LOGI – Scientific Journal on Transport and Logistics
Vol. 16 No. 1 2025 DOI: 10.2478/logi-2025-0010

Shopping Trip Choice Prediction for Assessing Store Relocation: a Joint Data-Driven and Behavioural Modelling Approach

Sonagnon Hounwanou[1], Antonio Comi[2], Jesus Gonzalez-Feliu[3*] and Natacha Gondran[4]

[1]*Institut VEDECOM, 23 bis allée des Marronniers, 78000 Versailles, France; Email: pierrechanel.h@gmail.com*
[2]*University of Rome Tor Vergata, Department of Enterprise Engineering, 00133, Rome, Italy; Email: comi@ing.uniroma2.it*
[3]*Excelia Business School, Center of Research in Innovation and Intelligence in Management, 102 rue de Coureilles, 17000 La Rochelle, France; Email: gonzalezfeliuj@excelia-group.com*
[4]*Mines Saint-Etienne, CNRS, Univ Jean Monnet, Univ Lumière Lyon 2, Univ Lyon 3 Jean Moulin, ENS Lyon, ENTPE, ENSA Lyon, UMR 5600 EVS, Institut Henri Fayol, F - 42023 Saint-Etienne France; Email: gondran@emse.fr*

***Corresponding Author:** Jesus Gonzalez-Feliu

Received: 17 January 2025; Revised: 24 April 2025; Accepted: 15 May 2025; Published: 21 August 2025

Abstract: This paper proposes and tests a methodology to analyse end consumers' choices in terms of shopping destination for a store selling culture products, comparing a city centre location to a peripheral one. The proposed methodology begins with a stated preferences survey and incorporates a conditional tree classification algorithm to pre-select the predictors (attributes), then used to develop a discrete choice model. To validate the methodology, a real-world case study was carried out, including a survey with over one thousand customer responses. The findings reveal noteworthy insights into customer attitudes toward relocation, distinguishing frequent from non-frequent users and examining factors such as travel distance and visit frequency. These results offer valuable guidance for retailers and policy makers in shaping city logistics scenarios, highlighting the potential transformations in urban freight flows driven by changes in retail land use.

Keywords: Shopping trip behaviour, conditional-inference-tree, logit model, destination choice

1. Introduction

The location of urban retailing activities is a major question in urban dynamics that involves various stakeholders and has a direct impact on urban transport economics. Indeed, a retailer's location impacts customer choices, not only in terms of the product offerings but also the ease (or difficulty) of access, which in turn affects freight vehicle flows. The decision to relocate a store is typically driven by market and cost considerations [1]. It represents a trade-off between attracting more customers (and thereby increasing revenues) and reducing operational costs (e.g., restocking expenses or rental fees). However, selecting an optimal location is far from straightforward [2]. Addressing those issues is important not only for retailers but also for public authorities, since they act as organizers of the urban space and can act to provide more sustainable land uses.

In this context, shopping trips, and more specifically end consumers' choices, are closely tied to the generation of purchases and travel activities [3]. Moreover, shopping trip behaviour is extremely related to the consumer's choices in terms of type of retailer and sometimes, quantity and quality of products to buy [4]. Moreover, the spatial patterns of retailing, i.e., its location and context, but also that of the consumer, play a major role on such choices [5].

The solutions to these limits may consist in making the link between customers' spatial behaviour and retail locations, including on one side the ideas and representation that customers have about the places, and on the other side, their mobility behaviours [5]. Although various methods can be used for predicting behaviour [4] (most bases in logit or probit models), in discrete choice methods, a high number of variables could be examined [6], so a pre-selection of variables (or attributes) can lead to more relevant and robust models since it reduces the computation efforts of the calibration process [7]. Since consumers' choices have a direct impact on the frequentation of a retailer [8], addressing them in the context of a retailer relocation has both a research and practice interest.

This paper aims at proposing a data-driven methodology to assess customer preference between two or more locations for a retail store, in a store relocation context, which combines a conditional inference tree algorithm to pre-select decisional variables and a logistics regression analysis. Section 2 presents the methodology, and its different stages and methods. Section 3 shows the case of application (which concerns a store moving from a French city-centre to a shopping mall in the periphery of the same urban area) and the main results of the framework applied to the case. It includes the comparison of the two proposed ranking algorithms, the selection of the chosen one and its combination with the logit, plus a cross-validation process. The results' discussion, including the main applications and evolutions of the proposed framework, are presented in Section 4. Finally, conclusions are drawn in Section 5.

