AN INVESTIGATION TO DISCUSS THE MODELLING OF MASSIVE CROSS EFFECTS IN PURCHASE INCIDENCE: COMPARING ARTIFICIAL NEURAL NETWORK TECHNIQUES AND MULTIVARIATE PROBIT MODELLING.

Cheng Nannan[1], Divya Midhunchakkaravarthy[1]

[1]Lincoln University College, Petaling Jaya, Malaysia.

ABSTRACT

Consumer behaviour across related product categories may be better understood by the modelling of huge cross-effects in purchase incidence, which is the focus of this work. In order to capture complicated cross-category interactions, it evaluates the performance of Artificial Neural Networks (ANNs) and Multivariate Probit (MVP) modelling methodologies. When it comes to big datasets and nonlinear interactions, ANNs are recognised for their flexibility, whereas MVP models provide a defined statistical framework based on economic theory. Research assesses these approaches by looking at how well they forecast outcomes, how easy they are to understand, and how well they reflect the dynamics of purchases across different categories. While ANNs do quite well in terms of prediction accuracy, the results show that MVP models provide far better interpretability and theoretical consistency. Practical insights for marketing strategies and decision-making are offered by the results, which also contribute to expanding approaches in consumer choice modelling. By contrasting and comparing the efficacy of Artificial Neural Network (ANN) methods with Multivariate Probit (MVP) modelling, this research investigates large-scale cross effects on sales incidence. This study aims to compare and evaluate two state-of-the-art approaches with the idea of revealing how well they understand complex linkages and can predict consumer behaviour in massive datasets. This research aims to evaluate the methodology's ability to deliver more accurate and practical insights into consumer behaviour by using ANN and MVP algorithms to purchase incidence data. Focussing on model correctness, computation efficiency, and interpretability, this comparison study may help researchers performing marketing or consumer research choose the right modelling methodologies.

Keywords: AI Neural Networks, Multimodal Probit Simulation, Sale Prevalence, Data Interpretation.

INTRODUCTION

Understanding the factors that impact purchase incidence is crucial for organisations to effectively adapt their strategy in the current consumer behaviour analysis landscape. Since there is an excess of data and complicated customer interactions, high-level modelling methodologies are essential for accurate analysis and forecasting. Two cutting-edge methods for researching large-scale cross effects in purchase incidence are Artificial Neural Networks (ANNs) and Multivariate Probit

Modelling. This research aims to analyse and compare these two methodologies. Using Artificial Neural Networks—renowned for its ability to depict intricate data patterns and non-linear interactions—was a varied way to understand consumer behaviour. When dealing with several simultaneous purchase options, however, Multivariate Probit Modelling provides a statistical framework for analysing many linked binary outcomes. By comparing and evaluating these two approaches, this study hopes to get a better understanding of the benefits and drawbacks of each in terms of estimating purchase incidence and capturing the cross effects that influence consumer choices. the researcher expected this study to help us better understand consumer behaviour and to improve the predictive power of marketing and economic studies. One of the main goals of marketing research is to understand how and when consumers buy different types of products. Interactions between similar items, called cross-category effects, often impact consumers' selections rather than consumers making purchases in isolation. In order to create successful marketing tactics like product bundling, pricing, and promotional campaigns, it is crucial to accurately simulate these interactions. There has been extensive use of more conventional approaches to studying consumer behaviour, such as Multivariate Probit (MVP) models. To account for connections between purchase occurrences across categories, MVP models use probabilistic techniques, providing a structured and theoretically informed framework. Nevertheless, when dealing with complicated consumer decision-making with enormous cross-effects or non-linear interactions, their capacity to capture the whole complexity may be limited due to their dependence on established assumptions about the connections among variables. A data-driven, adaptable alternative, Artificial Neural Networks (ANNs) can represent complicated interactions and very nonlinear connections without relying on rigid parametric assumptions. Because of their flexibility, they provide a potential substitute for investigating large-scale interactions in consumer behaviour. The lack of interpretability in ANNs is a common criticism, since it makes it difficult to glean useful insights from them, despite their flexibility. The results of this study can help academics and marketers choose the best methods for consumer choice analysis by shedding light on the benefits and drawbacks of using classic econometric models vs. more sophisticated machine learning approaches (Shi, 2019).

