MARKETING MIX MODELING

Case of an iGaming Company

Lappeenranta–Lahti University of Technology LUT

Master's Programme in Business Analytics

Master’s Thesis

2023

Mujtaba Anwer

Examiner(s): Post-doctoral Researcher Jyrki Savolainen,

Associate Professor Jan Stoklasa

ABSTRACT

Lappeenranta–Lahti University of Technology LUT

LUT School of Business and Management

Master's Programme in Business Analytics

Mujtaba Anwer

Marketing Mix Modelling: Case of an iGaming Company

Master’s thesis

2023

105 pages, 19 figures, 19 tables and 3 appendices

Examiner(s): Post-doctoral Researcher Jyrki Savolainen and Associate Professor Jan Stoklasa

Keywords: Marketing Mix Modelling, Marketing Measurement, MCMC, ANFIS, Regression

This thesis explores cutting-edge marketing mix modelling (MMM) methods and their application to a case-company's data from the iGaming industry. The objective was to quantify the impact of media channels on sales and develop an accurate sales forecasting model using marketing variables. Three methods were implemented: log-differenced OLS regression, Hamiltonian Markov Chain Monte Carlo simulation, and Adaptive Neuro Fuzzy Inference System trained using Particle Swarm Optimization.

OLS regression showed strong goodness of fit but moderate performance with test data, while exhibiting some multicollinearity and deviations from normality. In contrast, MCMC simulations fully converged and estimated the contribution of posterior distributions of media variables, to the target distribution of sales. The ANFIS model provided high forecasting accuracy and validated the rankings of media channels' contributions to sales provided by the other two methods. The research revealed that TV, Online Display, and Affiliates were the primary contributors to the case-company's sales, whereas the saturation points for TV advertising spend were far above other media channels. This confirms what is already known from literature (Fareniuk & Chornous, 2023) that TV tends to have a higher contribution than digital media, due to its wider reach, and a higher saturation point due to the various programmes during which advertisements can be placed to reach different age groups & segments of audience.

Overall, this thesis demonstrates the significance of using multiple MMM methods in conjunction to gain deeper insights and support decision-making in the marketing industry. The findings provide practical implications for allocating capital to media channels and highlight the potential of explainable Ai methods such as ANFIS with PSO for forecasting in MMM.

ACKNOWLEDGEMENTS

I would like to thank my thesis supervisors Jyrki Savolainen and Jan Stoklasa for their feedback and guidance.

I would like to thank the Director of Analytics at the Case-Company for giving me the opportunity to develop this thesis.

I would like to thank all the Marketing and Media Managers at the Case-Company, who spent time explaining to me the marketing mix at the company, the overall business dynamics and extended on-going support for the successful development of this thesis.

I would like to thank all the donors and funders of LUT, who sponsored the scholarship that I received for my studies at LUT.

I would also like to thank the teachers at LUT for all their teachings. In particular, I would like to thank Professor Jan Stoklasa for his remarkable efforts in teaching Financial Econometrics, Professor Kaisu Puumalainen for her teachings on Research Methods & Multivariate Econometrics, Professor Lasse Lensu for answering the many questions I asked during each lecture session on non-parametric Machine Learning methods, Professor Lassi Roininen for his teachings on Bayesian Statistical Parameter Estimation, Post-doctoral Researcher Jyrki Savolainen for his teachings on System Dynamics, Post-doctoral Researcher Heini Vanninen for her teachings on Strategic Issues in Digital Marketing, Professor Anssi Tarkiainen for his teachings on Marketing Analytics, Post-doctoral Researcher Jiri Musto for his teachings on Data Intensive Systems, Professor Mikael Collan for his teachings on Information Systems, Post-doctoral Researcher Shola Oyedeji for his teachings on Requirement Engineering, Post-doctoral Researcher Andrey Saltan for his teachings on Project Management, Professor Kari Smolander for his teachings on Software & System Architecture, Associate Professor Antti Knutas for his teachings on Software Engineering Models, Post-doctoral Researcher Azeem Akbar for his teachings on Software Maintenance, Associate Professor Matylda Jablonska for her teachings on programming, Lecturer Erno Vanhala for his teachings on DevOps, Senior Lecturer Tarja Saarnisto for her teachings on Finnish language and Senior Lecturer Sari Pärssinen for her teachings on Spanish language.