2. Data and Methods

This paper explores potential customer behavioural responses when choosing between different shopping destinations, with a specific focus on the comparison between a city centre location and a peripheral location. The analysis is grounded in shopping trip intentions, closely linked to urban transport considerations. The proposed methodology is outlined in Figure 1.

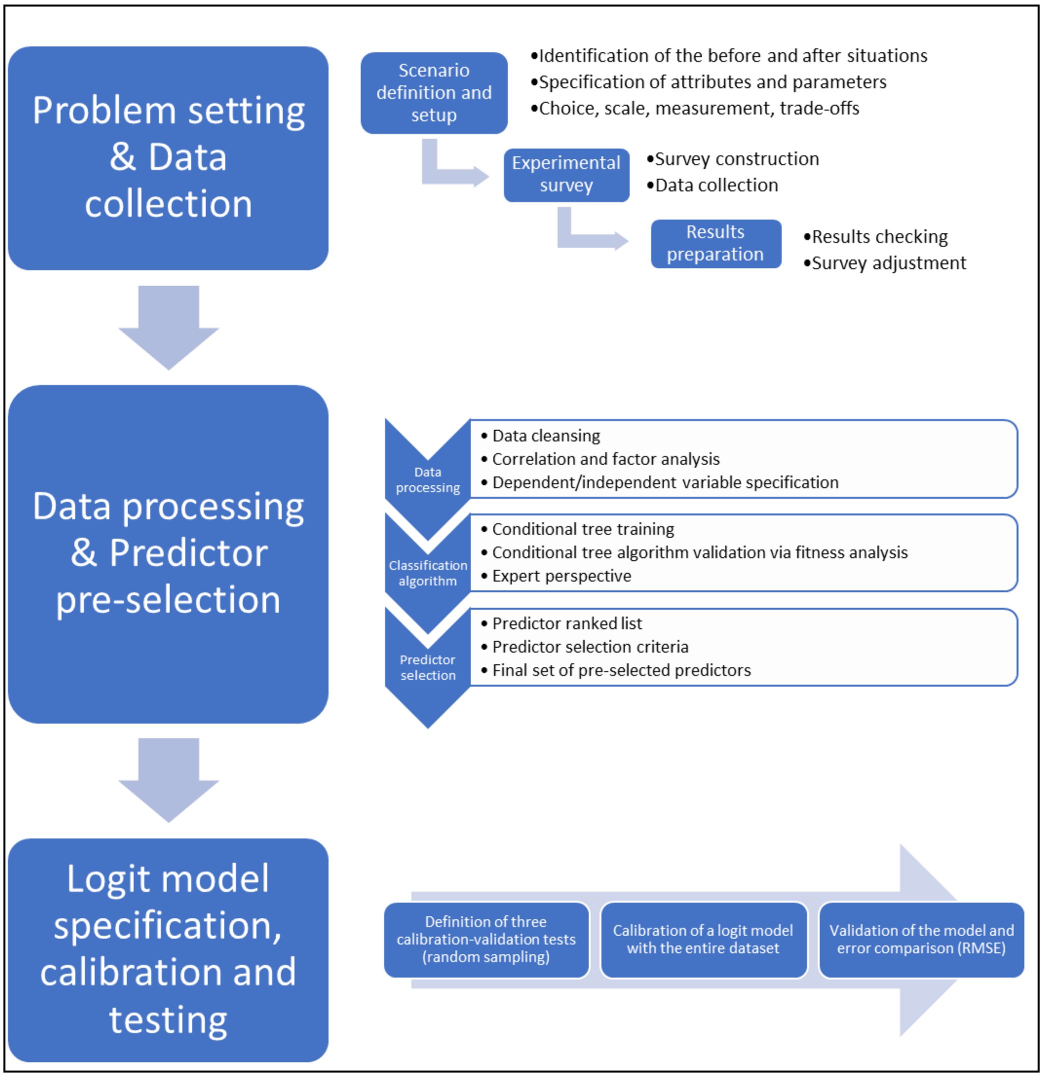

Fig. 1 Flowchart of the proposed methodology. Source: authors

The first phase is that of problem setting and data collection. A first field observation led to define the before-after survey and define the variables to collect, the questionnaires to develop and define the different scales, choices, measurement and trade-offs [9]. A stated preferences survey was developed to assess the after situation, including a set of questions to define the current (before) situation. The details on the survey can be seen in [5].