BACKGROUND OF THE STUDY

Developing efficient marketing strategies and understanding the market dynamics via consumer behaviour research relies heavily on correct estimation of the elements influencing buy frequency. Understanding the consequences of consumer decision-making required sophisticated analytical methods; this was no easy task due to the presence of non-linear interactions and the multiplicity of interdependent factors. A customer's "purchase incidence" is the number of times they make a purchase during a certain time frame, as used here (Kuruczleki, 2020). Capturing these cross-effects becomes more complex when working with large-scale datasets that comprise a range of items and customer attributes, which was crucial for accurate modelling. For investigations involving complicated datasets and a high number of binary outcomes, statistical methods like Multivariate Probit Modelling have traditionally been used. This method really comes into its own when used to scenarios with linked purchase decisions; doing so provides invaluable insights into the myriad of elements that influence choices made simultaneously (SİGEZE et al., 2019).

PURPOSE OF THE STUDY

The purpose of this research was to examine and evaluate two modern modelling techniques for analysing large-scale cross effects on purchase incidence: Multivariate Probit modelling and Artificial Neural Networks (ANNs). The purpose of this study was to compare and evaluate various methods according to how well they could extract the intricate network of factors that impact consumers' purchase choices from massive datasets. In order to better understand how effective artificial neural networks and multivariate probit models are at predicting purchases and uncovering hidden patterns in consumer behaviour, this research set out to examine and contrast the two approaches. Academic studies and real-world marketing and economic analysis both benefited from the results. Also, as a result of this, better models for assessing and predicting customer purchases will be created.

LITERATURE REVIEW

For this objective, a number of methods have been used. Researchers have been focussing on building advanced modelling methods to handle the large and complex modern consumer data. Using Multivariate Probit Modelling has become an essential method for analysing several related binary outcomes, such buying multiple things at once. The technique allows one to examine the interdependencies between various purchases, which in turn illuminates the factors that influence buying decisions. Its ability to forecast relationships between binary outcomes made it particularly useful for understanding complex consumer patterns and handling massive datasets with several dimensions of data (Kong et al., 2020). The emergence of Artificial Neural Networks constituted a significant challenge to the dominance of traditional statistical models. In order to identify complicated patterns and non-linear correlations in data, artificial neural networks (ANNs) were developed with inspiration from the structure and function of the human brain. They are perfect for modelling complex consumer interactions due to their ability to learn from large datasets and adjust to changing circumstances. Recent research has shown that artificial neural networks (ANNs) have the potential to outperform more traditional methods in predicting a variety of consumer outcomes, including purchase frequency, by making use of their ability to represent complex and non-linear interactions (Zheng et al., 2019). Although both ANNs and Multivariate Probit Modelling have their advantages, there has been less research comparing them in the context of large-scale cross effects on purchase incidence. When presented with data that is highly dimensional, ANNs outperform traditional models in terms of accuracy and adaptability, according to the research. Multivariate Probit Modelling remains significant, however, due to its interpretability and the insights it offers into the relationships between linked binary outcomes (Arzhenovskiy et al., 2020).

RESEARCH QUESTION

When evaluating the cadence of artificial neural networks, how does complementarity play a role?

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 & 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.