ABBREVIATIONS

Ai	Artificial Intelligence
AiC	Akaike Info Criterion
ANFIS	Adaptive neuro-fuzzy inference system
API	Application Programming Interface
ESS	Effective Sample Size
GRP	Gross Rating Points
HMCMC	Hamiltonian Markov Chain Monte Carlo
HMC	Hamiltonian Markov Chain Monte Carlo
MAPE	Mean Absolute Percentage Error
MCMC	Markov Chain Monte Carlo
Mdip	Media Impressions
Mdsp	Media Spend
MTA	Multi Touch Attribution
MMM	Marketing Mix Modeling
NUTS	No U-Turn Sampler
OLS	Ordinary Least Squares
OLV	Online Video
PPC	Pay Per Click
PS	Paid Social Media
PSO	Particle Swarm Optimization
Q-Q	Quantile Quantile
R^2	R Squared or Coefficient of Determination
Rhat	Gelman-Rubin potential scale reduction factor
RCTs	Randomized Control Trials
RMSE	Root Mean Squared Error
ROAS	Return on Advertising Spend
SEM	Structural Equation Modelling
TRP	Target Rating Points
VAR	Vector AutoRegression
VIF	Variance Inflation Factor

Table of contents

1. Introduction

1.1. Background and Motivation

Marketing Expenditures form the bulk of costs on most companies' balance sheet, often prompting senior leaders in organisations to think whether they would have achieved the same level of sales without allocating financial resources for marketing, or that perhaps a smaller marketing budget for a slightly longer time horizon might have yielded the same results. In essence, the value added by marketing spend is often questioned, which makes it important for marketing data scientists to develop techniques that accurately measure the return on marketing expenditure. This process of using advanced analytical techniques to measure marketing impact with an aim to optimize the spend, is known as 'Marketing Mix Modelling' (MMM).

The challenge with using simple analysis techniques is that accurate data are not always available on some of the key factors impacting the outcomes of marketing campaigns, and customers see a plethora of adverts each day, many of which could be for the same product but through different marketing channels. This makes it difficult for the marketer to identify which marketing channel is driving the sales and allocate budgets with confidence. This study aims to leverage supervised learning techniques such as Regression Analysis, Hamiltonian Markov Chain Monte Carlo Simulations, and an Adaptive Neuro Fuzzy Inference System (ANFIS), to understand marketing channels' contributions to sales, and thus aid in marketing budget allocation decisions.

Another key reason to conduct research on this subject is that simple linear regressions used for MMM often result in multicollinearity, because of the tendency of marketers to start spending on multiple marketing channels at the same time. This makes it difficult for the model to separate incremental sales due to marketing spend from baseline or repeat sales caused by loyal customers, and then to further attribute them to the correct marketing channel. Due to this, more complex techniques have been developed that use multiple statistical models or machine learning algorithms in an attempt to attribute outcome variables accurately to marketing channels and other variables included in the model. The aim of this study is to use a mix of these techniques to provide the marketer with a strong toolkit for understanding the impact of media channels and to forecast scenarios with high accuracy.

The General Data Protection Regulation (GDPR) has limited the ability of marketers to track consumers. With third party cookies also deprecating and tightening of privacy regulations, various marketing measurement methods that rely on user level tracking will be unusable in the future, thus increasing the importance of MMM and underpinning the motivation for this research.

1.2. Focus of this thesis

Majority of the work done in this thesis is based on methods developed by researchers in the past. The focus has been on building multiple approaches to marketing mix modelling based on methodologies used in shortlisted research papers. These include a log-log linear regression and Bayesian Markov Chain Monte Carlo simulations for determining channel wise Adstock, saturation points & contribution to sales.

The novel aspect of this thesis is the application of Particle Swarm Optimization (PSO) algorithm to train a Fuzzy Neural Network, known as Adaptive neuro fuzzy inference system (ANFIS) for forecasting sales using data on channel wise marketing budgets, seasonal variables and control variables. Various researchers have used the ANFIS model trained by PSO for forecasting purposes in different domains (Catalao et al., 2011a), (Catalao et al., 2011b), (Ghasemi et al., 2016), (Kassa et al., 2020), (Hasanipanah et al., 2018), (Band et al., 2019), (Ghordoyee et al., 2021), (Jiang et al., 2015), (Noushabadi et al., 2020), (Mottahedi et al., 2018), (Mahdevari & Khodabakhshi, 2021), (Shamshirband et al., 2019), however, it's usage for marketing mix modelling is not yet documented. Moreover, almost all literature reviewed for this thesis included 'after the fact' variables to build models, be they predictive or aiming to understand the past only. The novelty of the forecast provided by the ANFIS model in this thesis is that it is developed only using 'before the fact' variables, such as media spend, seasonal and control variables, on which data is available before a marketing campaign has been initiated. Moreover, the model developed is tailored to the iGaming industry. So far, no peer reviewed research papers are available that focus on carrying out Marketing Mix Modelling for the iGaming industry.

