A STUDY ON THE MODELING OF SIGNIFICANT CROSS EFFECTS IN PURCHASE INCIDENCE: ANALYZING ARTIFICIAL NEURAL NETWORK METHODS AND MULTIVARIATE PROBIT MODELING

[1]CHENG NANNAN, [2]Divya Midhunchakkaravarthy

Abstract

This study focuses on modeling enormous cross-effects in purchase incidence in hopes of better understanding consumer behavior across related product categories. It tests how well many modeling approaches, including Artificial Neural Networks (ANNs) and Multivariate Probit (MVP), capture complex cross-category interactions. While MVP models provide a clear statistical framework grounded on economic theory, ANNs are known for their adaptability when dealing with large datasets and nonlinear interactions. Studies evaluate these methods by comparing their accuracy in predicting results, clarity of presentation, and ability to capture the nuances of buying behavior across various product types. The findings demonstrate that MVP models provide much higher interpretability and theoretical consistency compared to ANNs, despite the fact that ANNs perform pretty well in terms of prediction accuracy. The findings add to broadening methodologies in consumer choice modeling and give practical insights into marketing strategies and decision-making. To examine large-scale cross effects on sales incidence, this study compares and contrasts the effectiveness of ANN approaches with MVP modeling. The goal of this research is to provide light on the relative strengths of two cutting-edge methods for understanding complicated relationships and making predictions about consumer behavior in large datasets by comparing and contrasting them. By using ANN and MVP algorithms to purchase incidence data, this study intends to assess the methodology's capacity to provide more precise and actionable insights into consumer behavior. This comparative study aims to assist researchers doing marketing or consumer research in selecting the most appropriate modeling approaches by focusing on model accuracy, computing efficiency, and interpretability.

Keywords: *Multimodal Probit Simulation, Artificial Intelligence Neural Networks, Sales Prevalence, Data Interpretation.*

Introduction

In the present consumer behavior analysis environment, it is critical for organizations to understand the elements that affect purchase incidence in order to successfully change their strategy. The need for high-level modeling approaches in order to conduct reliable analysis and forecasts is heightened by the abundance of data and the complexity of client interactions. Multivariate probit modeling and ANNs are two state-of-the-art ways to study large-scale cross effects in purchase incidence. Examining and contrasting these two approaches is the goal of this study. An alternative approach to comprehending customer behavior was to use Artificial Neural Networks, which are well-known for their capacity to represent complex data patterns and non-linear interactions (Wang, 2018). A statistical framework for analyzing numerous linked binary outcomes is provided by Multivariate Probit Modelling, however, when dealing with many simultaneous purchasing alternatives. In order to have a better grasp of the pros and disadvantages of each method, this research compares and evaluates two methodologies. The goals are to estimate purchase incidence and capture the cross effects that impact customer decisions. The study's anticipated benefits included enhancing the prediction capabilities of marketing and economic studies and providing a clearer picture of consumer behavior. Learning when and how people purchase various goods is a primary objective of marketing research. Rather than making purchases independently, customers are frequently influenced by interactions between related goods, which are known as cross-category effects. Marketing strategies such as product packaging, pricing, and promotional campaigns rely on precise simulations of these interactions. Traditional methods, such MVP models, have been widely used to examine customer behavior. Using probabilistic methodologies, MVP models provide a systematic and theoretically informed framework to account for links between purchase occurrences across categories. However, when it comes to complex consumer decision-making including non-linear interactions or large cross-effects, their reliance on known assumptions about the relationships among variables could restrict their ability to capture the whole complexity. Instead of depending on strict parametric assumptions, data-driven and flexible ANNs may depict complex interactions and very nonlinear connections. Their adaptability makes them a promising alternative for studying customer behavior on a scale. A typical complaint about ANNs is that, despite their versatility, they are not interpretable, which makes it hard to extract valuable insights from them. This study's findings might be useful for academics and marketers in making informed decisions on consumer choice research tools, since they compare and contrast traditional econometric models with more advanced machine learning techniques (Alptekinoğlu & Semple, 2021).

