049	1.00	0.292	5.41**	−0.418	−0.80
Spatial autocorrelation (λ)	0.218	9.24**	0.243	10.64**	0.840	101.16**
R^2	0.06		0.08		0.47	
# observations	9798		9798		12,024	

Notes: **p < 0.01 * p < 0.05.

Table 3
R squares of predictions by different models and for different samples.

	R square in-sample predictions	R square out-of-sample predictions
Tobit model with spatial effects		
Tobit predictor without residuals	0.309	0.274
Tobit predictor with residuals	0.473	0.429
Model including only store characteristics[a]	0.412	0.373
Model including only competitor characteristics[a]	0.411	0.372
Model including only consumer characteristics[a]	0.472	0.428
Tobit model without spatial effects		
Tobit predictor	0.396	0.342

[a] Characteristics added to the Tobit model with an intercept, D2004, D2005, and spatial effects.

out in Kelejian and Prucha (2007) and Goulard et al. (2017). As shown in Appendix B, these residuals depend on the size of the spatial autocorrelation coefficient, which is quite substantial for the total sales model ($\lambda = 0.84; p < 0.01$). More specifically, for the estimation period, the R-square of predictions if spatial autocorrelation is included is 0.473, while that for the standard Tobit model is 0.396. If we leave out the correction term for the spatially autocorrelated error terms from the predictions, the R-square drops to 0.309. Hence, this value is lower than the R-square for the standard Tobit model without spatial autocorrelation because the parameter estimates and thus the sales predictions change if we ignore spatial autocorrelation while present. We observe a similar pattern for the out-of-sample predictions. The R-square for the model with spatial autocorrelation is 0.429, which is higher than that for the standard Tobit model (0.342). This value further drops to 0.274 if we do not add the residuals observed for neighboring zip codes.

Table 3 further shows that the relative contribution of the consumer characteristics to the R-square dominates the contributions of the store and competitor characteristics, in line with the significance levels found in the last column of Table 2. This finding confirms the view that it is beneficial to account for heterogeneity in consumer locations and demographics within the trade area of the store rather than relying on one customer profile per store.

To examine to which extent the proposed model predicts sales correctly for new stores, we use out-of-sample data in 2006. This year marks the opening of two new stores in the mid-western part of the Netherlands (in Fig. 2, numbers 1 and 2), which have similar characteristics as the 28 stores in the estimation sample (2003–2005).

The left panels of Figs. 1 and 2 present the predicted sales levels at the zip code level for the hypothetical situation that the two new stores would not have been present in 2006. We obtained these predictions using the methodology described in Web appendix B and the coefficient estimates reported in Table 2. In this situation, the values for the explanatory variables in the model measure the travel distances to and characteristics of the nearest stores, assuming that only the existing 28 stores are present and not the two new stores. The right panels of Figs. 1 and 2 repeat this analysis but then after changing the explanatory variables in the model so that they reflect the situation after the new store openings. For this purpose, the set of explanatory variables observed in 2006 has been modified such that they represent the travel distances to

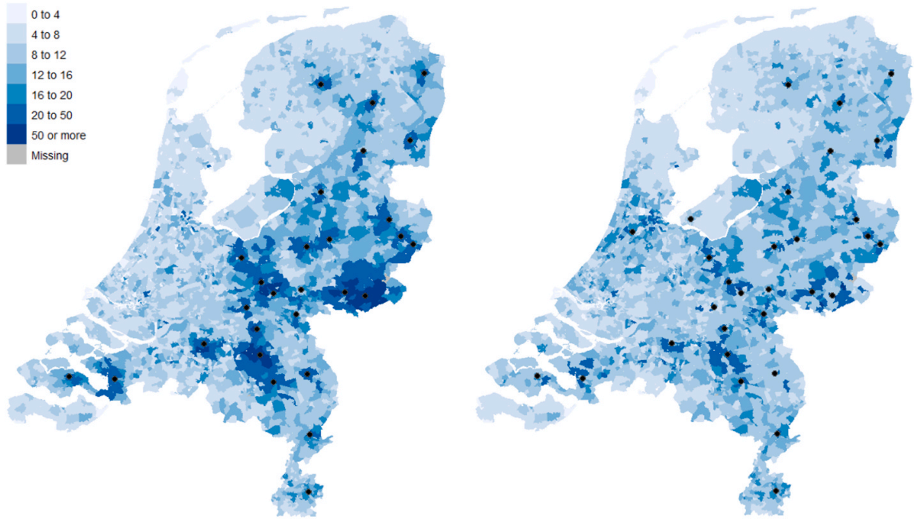

Fig. 1. Predicted sales levels for each zip code in the Netherlands in the year 2006.