Furthermore, an explainable AI approach is taken towards the (ANFIS) Neural Network, through the usage of only a single hidden layer for the ANFIS model. This approach was already implemented in a recent highly cited peer reviewed research paper on MMM by (Kumar et al., 2019). Deep Learning Neural Networks usually create new features and data in various layers, adding new exponential features to the model (Goodfellow et al., 2016) that cannot be explained to policy makers or marketing managers. A critical outcome of the usage of a single hidden layer in the ANFIS model is that the neural network does not create any new data or exponential features. Moreover, neural networks can use complex activation functions which render their explain-ability questionable. To resolve this problem, the neural network developed as part of this thesis only uses the Gaussian activation function which is more interpretable since it is based on the normal distribution. Furthermore, the data inputted into the neural network is transformed using Fuzzy logic in order to improve performance through usage of membership values. This approach was already implemented by (Kumar et al., 2019).

1.3. Research Questions

The main problem to solve is to inform the marketer about the impact of channel wise marketing spend on sales, so that the marketer can allocate budgets to more impactful channels. Secondly, the aim is to enable the marketer to simulate budget scenarios based on an accurate forecast. Based on these problems, the research questions are formulated as:

RQ1: What are the state-of-the-art machine learning or statistical analysis based Marketing Mix Modelling methods according to the literature and what are their key application areas?

RQ2: What are the cutting-edge methodologies that are best suited to case-company and provide a high degree of confidence in the results?

The case-company operates in the iGaming industry in North Europe, a region where the industry is highly regulated and only certain forms of marketing are permitted by the regulators, thus imposing constraints on the marketing channels that could be used and even limiting the maximum bet amount per day during lockdowns. On the other hand, the customer behaviour in the industry is also underpinned by risk-seeking and impulsiveness, thus possibly translating into clear signals in the data about marketing mix channels' impact. After performing the literature review, it was deduced that researchers are yet to publish peer reviewed articles exploring industries with such limitations. Hence, to answer the research questions posed in this thesis, approaches taken by various researchers to determine marketing channels' impact will be reviewed, which will then be followed by the implementation of a custom approach best suited to the case-company.

1.4. Structure of this thesis

The thesis begins with introducing the problem and research questions, followed by a theoretical background (Section 2) where key concepts that are needed to understand marketing mix modelling (MMM) in a marketing context are defined. Also defined are the concepts required for understanding the statistical methods used for building MMM models used in this thesis. This is followed by a wide-ranging literature review (Section 3) of methodologies used by various researchers to answer the marketing mix modelling problem and its industry specific applications and challenges. Section 4 describes the data used and provides a comprehensive justification for the methodology adopted to answer the research questions in this thesis. Section 5 presents the exact steps that can be carried out to replicate the three methods implemented in this thesis, while also providing a summary of the results at each step. Section 6 presents the exploratory analysis, results and findings, and their implications for the case-company. Section 7 presents the conclusion and the areas of further research.

2. Theoretical Background

In this section, a theoretical background will be provided based on the literature review, for the reader to develop knowledge about the conceptual phenomenon that underpin marketing mix modelling exercises, both from a marketing and statistical point of view.

2.1.Marketing Mix Modelling

Business revenues are generated through sales of goods or services. These sales are dependent not only on having an addressable market for the product but also on successfully reaching out to its potential consumers in an environment where there are likely competitors reaching out to the same addressable market with an alternative or substitute product.

Various studies (Tariq et al., 2017) have concluded that generating awareness amongst consumers or asking them for a 'call to action' through marketing advertisements, can influence consumers' purchase decisions. Even in Business-to-Business sales, marketing plays an important role through pre-empting or shaping product perceptions, that can make it easier for Business Development professionals to approach new prospects and persuade them to allocate their time to business proposals for the product.