Background of study

Accurate estimate of the factors impacting purchase frequency is crucial for developing effective marketing strategies and comprehending market dynamics via consumer behavior research. Due to the existence of non-linear interactions and a variety of interdependent elements, understanding the effects of consumer decision-making is needed for advanced analytical tools. The "purchase incidence" of a consumer is defined as the frequency with which they make a purchase during a certain time period (Yang & Sudharshan, 2019). It was essential for accurate modeling to capture these cross-effects, but doing so becomes more challenging when dealing with large-scale datasets that involve a variety of goods and customer variables. Statistical approaches such as Multivariate Probit Modeling have historically been used for studies that include complex datasets and a large number of binary outcomes. Applying this approach to cases with linked purchasing decisions yields very useful insights into the plethora of factors that impact choices made at the same time (Aouad & Désir, 2022).

Purpose of the study

Multivariate Probit modeling and ANNs are two state-of-the-art modeling approaches that this study aimed to investigate and assess for their ability to analyze large-scale cross effects on purchase occurrence. The researchers set out to determine which strategies were the most effective at mining large datasets for information on the complex web of variables that influence customers' purchasing decisions. The purpose of this study was to compare and contrast artificial neural networks with multivariate probit models in order to learn more about the relative merits of these two methods for analysing customer behavior and making purchase predictions. Results were useful for both theoretical and practical marketing and economic research. Additionally, this will lead to the development of more accurate models for evaluating and forecasting consumer purchases.

Literature review

Several approaches have been used to achieve this goal. In order to manage the massive amounts of complicated current consumer data, researchers have been concentrating on developing sophisticated modeling approaches. For the analysis of several linked binary outcomes, such as purchasing numerous products at once, Multivariate Probit Modeling has become an important tool. This method sheds light on the elements that impact purchasing choices by allowing one to analyze the interdependencies between different purchases. In particular, it was helpful for processing large datasets with several dimensions of data and comprehending complicated consumer patterns due to its capacity to predict correlations between binary outcomes. Conventional statistical models faced a formidable threat from the rise of ANNs (Zheng et al., 2019). The architecture and operation of the human brain served as the basis for the development of ANNs, which were designed to detect complex patterns and non-linear correlations in data. Because they can adapt to new situations and learn from big datasets, they are ideal for simulating complicated customer interactions. Because of its capacity to depict complicated and non-linear interactions, ANNs may be able to surpass more conventional approaches in forecasting a number of consumer outcomes, such as purchase frequency, according to recent study. When it comes to large-scale cross effects on purchase incidence, there is a lack of study comparing ANNs with Multivariate Probit Modeling, despite the fact that both have their merits. In terms of accuracy and flexibility, ANNs surpass conventional models when given highly dimensional data, as shown in the study. Because of its interpretability and the insights, it provides links between linked binary outcomes, multivariate probit modeling is still relevant (Aouad & Segev, 2021).

Research question

- What is the impact of personalization on comparing artificial neural networking cadence?

Research methodology

Quantitative research refers to studies that examine numerical readings of variables using one or more statistical models. The social environment may be better understood via quantitative research. Quantitative approaches are often used by academics to study problems that impact individuals. Objective data presented in a graphical format is a byproduct of quantitative research. Numbers are crucial to quantitative research and must be collected and analyzed in a systematic way. Averages, predictions, correlations, and extrapolating findings to larger groups are all possible with their help.

Research design: In order to analyse quantitative data, SPSS version 25 was used. When analysing the statistical association, the odds ratio and 95% confidence interval were used to determine its direction and size. A statistically significant threshold was suggested by the researchers at $p < 0.05$. The primary features of the data were identified by a descriptive analysis. Mathematical, numerical, or statistical evaluations using quantitative methodologies are often used for data gathered from surveys, polls, and questionnaires, or by modifying existing statistical data using computing tools.