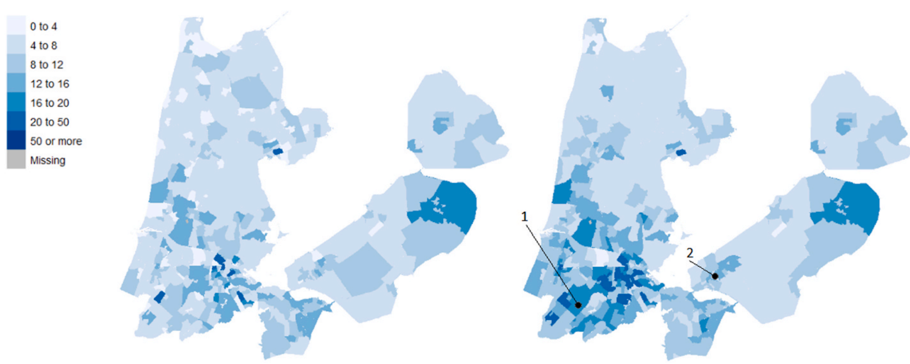

Fig. 2. Predicted sales levels for each zip code in the two provinces (Noord-Holland and Flevoland) in which the new stores opened in the year 2006.

and characteristics of the nearest store to each zip code including the two new stores. To evaluate and compare the predicted sales levels with the observed sales figures of the new stores, we subtract the sales amounts generated in each zip code before from that after the opening of the new store and sum the sales for all zip codes that belong to a store's trade area.

To determine the trade area of these stores, we first calculated the average size of the trade area size for all stores in 2005. We did so by sorting all zip codes in descending order of travel distance to a store and then select, for each store, the first sorted zip codes responsible for 85% of the sales. The perimeter of the trade area is defined as the travel distance to the last zip code assigned to a store. This travel distance is calculated as the fastest distance in miles a car can travel from (the centroid of) a four-digit zip code to the store under consideration. We calculated this trade area perimeter for all 28 stores present in 2005 and took the average of these values, which equals 16.44 miles. We subsequently used this value to predict the sales of the new stores by taking the sum of the predicted sales levels over all zip codes within 16.44 miles from each new store location. The predicted and realized sales levels are: €1,018,711 versus €1,057,353 for store 1, and €285,197 versus €105,344 for store 2. Hence, the model predictions are fairly accurate compared to the general range of sales levels across all stores. In addition, the prediction is more accurate for the relatively successful store than for the store with a relative low sales level. Overall, these model predictions for total sales approximate the observed values reasonably well, indicating that the model is useful for store location evaluations.

6. Conclusion and discussion

We introduce a model for store location evaluations that acknowledges the effects of location characteristics on the department sales of a store and that contributes to prior literature in several ways. We allow for more heterogeneity in consumer characteristics than existing models in that we use consumer data observed at the zip code level rather than aggregated socio-demographic profiles for each store (Kumar and Karande, 2000). On top of that, we account for spatially correlated error terms that may result from unobserved imitating behavior by consumers (Choi et al., 2010), retailers (Bronnenberg and Mahajan 2001), or other variables that cause spatial dependence in department sales levels across zip codes.

We test the proposed model using data from a Dutch clothing chain operating 30 stores, each offering a retail assortment to middle-class families. The empirical study confirms previous findings from Campo et al. (2000): Location variables affect each department's sales shares differently. Travel distance to the store, for example, affects the sales shares of the men's and women's department more strongly than that of the children's department, whereas in areas with more families with children, the sales share of the children's department is greater. Total sales levels are higher in areas closer to the store, where there is more intense competition and greater market potential (i.e., number of households). We further find evidence that retailers decide about the amount of floor space devoted to each department, based on each department's past performance, indicating that department sizes should not be treated as exogenous. On the other hand, such changes in floor space devoted to each department primarily appears to affect the sales share coming from that department rather than total sales, indicating a potential shortcoming of this study. But above all, we find that spatial error correlation is significant in those equations estimated based on zip code level data.