The second phase is that of data processing and predictor pre-selection. To do that, various methods can be used, as studied in [5], showing the potential of conditional inference trees. We propose in this work a ranking process based on conditional inference trees via a multiple random subsampling of the original data frame based on [10]'s framework. The method is developed and tested on the data collected through a stated preferences (SP) survey [11]. Moreover, this method seems more relevant when we have to deal with a large number of variables/attributes coming from land-use or socio-economic dataset as well as from the large number of sensors nowadays available on the network [12] but testing in on SP surveys results is a first step in the unified deployment of this method.

Finally, in phase 3, a discrete choice model based in regression logistics is proposed. More precisely, a binomial logit model is set up [13]. After obtaining data and selecting the predictors, the probabilistic-behavioural model calibration is performed. To construct the model, a first three-set cross pre-validation process helps to assess the stability of the prediction, and to estimate the mean prediction error of the model. More precisely, the sample is divided in two parts (*2/3* of the data are used for calibrating a first-choice model, and the remaining third to validate it [14]). However, we generate here *3* sets of data by partitioning the whole survey sample into one calibration set (*2/3* of data) and one validation set (*1/3* of the data) randomly, and this three times. Those three sets are used to define three different models, which are compared using mean of the mean square Out-Of-Bag errors (OOB) of each model [15]. This model is finally validated via a significance test and an OOB error analysis, where the Root Mean Square Error (RMSE) is calculated and analyzed.

3. Results

3.1 Field of Study and Data Frame

The field of study for testing this methodology involved a city-centre culture and high-tech store planning to relocate to a shopping centre on the outskirts of Saint-Étienne, France. The store belongs to one of the leading French store chains in this domain and is locally regarded as a historical and significant element within the city-centre retail structure. The relocation was scheduled for the fall of 2017, with the new location being an already existing peripheral shopping centre.

A stated preferences (SP) survey was carried out in November 2016. The survey asked customers which changes were induced by store relocation within the frequency of their shop visits, in a before-after assessment analysis [16]. In the present case, the "before" situation was the situation in 2016, with the store located in the city centre, and the "after" was the situation with the store located in a peripheral retail area. More details of the survey can be found in [5]. After an adjustment procedure to ensure data uniformity, the final dataset comprised 1044 customers' responses.

3.2 Predictor Selection

To model the end-consumer (customer) preferences between the two possible store locations, we define a dependent variable, here "*pref*", as a Boolean variable (it takes the value of "*1*" for the urban periphery location and "*0*" for the city centre store, i.e. the previous location). First, we apply the proposed tree-based conditional inference framework. The set of attributes used for modelling customer's choices is presented in Table 1. The decision (preference variable) is *pref* (binary variable of choosing city-centre or peripheral location for shopping) and a categorial variable to explain the main causes/purposes of this choice is *reason* (not used for modelling purposes). All other attributes are the potential explanatory variables used in the present analysis. The tree-based algorithm is deployed under R-project software, version 3.4.3.