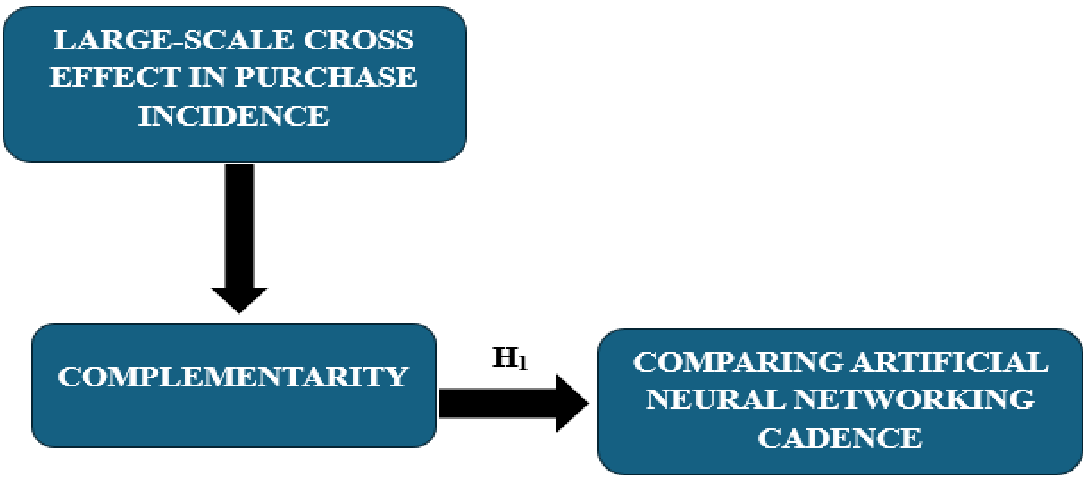
STATISTICAL SOFTWARE

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



RESULTS

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: .836

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 0.836 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 1: KMO and Bartlett's.

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.836
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 0.836 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: Many options exist for products that are comparable to or linked to a customer's purchases to be impacted by their purchasing habits. There are a lot of different ways this kind of conversation might take place, depending on the details. Phones often come with an assortment of add-ons, such as cases and headphones, that customers may purchase separately. The researcher may also consider investing in a high-end machine if they wanted to elevate their coffee-making experience. Quite a few situations call for this, not the least of which is when two separate items are promoted simultaneously. There may also have been a role for more broad-based trends and patterns. For instance, if a recently established health movement were to gain significant traction, the demand for organic foods and exercise equipment would surge. Customers' actions in one sector may have far-reaching consequences for their buying patterns across a wide range of products. That led to this as one of its outcomes (Wang, 2018).

FACTOR

Complementarity: When two or more items work together to improve each other's quality of life, the researcher say that they are complementary. This is a typical occurrence in consumer behaviour, where buying one thing usually makes buying another related one more likely. Products that work well together include sugar and coffee, bread and butter, and printer ink and cartridges. Commonly, these things work together to accomplish a common goal or improve the user experience as a whole. In the case of functionally reliant or more satisfyingly integrated items, complementarity is often the deciding factor for consumers. The cross-price elasticity of demand provides economic evidence of complementarity. In general, as the price of a product drops, consumers are more likely to buy the product that goes along with it since the combined cost is lower. If pasta were to go on sale, for instance, more people may buy pasta sauce. Strategies like product bundling, cross-promotions, and co-branding rely on marketers' grasp of complementarity. Businesses may increase sales, boost customer value, and forge synergies across product categories by capitalising on the complimentary nature of their goods (Hirk et al., 2018).

DEPENDENT VARIABLE

Comparing Artificial Neural Networking Cadence: Various architectures of artificial neural networks were used to tackle various types of issues. How well they handle learning rates, training periods, generalisability, and task suitability may be used to measure their efficiency, or cadence. For instances when the data links aren't sequential, one example is the easily trainable Feedforward Neural Networks (FNNs). However, their lack of complexity might make them less than ideal when dealing with complex patterns. Conventional Neural Networks (CNNs) weren't up to the task of image recognition or any other application that needed geographic data. To recognise hierarchical visual features and patterns, they use convolutional layers. The inference and training durations of convolutional neural networks are slowed down because they are computationally and resource-intensive, even if they get visual tasks very accurately. Recurrent Neural Networks (RNNs) and its more advanced progeny, Long Short-Term Memory Networks (LSTMs), were built to operate with sequential input. Tasks like language modelling or time-series prediction are well-suited to their abilities to examine sequences while preserving an internal state. Long short-term memory (LSTM) models aim to address the concerns raised by RNNs' complex design, which causes them to train more slowly and perhaps struggle with long-term dependencies (Mueller et al., 2019).