Although this study increases our understanding of the impact of spatial autocorrelation, one limitation is that we consider only one chain of stores; the findings are therefore peculiar to the positioning of this chain and difficult to generalize. On the other hand, we have tried to present the model in such a general form that it can also be applied to other retailers or other settings that require evaluations of their locations and store attributes.

Despite this limitation, we believe that the proposed model is valuable to retailers that want to open new chain stores and tailor their assortments to local conditions. As we have shown in our empirical study, this model effectively predicts potential sales by new store locations and the sales impacts of changes in department sizes, which makes it a useful tool to support these decisions.

The left panel shows the sales distribution if there are no new stores and the right panel when including the two new stores. Sales figures are measured in 1000 euros.

Statements and declarations

• The authors did not receive financial support from any organization for the submitted work.

• The authors have no competing interests to declare that are relevant to the content of this article.

Declarations of interest

None.

Data availability

The data that has been used is confidential.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jretconser.2023.103355.

References

Albuquerque, A., Bronnenberg, B.J., 2012. Measuring the impact of negative demand shocks on car dealer networks. Market. Sci. 31, 4–23.
Bawa, K., Ghosh, A., 1999. A model of household grocery shopping behavior. Market. Lett. 10, 149–160.
Bhatnagar, A., Ratchford, B.T., 2004. A model of retail format competition for non-durable goods. Int. J. Res. Market. 21, 39–59.
Blut, M., Teller, C., Floh, A., 2018. Testing retail marketing-mix effects on patronage: a meta-analysis. J. Retailing 94, 113–135.
Bonfrer, A., Chintagunta, P.K., Dhar, S., 2022. Retail store formats, competition and shopper behavior: a systematic review. J. Retailing 98, 71–91.
Bradlow, E.T., Bronnenberg, B.J., Russell, G.J., Arora, N., Bell, D.R., Duvvuri, S.D., Ter Hofstede, F., Sismeiro, C., Thomadsen, R., Yang, S., 2005. Spatial models in marketing. Market. Lett. 16, 267–278.
Briesch, R.A., Chintagunta, P.K., Fox, E.J., 2009. How does assortment affect grocery store choice? J. Market. Res. 46, 176–189.
Bronnenberg, B.J., Mahajan, V., 2001. Unobserved retailer behavior in multimarket data: joint spatial dependence in market shares and promotion variables. Market. Sci. 20, 284–299.
Campo, K., Gijsbrechts, E., 2004. Should retailers adjust their micro-marketing strategies to type of outlet? An application to location-based store space allocation in limited and full-service grocery stores. J. Retailing Consum. Serv. 11, 369–383.
Campo, K., Gijsbrechts, E., Goossens, T., Verhetsel, A., 2000. The impact of location factors on the attractiveness and optimal space shares of product categories. Int. J. Res. Market. 17, 255–279.
Campo, K., Gijsbrechts, E., Nisol, P., 2003. The impact of retailer stockouts on whether, how much, and what to buy. Int. J. Res. Market. 20, 273–286.
Chernev, A., Hamilton, R., 2009. Assortment size and option attractiveness in consumer choice among retailers. J. Market. Res. 46, 410–420.
Chintagunta, P.K., Chu, J., Cebollada, J., 2012. Quantifying transaction costs in online/off-line grocery channel choice. Market. Sci. 31, 96–114.
Choi, J., Hui, S.K., Bell, D.R., 2010. Spatiotemporal analysis of imitation behavior across new buyers at an online grocery retailer. J. Market. Res. 47, 75–89.
Cooper, L.G., Nakanishi, M., 1988. Market-Share Analysis: Evaluating Competitive Marketing Effectiveness. Kluwer Academic Publishers, Boston.