Table 1 Main variables involved in the analysis. Source: authors

Variable Name	Description	Type
Pref	Binary variable showing the preference of the customer between the city centre location of the store and the urban peripheral one	Factor, binary
Reason	Reason for making this choice of preference	Factor, 10 values
Day	Day of the week when the respondent answered the survey	Factor
motive	Trip purpose, according to French trip survey standards [18]	Factor with 3 values
act_purch	Product purchasing at the store (or not) the present day (binary)	Factor, binary
dist_resid	Distance from customer residence to the city-centre store	Ordinal
dist_cov	Distance actually covered from the trip origin to the city-centre store	Ordinal
trip_mode	Trip mode used to make the trip	Factor with *3* values
trip_chain	Detail of the whole trip-chain of the customer including all the trips made between the first departure and its final destination [9,18]	Factor
s_club	Membership in the store's fidelity program (club)	Factor, binary
freq_dor	How often the customer visits the store located in the city-centre	Ordinal with *5* levels
d_mode	Trip mode used for daily commuting (i.e. where no shopping is made)	Factor with *4* values
reloc_info	Customer's awareness about the store relocation before the survey (binary)	Factor, binary
gender	Respondent's gender	Factor, binary
Age	Respondent's age, grouped into *4* categories, adapted from CEREMA [18]	Ordinal with *4* levels
profession	Respondent's profession, adapted from CEREMA [18]	Factor with *7* levels
stud_lev	Respondent's study level, adapted from CEREMA [18]	Ordinal with *4* levels

The results of the ranking proposed by the conditional inference tree algorithm are presented in Table 2. To make a restrained set of variables, we selected only those whose appearance statistics are higher than *50*%, leading to the selection of *5* variables. Two variables are identified as being the primary ones (with an appearance statistic of *100*%), distance to the location ("*dist_cov*") and daily mode ("*d_mode*", i.e. the mode of transport users take for daily trips). Then, the frequency to the surveyed location (i.e. the city centre one) is third, with *90*%, followed by the mode of the actual shopping trip (*77*%) and the purpose of the trip, named "*motive*" here (*58*%).

Table 2 Variable ranking with conditional inference trees on 100 sub-samples. Source: authors

Variable	Appearance statistics	Importance ranking order
dist_cov	100%	1
d_mode	100%	2
freq_dor	90%	3
trip_mode	77%	4
Motive	58%	5
Day	46%	6
stud_lev	31%	7
dist_res	14%	8
Age	11%	9
reloc_info	10%	10
Profess	9%	11
Gender	4%	12
trip_ch	3%	13
S_club	2%	14
act_purch	1%	15

The five predictors (attributes) retained and involved in the following modelling process are related into a conditional tree, as shown in Figure 2. However, when deploying the logit model using the dataset made of all the *1044* individuals, it can be stated that the variable "*trip-mode*" has a very high correlation with the variable "*d_mode*", leading to the removal of this last attribute. Finally, the remaining predictors are: "*dist_cov*", "*d_mode*", "*freq_dor*" and "*motive*".