Hence marketers play a vital role in generating revenues for companies by using various types of media to reach out to consumers. These media types traditionally included the likes of outdoor banners, lamp posts, events, newspapers, radio, TV, mail-order, partnership discounts, SMS etc. With rapid adoption of internet & e-commerce, new digital channels became available to marketers for reaching out to potential consumers. Further complexity was added through the various targeting objectives, rule sets and techniques offered by them for reaching potential audiences. These channels included social media platforms (e.g. Facebook, Snapchat etc.), pay per click (e.g. Google Display Network, RTB House etc.), search engine marketing (e.g. Google Search, Bing Search etc.), email, Web Push Notifications, App Push Notifications, Online Video (e.g. YouTube), programmatic platforms that buy digital ad-space in real time across websites and apps (e.g. Adform, Betgenius, BidManager etc.), iMessage for Apple devices, affiliated websites & aggregators, and social media influencers that market products through posts and stories on their social media profile to their unique audience.

In such an environment where the range of channels available is expanding rapidly and the importance of channels is also evolving over time, it is difficult for a marketer to determine the most optimal way to allocate budgets to the various channels. This process is known as (deciding on) the Marketing Mix and to this day, to many marketers, remains an 'art' based on a mix of intuition and judgement.

Marketers tend to have quarterly revenue targets in business to consumer organisations, or 'lead generation' targets in business-to-business organisations, therefore most of the time it is not possible for them to conduct Randomized Control Trials (RCTs) that enable causal inference (Ponzini & Schofield, 2019), since they involve curbs on spending on selective channels to identify the most effective channel. Since curbing marketing spend could result in marketers failing to meet the revenue targets defined in the company's business plan, it poses a risk of adversely impacting their individual performance as well as the company's financial performance and stock price. Due to this, data driven marketers rely on insights from marketing mix models to optimize and reallocate spends.

To aid marketers' understanding of effectiveness of marketing channels, and their decision making about budget allocation, various statistical approaches have been used over the last few decades, especially as more and more data has become available about both online and offline media channels. This process of using advanced statistical methods to determine the effectiveness of marketing channels and aid marketing decisions about budget allocation is known as 'marketing mix modelling' (MMM).

Typically, in an MMM, a multivariate regression is built using time series marketing data, where the target variable is either sales or customer visits and the explanatory variables include the likes of spend on each marketing channel, seasonality, promotions, price and other control variables (Chan & Perry, 2017). The regression results provide weights for each of the variables that help the marketer identify the channel that is more heavily impacting the outcome. The idea being that marketers would be able to generate more sales or leads, if they reallocated budgets from a poorly performing marketing channel to a more impactful marketing channel based on the weightage of the regression coefficients.

2.1.1. Control Variables

In context of MMM, a control variable is a variable that is included in the analysis to account for the impact of external factors or other marketing activities that could influence the relationship between marketing inputs (such as media spend, media impressions, TRPs) and the marketing outcomes (such as sales, revenue etc.). Examples of control variables include factors such as seasonality, weather, sunshine hours, public holidays, sporting events and market competition (Chan & Perry, 2017).

2.1.2. Measures for Measuring Marketing Activity in MMM Models

In order to build advanced statistical models to aid marketers in budget allocation decisions, certain measures related to marketing channels are used as explanatory variables, to quantify the channels' impact on a target variable such as sales. The kind of measures used depends on the nature of the media channel, as well as the aim of the model.

For online media channels, different marketers use different measures for quantifying the channels' impact. One commonly used metric is 'impressions' (Chan & Perry, 2017), defined as the number of times an advertisement is viewed by the target audience, regardless of whether the advertisement was clicked upon or not. While some marketers alternatively use 'clicks' as the representative metric of the channels' performance, using expenditure as a performance measure in MMM models is discouraged (Think with Google, 2022) as it can be a noisy metric because the price charged per click or per thousand impressions by various online marketing platforms varies almost daily, depending on various factors.