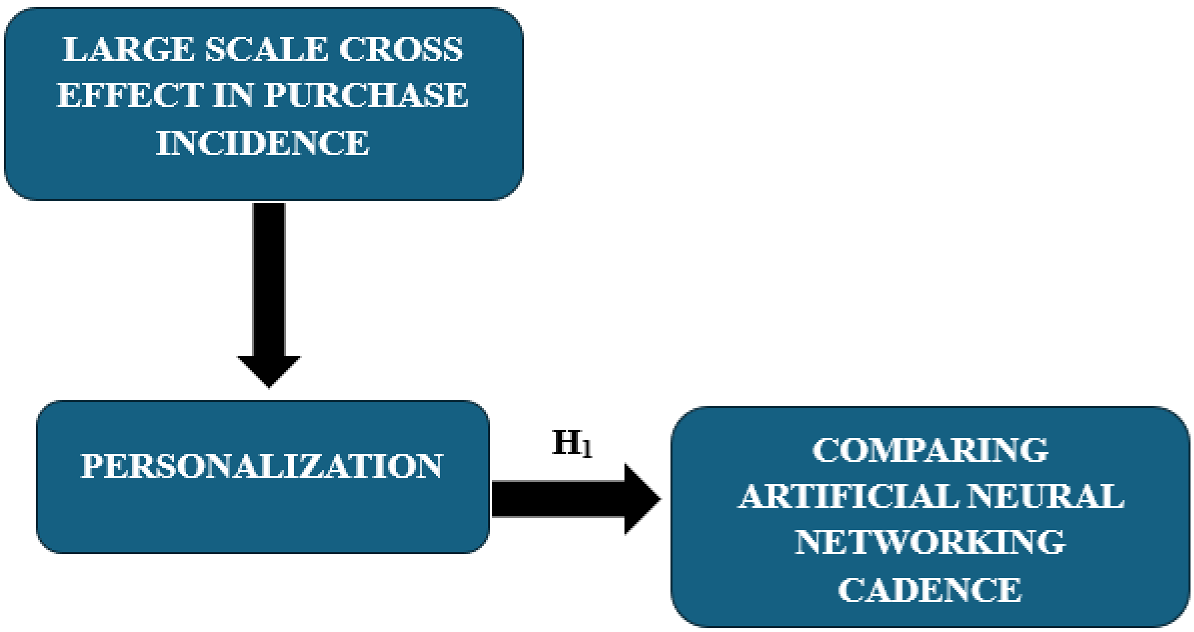
Sampling: Research participants filled out questionnaires to provide information for the research. Using the Rao-soft programme, researchers determined that there were 735 people in the research population, so researchers sent out 850 questionnaires. The researchers got 810 back, and they excluded 32 due to incompleteness, so the researchers ended up with a sample size of 778.

Data and Measurement: A questionnaire survey functioned as the primary data collection instrument for the investigation. The survey had two sections: (A) General demographic information and (B) Responses on online and non-online channel factors on a 5-point Likert scale. Secondary data was obtained from many sources, mostly on internet databases.

Statistical software: The statistical analysis was conducted using SPSS 25 and MS-Excel.

Statistical Tools: To grasp the fundamental character of the data, descriptive analysis was used. The researcher is required to analyse the data using ANOVA.

Conceptual framework



1. RESULT

❖ **Factor analysis**

One typical use of Factor Analysis (FA) is to verify the existence of latent components in observable data. When there are not easily observable visual or diagnostic markers, it is common practice to utilize regression coefficients to produce ratings. In FA, models are essential for success. Finding mistakes, intrusions, and obvious connections are the aims of modelling. One way to assess datasets produced by multiple regression studies is with the use of the Kaiser-Meyer-Olkin (KMO) Test. They verify that the model and sample variables are representative. According to the numbers, there is data duplication. When the proportions are less, the data is easier to understand. For KMO, the output is a number between zero and one. If the KMO value is between 0.8 and 1, then the sample size should be enough. These are the permissible boundaries, according to Kaiser: The following are the acceptance criteria set by Kaiser:

A dismal 0.050 to 0.059, subpar 0.60 to 0.69
Middle grades often range from 0.70 to 0.79.
Exhibiting a quality point score between 0.80 and 0.89.
They are astonished by the range of 0.90 to 1.00.
Table 1: KMO and Bartlett's Test for Sampling Adequacy Kaiser-Meyer-Olkin measurement:

.989

The outcomes of Bartlett's test of sphericity are as follows: Approximately chi-square degrees of freedom = 190 significance = 0.000

This confirms the legitimacy of claims made just for sampling purposes. Researchers used Bartlett's Test of Sphericity to ascertain the significance of the correlation matrices. A Kaiser-Meyer-Olkin value of 0.989 indicates that the sample is sufficient. The p-value is 0.00 according to Bartlett's sphericity test. A positive outcome from Bartlett's sphericity test indicates that the correlation matrix is not an identity matrix.

Table 10: KMO and Bartlett's

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.989
Bartlett's Test of Sphericity	Approx. Chi-Square	3252.968
	df	190
	Sig.	.000

The overall significance of the correlation matrices was further confirmed by using Bartlett's Test of Sphericity. A value of 0.989 was the Kaiser-Meyer-Olkin sampling adequacy. By using Bartlett's sphericity test, researchers found a p-value of 0.00. A significant test result from Bartlett's sphericity test demonstrated that the correlation matrix was not a correlation matrix.