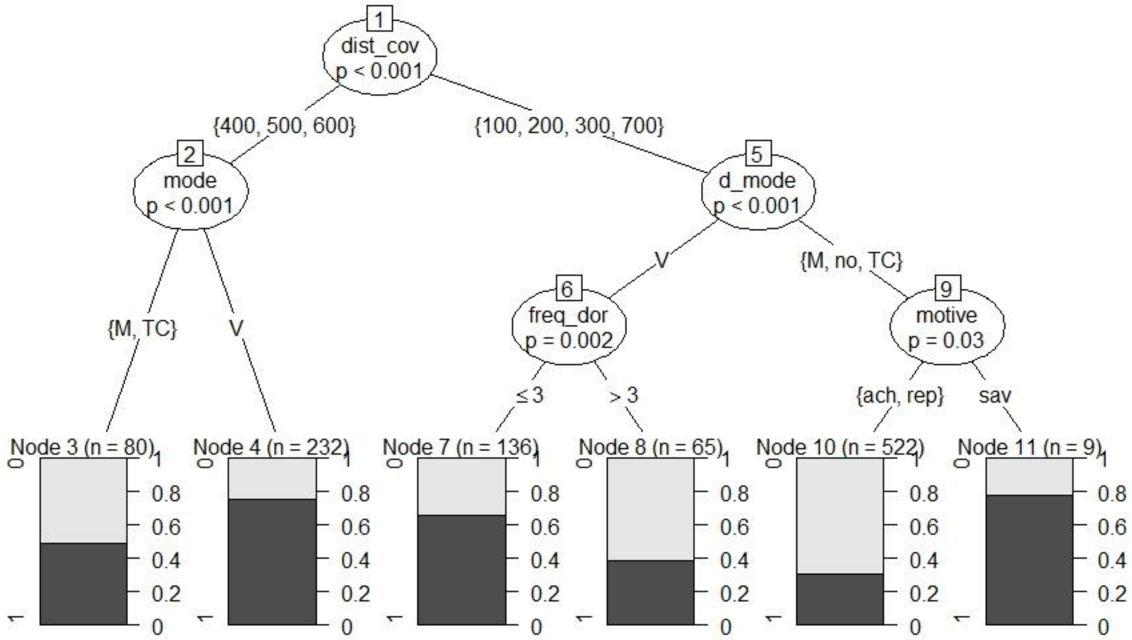

Fig. 2 Conditional inference tree (entire sample) with five best ranked variables. Source: authors

These variables are intuitive, as existing literature identifies trip mode and distance as key predictors of shopping location choice. However, the mode considered by the conditional tree algorithm is not the specific trip mode for shopping but rather the users' usual mode of daily transport. Both modes, though inherently correlated, influence shopping location choice. The selection of a mode for a shopping trip is closely tied to the trip's purpose: while household-work commuting trips are often made using public transport, work-shopping trips, typically occurring once a week or twice a month, are more likely to be by car. Distance is another critical factor, as it interacts with mode choice: car users tend to perceive distance as less constraining than pedestrians, which often translates to reduced accessibility of city centres compared to peripheral, car-oriented malls.

3.3 Logit Model Calibration and Validation

After selecting the four attributes, the calibration and validation processes were carried out. First, we present the three-set cross-validation process. The original data frame is thus divided into three subsamples $DataV_1$, $DataV_2$ and $DataV_3$, respectively, of *348* individuals. Table 3 shows the Out-Of-Bag (OOB) errors for each of the three cross-validation models and provides the mean error expected for the final model.

Table 3 Cross validation process and mean error expected from the final model. Source: authors

Model	**Learning-sample**	**Testing-sample**	**OOB with the learning sample**	**OOB with the testing sample**	**RMSE**
M_1	696 individuals	348 individuals	32.0%	32.8%	32.4%
M_2	696 individuals	348 individuals	33.2%	28.4%	30.9%
M_3	696 individuals	348 individuals	30.9%	33.6%	32.3%
Mean error expected for the final model (RMSE)					**31.8%**

We observe a Root Mean Square Error (RMSE) of around *31-32*% for all three models, with an average of *31.8*%. These results are consistent with the literature [4,6] and show the robustness of the approach, since the three models present the same attributes and the RMSE remains in the same order of magnitude and is acceptable for this type of models. Indeed, although those rates are moderate to high with respect to various choice models [13,14], in urban mobility, mainly in trips other than home-work, which are less systematic [19], a RMSE between *15* and *30*% remains acceptable [9,11]. After that analysis, we develop our logit model with the entire dataset. In order to provide an easy reading of the model parameters.

Table 4 Final model parameters. Source: authors

```
Deviance Residuals:     Min       1Q         Median       3Q      Max
                      -2.0013  -0.9624      -0.6351    1.0349   1.9106

Coefficients:
                Estimate   Std. Error    z value     Pr (>|z|)
(Intercept)    -0.34120   0.26978      -1.265      0.20596
motiverep       0.15079   0.15532       0.971      0.33165
motivesav       2.31379   0.81796       2.829      0.00467 **
dist_cov200    -0.06370   0.20317      -0.314      0.75389
dist_cov300     0.59919   0.19007       3.152      0.00162 **
dist_cov400     1.15760   0.23910       4.842      1.29e-06 ***
dist_cov500     1.22035   0.22899       5.329      9.86e-08 ***
dist_cov600     0.62203   0.31250       1.990      0.04654 *
dist_cov700    -0.19589   1.27824      -0.153      0.87820
d_modeno        0.52085   0.23413       2.225      0.02611 *
d_modeTC        0.12729   0.20918       0.609      0.54284
d_modeV         1.08934   0.20914       5.209      1.90e-07 ***
freq_dor       -0.26166   0.06188      -4.228      2.35e-05 ***
---
Significance codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

    Null deviance: 1444.5  on 1043  degrees of freedom
Residual deviance: 1271.7  on 1031  degrees of freedom
AIC: 1297.7             Number of Fisher Scoring iterations: 4
```

Table 4 presents the parameters of the binomial logit model, as well as that significance test, which shows how significant the continuous variables and the values of the discrete ones identified by the model are. Making the specification $\alpha = 0.1$, we assume that the continuous variable or the value of the discrete variable is significant if the condition is Pr < α. Therefore, variables "*motivesav*", "*dist_cov300*", "*dist_cov400*", "*dist_cov500*", "*dist_cov600*", "*d_modeno*", "*d_modeV*" and "*freq_dor*" are statistically significant as shown in Table 4.