For measuring the effectiveness of SMS campaigns, researchers could use the number of SMS sent (Rettie et al., 2005), or SMS opens (for tracked SMS) instead of using SMS expenditure, as explanatory variables in the marketing mix model. This is because research (Drossos & Giaglis, 2005) suggests that SMS effectiveness can depend not only on the quality of the customer database and customers' desire towards the brand, category or product, but also on the timing when the SMS was sent, as customers are more likely to open SMS or email offers received at certain times of week and day. Using SMS opens in the MMM model as a proxy to measure SMS media impact hence isolates the impact of the SMS channel to only in cases where the SMS is opened, while marketers could conduct separate analyses to discover the timings when users are more likely to open an SMS.

For measuring TV advertising, instead of net expenditure on TV campaigns, alternative measures that are widely used include 'Gross Rating Points' (GRPs) (Neijens & Voorveld, 2015) and 'Target Rating Points' (TRPs). GRP and TRP data are collected and sold for a fee, by marketing research companies such as Nielsen in the USA or Kantar in Europe. One GRP is described as being equal to an ad being exposed only once, to 1% of the total possible TV audience, or any household with a television (Cowling et al., 2010). Whereas one TRP is equal to an ad being exposed only once, to 1% of the targeted audience.

2.1.3. Lag Effects

If an advertisement with a promotional offer for a product is shown today to a potential buyer, they might buy the product the following day or a few days after seeing the advertisement. This phenomenon is known 'carry-over' or 'lag effect'. The number of days or weeks that the lag effect lasts, depends upon the media channel and the industry as well as any timelines involved in case special offers are made to customers. For example, it is widely believed that the lag effect of social media platforms is less than a week, whereas for TV, it could be between one week to three weeks and beyond (Broadbent & Fry, 1995). In MMM models, lag effects are accounted for by transforming the data using an 'Adstock' function (see section 2.2.20), which introduces a decay or carry-over co-efficient. The magnitude of the Adstock can be exponential or multiplicative, depending on the marketers' prior beliefs about a media channel.

Some advanced MMM approaches determine the lag effects using Bayesian simulations (Jin et al., 2017), where the researcher's or marketer's prior beliefs are incorporated into the Bayesian 'prior', leaving the model to find the posterior and thus eliminating the bias. Such is the approach used in this thesis to determine lag effects.

2.1.4. Saturation or Shape Effects

It is widely believed by marketers that repeatedly showing an advertisement to a given audience, for example during a certain TV show, can lead to diminishing marginal returns. This is because once the audience is already aware of a given brand, product or offer, then showing them the same advertisement again and again would not lead to additional sales or revenue. Some advanced MMM models such as the one proposed by (Jin et al., 2017) and (Meta Platforms, 2023) use a Hill Function (see section 2.2.21) in order to capture the saturation points for spend on each media channel, beyond which the return on advertising spend is very small or negative, hence not justifying the expenditure. These saturation points or ‘shape effects’ can also be scientifically determined using Bayesian simulations, where the researcher’s or marketer’s prior beliefs are incorporated into the Bayesian ‘prior’, leaving the model to find the posterior and, hence eliminating any possible bias. Such is the approach used in this thesis in order to determine shape effects.

2.1.5. Baseline

It is widely believed by marketers that established brands have loyal customers i.e., those who would purchase the company’s products or services regularly regardless of whether an advertisement is shown to them or not. Hence their purchases would result in a minimum level of sales even if marketing was to be switched off. This minimum level of sales is also known as ‘baseline’ (Cain, 2014).

Traditionally, MMM models using multivariate techniques used the intercept term in the regression as a representation of the baseline. However, more dynamic approaches such as Facebook Prophet (Taylor & Letham, 2017) have been built to model baselines that account for customers switching loyalties repeatedly for e.g., due to offers, or in cases where the advertisements end up generating sales from customers who would have purchased anyway (e.g. existing customers) regardless of the advertising. These dynamic approaches use control variables and seasonal variables, as part of time series decomposition of sales into trend, seasonality, and random components, with the trend component representing the baseline, and the random component representing the incremental sales generated due to spending budget on advertising. Other approaches use data collected by market research agencies on unaided recall and consideration (Cain, 2022) or data from ‘Google Trends’ (Varian & Choi, 2009) for digital native businesses, in order to represent the baseline in the MMM model.