Test For Hypothesis:

- ❖ **INDEPENDENT VARIABLE**
  - ➢ **Large-Scale Cross Effect on Purchase Incidence**

You may influence goods that are similar to or related to a customer's purchases in several ways. The specifics dictate the manner in which this kind of discussion could unfold. Cases and headphones are only two of the many optional extras that come with most phones. In order to improve their coffee-making experience, the researcher can think about purchasing a high-end machine. When promoting two distinct goods at the same time is only one of many instances when this is necessary. More generalized patterns and trends may possibly have had a part. Organic food and exercise gear, for example, would see a meteoric rise in demand if a newly formed health movement were to get substantial support. What consumers do in one area could affect their spending habits across several other categories. One of its results was this, which followed (Arzhenovskiy et al., 2020).

- ❖ **FACTOR**
- ➢ **Personalization**

The term "personalization" describes the practice of adapting goods and services to each customer by taking their unique tastes, habits, and requirements into account. Personalized suggestions, targeted marketing messages, and content are created by analyzing data such purchase history, browsing activity, demographics, and interests. Customer engagement, happiness, and the chance of purchase are all boosted by personalization since it makes interactions more relevant and meaningful. Online retailers' product suggestion engines, email marketing campaigns, and digital apps' adaptable user experiences are just a few examples (Bai et al., 2020).

- **DEPENDENT VARIABLE**
- **Comparing Artificial Neural Networking Cadence**

To address different kinds of problems, several types of artificial neural network topologies were used. researcher may gauge their efficiency, or cadence, by looking at how they deal with learning rates, training durations, generalizability, and task appropriateness. One example is the readily trainable Feedforward Neural Networks (FNNs), which are used when the data linkages aren't sequential. On the other hand, when dealing with intricate designs, their simplicity may be a drawback. For tasks requiring geographic data, such as picture identification, conventional Neural Networks (CNNs) just couldn't cut it. They use convolutional layers to identify visual elements and patterns that are hierarchical. Despite their excellent performance on visual tasks, convolutional neural networks are computationally and resource-intensive, which slows down their inference and training times. Sequential input was the original intent of Recurrent Neural Networks (RNNs) and their more sophisticated offspring, Long Short-Term Memory Networks (LSTMs). They excel at tasks that require them to maintain an internal state while examining sequences, such as language modeling or time-series prediction. Because of their intricate architecture, RNNs train more slowly and may have trouble with long-term dependencies; LSTM models attempt to solve these issues (Kuruczleki, 2020).

- **Relationship between personalization and Comparing Artificial Neural Networking Cadence**

The connection between personalization and comparing the cadence of ANNs is that ANNs analyze patterns of customer behavior to optimize and improve tailored experiences. Machine learning, and ANNs in particular, are crucial to personalization because they can sift through massive datasets and dynamically modify content, ads, and suggestions. The efficiency of customization is directly influenced by an ANN's cadence, which is the rate and frequency of its learning, updating, and adaptation to new input. Instant product suggestions on e-commerce platforms are just one example of how a high-cadence ANN may provide real-time customisation. While higher-cadence networks update their customization tactics in real-time, lower-cadence networks update them more often in response to long-term patterns. While quicker learning enhances responsiveness, it demands more processing power; slower adaptation may be more cost-efficient, but it runs the danger of falling behind growing customer preferences. Finding the right balance between the two is vital. In addition, the cadence of ANNs affects the time of interaction, making sure that consumers get individualized information at the most appropriate periods. Companies may improve their customization tactics by comparing ANN cadences. They can use lower-cadence models for wider trend analysis and high-cadence networks for dynamic interactions. Due to its optimizing effects on user experiences, engagement, and conversion rates, ANN cadence is an essential component of efficient customization (Shi, 2020).

Analysis of the above discussion, the researcher formulated the following hypothesis, which was to analyse the relationship between Personalization and Comparing Artificial Neural Networking Cadence.

H_{01}: There is no significant relationship between Personalization and Comparing Artificial Neural Networking Cadence.

H_1: There is a significant relationship between Personalization and Comparing Artificial Neural Networking Cadence.