Relationship between Complementarity and Comparing Artificial Neural Networking Cadence: The capacity of Artificial Neural Networks (ANNs) to represent complicated, non-linear interactions among connected items is the link between ANNs and complementarity in consumer behaviour. When two products work well together, consumers are more likely to buy both of them. This phenomenon is known as complementarity. Many variables, like price, promotions, and customer preferences, impact these connections, making them often non-linear. Because of their ability to handle massive datasets and develop hidden representations without prior assumptions, ANNs are great at capturing such complex patterns. Because of this, they work wonders when trying to simulate the huge cross-effects in consumer behaviour across different types of products. When it comes to complicated dynamics, ANNs have more predictive power than more conventional approaches such as Multivariate Probit (MVP) models, which depend on pre-specified connections. But they are sometimes called "black-box" models and have problems with interpretability. Minimal viable product (MVP) models may have trouble scaling and being flexible, but they provide better insights into cross-effects. In conclusion, ANNs provide a promising method for detecting and forecasting complementarity; nevertheless, it is essential to strike a balance between prediction accuracy and interpretability when using ANNs for consumer behaviour research (Bai et al., 2020).

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

H01: There is no significant relationship between Complementarity and Comparing Artificial Neural Networking Cadence.

H1: There is a significant relationship between Complementarity 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	334	5655.517	2549.115	.000
Within Groups	492.770	443	5.356		
Total	40081.390	777			

In this study, the result was significant. The value of F was 2549.115, which reaches significance with a p-value of .000 (which is less than the .05 alpha level). This means the “H1: There is a significant relationship between Complementarity and

Comparing Artificial Neural Networking Cadence.” was accepted and the null hypothesis was rejected.

DISCUSSION

The researcher may observe the strengths and weaknesses of ANNs and Multivariate Probit Modelling in terms of simulating large-scale cross interactions in purchase incidence. Both methods have their advantages and disadvantages that impact how well they capture and comprehend consumer behaviour. The ability to simulate high-dimensional datasets with complex, non-linear interactions is what distinguishes ANNs from other models. Their flexibility allows them to uncover intricate relationships and patterns even when more traditional statistical methods have failed. Situations with complicated interdependencies, such as those needing many connected purchase decisions, are ideal for ANNs because to their flexibility. However, hyperparameter tweaking, data amount, and data quality all have a significant impact on ANN performance, which might make ANN implementation and interpretation more difficult. On the other hand, Multivariate Probit Modelling offered a more structured approach to understanding the relationships between several binary outcomes. This approach worked well for illuminating the connection between contemporaneous purchasing decisions and for understanding the statistical connections that were at work. One of its strengths is that it can handle cases where the dependent variables were naturally correlated, and another is that it produces results that are simple to grasp and comprehend. Multidisciplinary probit modelling may miss certain nuanced and intricate interactions in large-scale data sets because of their complexity and non-linearity.

CONCLUSION

In order to analyse large-scale cross effects on purchase incidence, the study found that both Multivariate Probit Modelling and Artificial Neural Networks function equally well. It was necessary to think about a number of things, such as the importance of interpretability and the need for accurate prediction, before deciding on a method. It was suggested that these strategies should be further investigated and improved in future studies, with a focus on using hybrid models and robust computational tools to better understand and anticipate consumer behaviour.

REFERENCES

1. 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.
2. 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.

3. Hirk, Rainer, Kurt Hornik, and Laura Vana. "Multivariate ordinal regression models: an analysis of corporate credit ratings." Springer Berlin Heidelberg, 2018.
4. Kong, Shufeng, Junwen Bai, Jae Hee Lee, Di Chen, Andrew Allyn, Michelle Stuart, Malin Pinsky, Katherine Mills, and Carla Gomes. "Deep Hurdle Networks for Zero-Inflated Multi-Target Regression: Application to Multiple Species Abundance Estimation." 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.
5. 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.
6. Mueller, Nikola S., Ivan Kondofersky, Gökcen Eraslan, Karolina Worf, and Fabian J. Theis. "Bioinformatics in Psychiatric Genetics." In Psychiatric Genetics, 124–43. Oxford University Press, 2019.
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. SİGEZE, Çiler, and Mehmet Sedat UĞUR. "Analyzing Financial Insufficiency of Households in Turkey with Multivariate Probit Model." Ege Akademik Bakis (Ege Academic Review) 19, no. 4 (October 26, 2019): 401–10.
9. 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.
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.