2.1.6. Factors Impacting Marketing Outcomes

Apart from media expenditure or impressions, researchers include various other variables in MMM models, that can impact the outcome of marketing activity. Summarizing from (Chan & Perry, 2017) (Kumar et al., 2019) (Jin et al., 2017) (Koenig-Lewis & Bischoff, 2005) (Chenavaz et al., 2019) (Sun et al., 2017) (Chen et al., 2021) these can include:

- Average weekly Temperature or sunshine hours, as consumers in countries with cold temperatures are likely to spend their leisure time outdoors enjoying the sunlight, as opposed to shopping or responding to advertising.
- Seasonal events such as ‘Black Friday’ or ‘Singles Day’ or ‘World Cup Final’ are included in MMM models as dummy variables, in order to account for the positive sales momentum generated in the run up to these events.
- Promotions and pricing related variables are also carefully integrated through sophisticated feature engineering, or through development of separate models. Since a company could have many products with varying pricing and promotional trends,

accurately capturing the underlying trends in a high-level model such as MMM can be challenging. Some industries such as iGaming are governed by strict regulations that refrain them from advertising offers to consumers, hence integrating promotions related data is not a challenge in such industries.

- Structural breaks such as financial crises or Covid19 lockdowns or a major change in regulations are also accounted for, using dummy variables.
- Indicators that cause a shift in demand curve, such as economic growth or drop in overall industry revenues.
- Product related variables can also be included. In case of a new launch, a dummy variable can be included, whereas specific offerings such as the 'odds' offered to players in iGaming can also be included.
- Website errors or 'downtime' or 'store closure' data can also be included.

Various other variables that impact marketing outcomes can be included in cases where data has been captured on them. However, one key limitation is that MMM models tend to have few rows, allowing for limited degrees of freedom, while they also need to be parsimonious, due to which it is not possible to include every possible variable.

2.1.7. Limitations of Marketing Mix Modelling

Marketing Mix Modelling is a strategic method (Pappa et al., 2022), deployed for modelling aggregate data at a monthly or weekly level and drawing conclusions about the past performance of the media channels. The future is different to the past and considering that data is not available on every possible variable impacting marketing outcome, MMM models are likely to perform poorly in out of sample forecasts even if they do well in testing.

Poor quality data or lack of data on important variables can also make it challenging to fit the model. For e.g., available data on online impressions from tracking tools is not always accurate, whereas for some significant channels such as social media influencers, no data is available apart from the revenue share offered to them, the magnitude of discount on their advertisements and the number of sales brought in by them. Hence collecting and integrating data appropriately in MMM models can be challenging.

Moreover, MMM models do not account for the various parameters used in targeting algorithms in each online marketing campaign, or the TV programmes during which the advertisements were shown to consumers. These unaccounted factors could themselves act as covariate or confounding variables impacting price elasticities through quality of audience attracted, and hence the sales. Hence due to the high-level modelling in MMM, it can be difficult to draw tactical insight without further analytical modelling. For the Digital marketing channels, this could mean the development of Multi Touch Attribution (MTA) models that model customer journeys, and their usage in conjunction with MMM (Witte et al., 2022).

Therefore, to ensure that MMM's insights are valid, researchers recommend validating the results of MMM using Randomized Control Trials (Chan & Perry, 2017) (Chen & Au, 2019), since MMM models capture correlation which does not infer causation, while Randomized Control Trials are a causal method (Ponzini & Schofield, 2019). Hence a widespread practice is to use MMM as one of the many measurement techniques from which results are obtained before making marketing investment decisions.

2.2. Statistical Approaches used in this thesis

For building MMM models, various statistical approaches have been used by researchers over the last few decades. For the approaches adopted as part of this thesis, a theoretical background to the statistical methods that underpin the approaches has been provided in this section.

2.2.1. Ordinary Least Squares Regression

Regression analysis is a technique that allows for the quantification of the strength of the relationship between explanatory variables and a target variable (Bishop, 2006). It involves the fitting of a curve which can be a function of a linear or polynomial form. The formula for linear regression is (Bishop, 2006):

$$y_i = \alpha_0 + \alpha_1 * x_1 + \alpha_2 * x_2 + ... + \alpha_n * x_n \quad \textbf{(1)}$$

Where α_0 is the intercept, α_1 to α_n are the coefficient weights of each variable that have been estimated by the regression, x_1 to x_n are the explanatory variables, and y_i is *i*th value of the target variable y, such as sales data.