After calibration of the model with the whole dataset, the OOB error obtained (see Table 5) is *31.30* and the prediction error for both locations remains almost the same (*31.25*% and *21.41*%) respectively for city centre and periphery location.

Table 5 OOB error of the final model. Source: authors

Value in the dataset	Predicted as "*0*"	Predicted as "*1*"	OOB error
0 = city centre location	407	185	31.25%
1= periphery location	142	310	31.41%
Total OOB error			**31.30%**

For the model fitting, it is assumed that the fit is satisfying if $\frac{deviance}{degrees\ of\ freedom} > \mathbf{1}$ [17]. As shown in the table of deviance (Table 6), that condition is fulfilled for each of the predictors and for the null deviance, because the residual deviance (*Resid. Dev*) is always higher than the residual degree of freedom (*Resid. Df.*).

Table 6 Analysis of deviance. Source: authors

```
Terms added sequentially (first to last)

          Df  Deviance         Resid.Df       Resid.Dev   Pr(>Chi)
NULL                           1043           1444.5
motive    2    6.561           1041           1437.9       0.03761 *
dist_cov  6   98.922           1035           1339.0      <2.2e-16 ***
d_mode    3   49.158           1032           1289.9      1.207e-10 ***
freq_dor  1   18.186           1031           1271.7      2.003e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.
```

4. Discussion

The model's parameters highlight some indications that allow understanding the behaviour of urban inhabitants concerning shopping location choice when a store is relocated, regarding shopping trips. The "*intercept*" has a non-significant parameter, which that the predictors we involve in the modelling process can provide a sufficiently accurate prediction of the dependent variable, without any constant. The variable "*motivesav*" describes the customers are visiting the store for "after-sale service". It has a coefficient of *2.3* (positive and higher than *0*) with an odds ratio of $e^{2.3} = 10$. The probability of accepting the new location of the store is then positively evaluated by customers who have already purchased ("*motivesav*") probably because they have constrained to post-sale services.

The calibration results also show the inclination of customers to reach the new store location according to the distance covered to reach the city-centre and to the mode generally used for commuting. Indeed, "*d_modev*", considering that as reference value has been chosen "no commuting for work", commuters show a propensity to accept the new location; specifically, the value of the variable "*d_mode*" describing the car-use for daily commute has a positive parameter, which is *1.1* which means that the probability of the alternative "*1*" increases by a factor of $e^{1.1} = 3$ when customers are dailycar users. Customers also accept this new location if the distance covered to reach the city centre is more than *2* km, without exceeding 80 km (*dist_cov300* - *dist_cov600*). Users covering a distance longer than *80* km (*dist_cov700*) tend to not prefer the new location, given that they could be attracted by other possible store options. The results confirm that the probability to accept the new store location decreases if customers are frequent visitors ("*freq_dor*"). This certainly means that in the city centre frequent visitors visit the store more often only to discover new articles or to ask for special discounts. Moreover, results confirm that the probability of accepting the new store location decreases if customers are frequent visitors ("*freq_dor*"). This certainly means that in the city centre, frequent visitors visit the store more often only to discover new articles or to ask for special discounts. In this context, the framework indicates the significance of how often customers visit the city centre and the trip mode they use for, but also the distance covered to reach the store, and the trip scope.