Desmet, P., Renaudin, V., 1998. Estimation of product category sales responsiveness to allocated shelf space. Int. J. Res. Market. 15, 443–457.
Dholakia, R.R., 1999. Going shopping: key determinants of shopping behaviors and motivations. Int. J. Retail Distrib. Manag. 27, 154–165.
Elhorst, J.P., 2017. Spatial models. In: Leeflang, P.S.H., Wieringa, J.E., Bijmolt, T.H.A., Pauwels, K.H. (Eds.), Advanced Methods for Modeling Markets. Springer, Cham Switzerland, pp. 173–202.
Fildes, R., Ma, S., Kolassa, S., 2022. Retail forecasting: research and practice. Int. J. Forecast. 38, 1283–1318.
Fok, D., Franses, P.H., Paap, R., 2002. Econometric analysis of the market share attraction model. In: Franses, P.H., Montgomery, A.L. (Eds.), Econometric Models in Marketing. Elsevier, Amsterdam, pp. 223–256.
González-Benito, O., González-Benito, J., 2005. The role of geodemographic segmentation in retail location strategy. Int. J. Mark. Res. 47, 295–316.
Goulard, M., Thibault, L., Thomas-Agnan, C., 2017. About predictions in spatial autoregressive models: optimal and almost optimal strategies. Spatial Econ. Anal. 12, 304–325.
Hoch, S.J., Kim, B.D., Montgomery, A.L., Rossi, P.E., 1995. Determinants of store-level price elasticity. J. Market. Res. 32, 17–29.
Hwang, M., Bronnenberg, B.J., Thomadsen, R., 2010. An empirical analysis of assortment similarities across U.S. supermarkets. Market. Sci. 29, 858–879.
Inman, J.J., Shankar, V., Ferraro, R., 2004. The roles of channel-category associations and geodemographics in channel patronage. J. Market. 68, 51–71.
Kelejian, H.H., Prucha, I.R., 2007. The relative efficiencies of various predictors in spatial econometric models containing spatial lags. Reg. Sci. Urban Econ. 37, 363–374.
Kumar, V., Karande, K., 2000. The effect of retail store environment on retailer performance. J. Bus. Res. 49, 167–181.
Ma, Y., Ailawadi, K.L., Gauri, D.K., 2011. An empirical investigation of the impact of gasoline prices on grocery shopping behavior. J. Market. 75, 18–35.
Newing, A., Clarke, G.P., Clarke, M., 2015. Developing and applying a disaggregated retail location model with extended retail demand estimations. Geogr. Anal. 47, 219–239.
Oppenwal, H., Holyoake, B., 2004. Bundling and retail agglomeration effects on shopping behavior. J. Retailing Consum. Serv. 11, 61–74.
Reinartz, W.J., Kumar, V., 1999. Store-, market-, and consumer-characteristics: the drivers of store performance. Market. Lett. 10, 5–23.
Rooderkerk, R.P., Van Heerde, H.J., Bijmolt, T.H.A., 2013. Optimizing retail assortments. Market. Sci. 32, 679–826.
Singh, V.P., Hansen, K.T., Blattberg, R.C., 2006. Market entry and consumer behavior: an investigation of a Wal-Mart supercenter. Market. Sci. 25, 457–476.
Thomadsen, R., 2007. Product positioning and competition: the role of location in the fast food industry. Market. Sci. 26, 792–804.
Van Dijk, A., Van Heerde, H.J., Leeflang, P.S.H., Wittink, D.R., 2004. Similarity-based spatial methods to estimate shelf space elasticities. Quant. Market. Econ. 2, 257–277.
Van Heerde, H.J., Bijmolt, T.H.A., 2005. Decomposing the promotional revenue bump for loyalty program members versus nonmembers. J. Market. Res. 42, 443–457.
Van Nierop, E., Fok, D., Franses, P.H., 2008. Interaction between shelf layout and marketing effectiveness and its impact on optimizing shelf arrangements. Market. Sci. 27, 1065–1082.
Ye, Z., Clarke, G., Newing, A., 2021. Estimating small-area demand of urban tourist for groceries: the case of Greater London. J. Retailing Consum. Serv. 58, 1–17.
Zhu, T., Singh, V., Manuszak, M.D., 2009. Market structure and competition in the retail discount industry. J. Market. Res. 46, 453–466.