Since polynomial regressions can result in overfitting (Hawkins, 2004) in cases where data has high variance or a change in the nature of non-linearity occurs over time, such as in marketing channels' performance, linear regressions remain the preferred method, if it is possible to fit them. In order to capture the non-linearities, log transformations can be applied to the data used in linear regressions, due to log's useful property that transforms exponential relationships into linear relationships (Blum et al., 2018). This leads to the contraction of the distance between the actual and fitted data points, allowing for a better fit. If the data has multiplicative seasonality along with a trend component, transforming the dataset to first differences of logs in the regression model allows for modelling of multiplicative seasonality in the output (Miron & Zeldes, 1988). For small values, first differences of logs can be interpreted as percent changes, which is also a useful property of the log.

It is pertinent to note that the curve fitting process in an OLS regression is achieved using Ordinary Least Squares, the aim of which is to minimize the sum of squared distances between the actual & fitted data points in order to provide a better fitting linear line (Davidson, 2000). For the OLS regressions' results to be deemed valid, they need to not only provide high goodness of fit statistics along with low RMSE in testing, but also need to meet various OLS assumptions or conditions. The first condition being that the error term is normally distributed. The second condition is that residuals have zero mean and constant variance. The third condition is that the residuals should do not suffer from heteroskedasticity, a condition where the variance is not constant and changes with time periods (Gujarati & Porter, 2009). The fourth and fifth conditions are that there should not be any autocorrelation amongst the residuals, and they should not be correlated to the explanatory variables. And the final condition is that there should be no multicollinearity i.e., the explanatory variables should not be linearly dependent on each other.

2.2.2. F-Statistic

The F-statistic (Gujarati & Porter, 2009) represents the ratio of the variability explained by the regression model (explained variance) to the variability that is not explained by the model (unexplained variance). It helps in determining whether the overall model is statistically significant (Gujarati & Porter, 2009).

The null hypothesis in the context of the F-test states that all the coefficients of the variables included in the regression (Gujarati & Porter, 2009), except the intercept, are equal to zero. It means that the model has no predictive power, and that none of the independent variables contribute significantly to the dependent variable. Moreover, the p-value associated with the F-statistic represents the probability of obtaining the observed F-statistic or a more extreme value under the assumption that the null hypothesis is true (Gujarati & Porter, 2009). The formula for F statistic is:

$$F_0 = \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-(k+1))}, \tag{2}$$

Where, SSR_r is the sum of squared residuals of the intercept only model (Gujarati & Porter, 2009), and SSR_{ur} is the sum of squared residuals of a multivariate regression model with multiple variables, n represents the number of rows and k and q represent the number of explanatory variables in the model.

2.2.3. Co-Efficient of Determination (R2) and Adjusted R2

R-squared, also known as the coefficient of determination (Gujarati & Porter, 2009), represents the proportion of the total variation in the dependent variable that is explained by the regression model. It ranges from 0 to 1, where 0 indicates that the model explains none of the variability, and 1 indicates that the model explains all of the variability (Gujarati & Porter, 2009). The formula for R squared is:

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \overline{y})^2} \tag{3}$$

Where, $\hat{y}_i$ is the predicted value of the dependent variable based on the regression model (Gujarati & Porter, 2009), $\bar{y}$ is the mean of the observed values of the dependent variable and y_i is the observed value of the dependent variable.

Adjusted R-squared is a modified version of R-squared (Gujarati & Porter, 2009) that takes into account the number of independent variables in the model and adjusts R-squared for the addition or removal of variables. It penalizes the inclusion of unnecessary variables in the model. The adjusted R-squared is generally lower than the regular R-squared if there are multiple independent variables in the model. It is preferred when comparing models with different numbers of predictors (Gujarati & Porter, 2009) because it takes into account the model's complexity. A higher adjusted R-squared indicates a better fit only if the additional predictors contribute significantly to the model's performance.

2.2.4. Mean Absolute Percentage Error (MAPE)

Data Scientists focused on building predictive models train machine learning algorithms on a percentage of the data, such as 80% of the entire number of rows in the dataset (Kumar et al., 2019), and then test the model on the remaining data to see if the model can predict the remaining observations. To measure the testing accuracy, metrics such as Mean Absolute Percentage Error are used (Bishop, 2006) that take absolute values to calculate the average percentage difference between predicted and actual values (Bishop, 2006). The formula for MAPE is:

$$\mathrm{M} = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right| \tag{4}$$

Where, n is the total number of observations, A_t is the actual value, and F_t is the forecasted value by the machine learning model.