The test case involved in this research is the relocation of a city centre store moving to settle down in a shopping centre located in the eastern outskirts of the city. The identification of car usage as a significant factor influencing trip mode is consistent with findings in the existing literature [4-6]. However, during the predictor selection process, the model establishes a link between car usage and travel distance. Interestingly, the decision of end-consumers traveling less than *5* kilometres to reach the city centre does not consistently correlate with car usage. This highlights that distance plays a more critical role. Distance is inherently tied to travel time, trip origin, and trip chain. Recent works such as [20] demonstrated that land use relationships of destinations in a trip chain significatively impact customer decision making. Travel time to reach a retail store is already known to be very critical, especially when uncertainty exists regarding product availability [21]. Additionally, the model emphasizes the importance of car usage during daily commuting. This finding reinforces the structuring impact of work commuting on general mobility, as recently observed by [20].

These insights lead to two key implications. The first is managerial: the model underscores the necessity for retailers to thoroughly evaluate the suitability of relocation strategies, considering then customers' behaviours and choices in a scenario assessment logic [9]. The second implication pertains to public policymakers: the findings suggest that retailers are more inclined to relocate to urban peripheries if urban planning policies promote extensive car use in daily commuting [22].

The significance of the rhythm customers visit the city centre store can be related to the store's peculiarities or to the store's format. Indeed, the store under study sells culture, library and high-tech, and is not prone to regular visits such as grocery stores. Therefore, when it moves 3 kilometres farther in the periphery, customers who comes about once a month or less can still continue visiting, as shown by the model and the survey's responses. However, the significance of the rhythm can also mean the preference the customer has to visiting the city centre, rather than shopping centres outskirts of the city.

5. Conclusion

This paper proposed a data-driven methodology to calibrate a discrete choice model that integrates a large set of attributes to model the customer choice between a central store and a periphery one. It combined a SP survey, a tree-based procedure to pre-select attributes, and a binomial logit model. The model was developed on the basis of a real case study of a store relocation at Saint-Etienne, France. Results show that the selection of attributes has an impact on the final model, but, with a good pre-selection algorithm, the computational times of calibration can be improved, leading to the deployment of a suitable model able to be used on applications in a reduced amount of time.

The proposed modelling process serves as a straightforward decision-support tool, as demonstrated by the test case. The most influential factors identified in this specific scenario, comparing a peripheral location with a city centre one, appear realistic. These findings provide public decision-makers with actionable levers to address or mitigate retail sprawl. For retailers, the model offers insights into the potential effects of relocating from a city centre to a peripheral area. However, the test case focuses on a single, unique store. Retailers could gain more precise and practical results if the methodology were applied to stores from different classes of trade. This would also yield valuable insights for public decision-makers. A broader generalization encompassing all trade classes could contribute to retail location planning, helping preserve the economic viability of retailing.

Two key improvements stand out as particularly beneficial. First, enhancing the data collection process would strengthen the model's predictive reliability. Integrating stated preferences (SP) surveys with revealed preferences (RP) data could reduce the impact of hypothetical factors on user choices. Second, the framework requires further validation through use cases involving multiple potential new locations, enabling the application of multinomial and nested logit models. This would provide retailers with a clearer set of considerations for selecting a new location. Future research directions should focus on combining SP and RP surveys to develop a more holistic and systematic framework. Additionally, expanding the methodology to evaluate more than two potential locations would offer valuable decision-making support for both retailers and policy makers.

Acknowledgments

This research is part of first author's PhD. Thesis and was carried out between 2015 and 2019 at Ecole des Mines de Saint-Etienne, France, funded via an internal scholarship, with a 6-months stay at Tor Vergata University, Italy. The authors thank the editor and reviewers whose suggestions were very useful in revising the paper.

References

[1] Formánek, T. & Sokol, O. (2022). Location effects: Geo-spatial and socio-demographic determinants of sales dynamics in brick-and-mortar retail stores. Journal of Retailing and Consumer Services 66, 102902. DOI: 10.1016/j.jretconser.2021.102902.

[2] Nahiduzzaman, K.M., Holland, M., Sikder, S.K., Shaw, P., Hewage, K. & Sadiq, R. (2021). Urban transformation toward a smart city: an e-commerce–induced path-dependent analysis. Journal of Urban Planning and Development 147(1), 04020060. DOI: 10.1061/(ASCE)UP.1943-5444.0000648.