Table 2: H_1 ANOVA Test

ANOVA					
Sum					
	Sum of Squares	**df**	**Mean Square**	**F**	**Sig.**
Between Groups	39588.620	255	5655.517	1055.921	.000
Within Groups	492.770	522	5.356		
Total	40081.390	777			

In this study, the result was significant. The value of F was 1055.921, which reaches significance with a p-value of .000 (which is less than the .05 alpha level). This means the ***"H_1: There is a significant relationship between Personalization and Comparing Artificial Neural Networking Cadence."*** was accepted and the null hypothesis was rejected.

Discussion

When doing research on the topic of modeling large-scale cross interactions in purchase incidence, the researcher may compare and contrast ANNs with Multivariate Probit Modeling. The relative merits of the two approaches determine how well they record and understand customer behavior. What makes ANNs unique among models is their capacity to mimic complicated, non-linear interactions in high-dimensional datasets. Even when more conventional statistical approaches have failed, their adaptability enables them to reveal complex patterns and linkages. Due to their adaptability, ANNs excel in scenarios with complex interdependencies, such as those requiring several linked purchasing choices. It may be more challenging to apply and comprehend ANN due to the fact that hyperparameter tinkering, data quantity, and data quality all significantly affect ANN performance. However, a better organized method for comprehending the correlations between several binary outcomes was provided by Multivariate Probit Modeling. This method was effective in shedding light on the relationship between immediate purchase choices and elucidating the statistical relationships that were operating. The fact that it generates findings that are easy to understand, and grasp is one of its benefits, and another is that it can manage situations where the dependent variables were organically connected. Due to the complexity and non-linearity of large-scale data sets, multidisciplinary probit models may fail to capture certain subtle and complicated relationships.

Conclusion

The research discovered that both ANN and Multivariate Probit Modeling perform equally well for analyzing large-scale cross effects on purchase incidence. Before settling on a strategy, several factors had to be considered, including the significance of interpretability and the need for precise prediction. In order to better understand and predict consumer behavior, it was proposed that these tactics be studied in more depth in future research, with an emphasis on hybrid models and powerful computational tools.

REFERENCES

1. Alptekinoğlu, A. and J. H. Semple. (2021). "Heteroscedastic exponomial choice". Operations Research. 69(3): 841–858.
2. Aouad, A. and A. Désir. (2022). "Representing Random Utility Choice Models with Neural Networks". arXiv preprint arXiv:2207.12877.
3. Aouad, A. and D. Segev. (2021). "Display optimization for vertically differentiated locations under multinomial logit preferences". Management Science. 67(6): 3519–3550.
4. Arzhenovskiy, S., T. Sinyavskaya, and A. Bakhteev. "Multivariate probit model for a priori assessment of behavioral risks in audit." Applied Econometrics 60 (2020): 102–14.
5. Bai, Junwen, Shufeng Kong, and Carla Gomes. "Disentangled Variational Autoencoder based Multi-Label Classification with Covariance-Aware Multivariate Probit Model." In Twenty-Ninth International Joint Conference on Artificial Intelligence and Seventeenth Pacific Rim International Conference on Artificial Intelligence {IJCAI-PRICAI-20}. California: International Joint Conferences on Artificial Intelligence Organization, 2020.
6. Kuruczleki, Éva. "Overcoming methodological issues in measuring financial literacy of companies, a proposed measurement model." In The Challenges of Analyzing Social and Economic Processes in the 21st Century. Szeged: Szegedi Tudományegyetem Gazdaságtudományi Kar, 2020.
7. Shi, Lei. "Bayesian analysis of multivariate ordered probit model with individual heterogeneity." AStA Advances in Statistical Analysis 104, no. 4 (June 23, 2020): 649–65.
8. Wang, Hai-Bin. "A Bayesian multivariate partially linear single-index probit model for ordinal responses." Journal of Statistical Computation and Simulation 88, no. 8 (February 25, 2018): 1616–36.
9. Yang, Z. & Sudharshan, D. (2019). Examining multi-category cross purchases models with increasing dataset scale–An artificial neural network approach. Expert Systems with Applications, 120, 310-318.
10. Zheng, Weikang, Zhigang Liu, and Junkang Guo. "An Optimized 3D Probe Using Sensitivity and Compliance Analysis." In ASME 2019 International Mechanical Engineering Congress and Exposition. American Society of Mechanical Engineers, 2019.