[3] Dellaert, B.G., Arentze, T.A. & Timmermans, H.J. (2008). Shopping context and consumers' mental representation of complex shopping trip decision problems. Journal of Retailing 84(2), 219-232. DOI: 10.1016/j.jretai.2008.02.001.

[4] Ortúzar, J.D. & Willumsen, L.G. (2024). Modelling transport. 5th edition. Chichester: John Wiley & Sons. DOI: 10.1002/9781119993308.

[5] Hounwanou, S., Comi, A., Gonzalez-Feliu, J. & Gondran, N. (2018). Inner city versus urban periphery retailing: store relocation and shopping trip behaviours. Indications from Saint-Etienne. Transportation Research Procedia 30, 363-372. DOI: 10.1016/j.trpro.2018.09.039.

[6] Hensher, D.A. & Johnson, L.W. (2018). Applied discrete-choice modelling. London: Routledge. DOI: 10.4324/9781351140768.

[7] Ölmez, M. & Lindemann, U. (2014). Managing attribute complexity for user-centered decision support systems. Procedia Computer Science 28, 130-137. DOI: 10.1016/j.procs.2014.03.017.

[8] Timmermans, H. (2012). Retail and spatial consumer behaviour. In Wells, V. & Foxall, G. (eds.), Handbook of Developments in Consumer Behaviour (pp. 121-146). London: Edward Elgar. DOI: 10.4337/9781781005125.

[9] Gonzalez-Feliu, J. (2018). Sustainable urban logistics: Planning and evaluation. London: John Wiley & Sons. DOI: 10.1002/9781119421948.

[10] Hothorn, T., Hornik, K. & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. Journal of Computational and Graphical statistics 15(3), 651-674. DOI: 10.1198/106186006X133933.

[11] d'Arcier, B.F., Andan, O. & Raux, C. (1998). Stated adaptation surveys and choice process: some methodological issues. Transportation 25(2), 169-185. DOI: 10.1023/A:1005042505063.

[12] Ang, K.L.M., Seng, J.K.P., Ngharamike, E. & Ijemaru, G.K. (2022). Emerging technologies for smart cities' transportation: geo-information, data analytics and machine learning approaches. International Journal of Geo-Information 11(2), 85. DOI: 10.3390/ijgi11020085.

[13] Fullerton, A.S. (2020). Logit and probit: binary and multinomial choice models. Ew York: SAGE Publications Ltd. DOI: 10.4135/9781526421036845007.

[14] Cramer, J.S. & Ridder, G. (1988). The logit model in economics. Statistica Neerlandica 42(4), 297-314. DOI: 10.1111/j.1467-9574.1988.tb01241.x.

[15] Breiman, L. (1996, January). Out-of-bag estimation. Technical report, Department of Statistics, University of California, Berkeley. Retrieved January 31, 2025, from https://www.stat.berkeley.edu/pub/users/breiman/OOBestimation.pdf

[16] Bagul, A. & Rajhans, R.K. (2024). How to Conduct Before-and-after Studies. New York: SAGE Publications Ltd. DOI: 10.4135/9781529685626.

[17] Kuhn, M. & Johnson, K. (2013). Applied predictive modeling. New York: Springer. DOI: 10.1007/978-1-4614-6849-3.

[18] CEREMA (2021). Mobility surveys certified by Cerema (EMC^2) (in French). Principes méthodologiques. CEREMA, Collection Retrieved from https://www.cerema.fr/fr/centre-ressources/boutique/

[19] Ortúzar, J. de D. & Willumsen, L.G. (2011). Modelling transport. 4th Edition. John Wiley & Sons. DOI: 10.1002/9781119993308.

[20] McCullagh, P. & Nelder, J.A. (1998). Generalized linear models, 2nd ed. Monographs on statistics and applied probability. Boca Raton: Chapman & Hall/CRC. DOI: 10.1201/9780203753736.

[21] Huang, A. & Levinson, D. (2017). A model of two-destination choice in trip chains with GPS data. Journal of Choice Modelling, 24, 51–62. DOI: 10.1016/j.jocm.2016.04.002.

[22] Pumain, D. & Reuillon, R. (2017). Urban dynamics and simulation models. Cham: Springer International Publishing. DOI: 10.1007/978-3-319-46497-8.