2.2.5. Root Mean Squared Error (RMSE)

Data Scientists focused on building predictive models train machine learning algorithms on a percentage of the data, such as 80% of the entire number of rows in the dataset, and then test the model on the remaining data to see if the model can predict the remaining observations. To measure the testing accuracy (Bishop, 2006), metrics such as Root Mean Squared Error (RMSE) are used. The formula for RMSE is:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - p_i)^2} \tag{5}$$

Where n is the number of rows in the dataset (Bishop, 2006), y_i is the actual value and p_i is the predicted value by the machine learning model (Bishop, 2006).

2.2.6. Jarque Bera Test

The Jarque-Bera test (Enders, 2004) is a statistical test used to assess whether a given dataset has a normal distribution. It checks if the skewness and kurtosis of the data match the values expected under a normal distribution. In a normal distribution, the skewness is expected to be close to zero, and the kurtosis is expected to be close to 3. If the dataset deviates significantly from these expected values, it may suggest that the data is not normally distributed (Enders, 2004).

The Jarque-Bera test computes a test statistic, which is then compared to a critical value from a chi-squared distribution. If the test statistic is greater than the critical value, it indicates that the data significantly deviates from a normal distribution (Enders, 2004). Alternatively, if the p-value for the test is greater than the chosen significance level (commonly 0.05), the null hypothesis that there is no strong evidence to conclude that the residuals significantly deviate from a normal distribution, is rejected (Enders, 2004). The formula for the test is:

$$JB = \frac{n}{6}\left(s^2 + \frac{(k-3)^2}{4}\right), \quad \textbf{(6)}$$

Where n represents the number of rows, s represents the skewness and k represents the kurtosis.

2.2.7. Q-Q plot

A Q-Q plot (Quantile-Quantile plot) is a graphical tool (Bishop, 2006) used to assess whether a dataset follows a normal distribution. The Q-Q plot compares the quantiles of the data to the quantiles of the normal distribution, allowing for the visual inspection of the similarity between the two distributions (Bishop, 2006). If the data closely follows the theoretical distribution, the points on the Q-Q plot will fall approximately along a straight line (Bishop, 2006).

2.2.8. Durban Watson Test

The Durbin-Watson statistic (DW statistic) is a test used to assess the presence of autocorrelation in the residuals of a regression model (Davidson, 2000). Autocorrelation occurs when there is a correlation between the residuals at different points in time (Davidson, 2000), indicating that the errors in the model are not independent of one another. The DW statistic ranges from 0 to 4. A value close to 2 indicates no autocorrelation (Davidson, 2000), a value greater than 2 indicates positive autocorrelation and a value below 2 indicates negative autocorrelation. The formula for the test is:

$$DW = \frac{\sum_{i=1}^{n-1}(r_{i+1} - r_i)^2}{\sum_{i=1}^{n} r_i^2}, \quad \textbf{(7)}$$

Where, residuals are represented by r_i and number of rows in the dataset by n (Davidson, 2000).

2.2.9. Breusch-Pagan Test for Heteroskedasticity

The Breusch-Pagan test is used to detect the presence of heteroskedasticity (Enders, 2004) in the residuals of a regression model. Heteroskedasticity occurs when the variance of the residuals is not constant across all levels of the independent variables. If the test statistic is greater than the critical value, it suggests evidence of heteroskedasticity in the regression model. Alternatively, insignificant p values for the test statistic indicate presence of heteroskedasticity (Enders, 2004).

2.2.10. Breusch-Godfrey Test for Autocorrelation

The Breusch-Godfrey test is commonly used in time series analysis to detect autocorrelation in regression models (Enders, 2004). If the test statistic is smaller than the critical value, there is no significant evidence of autocorrelation, and the assumption of no autocorrelation is satisfied. Alternatively, insignificant p values for the test statistic indicate presence of autocorrelation (Enders, 2004).

2.2.11. Goldfield Quandt Test

Goldfeld-Quandt test is specifically designed for cases where it is suspected that the variance of the residuals is not constant across different segments of the data (Enders, 2004). The null hypothesis is that there is homoskedasticity (constant variance) in the residuals. If the test statistic is significantly different from 1, it suggests evidence of heteroskedasticity, and hence the null hypothesis of homoskedasticity is rejected. Alternatively, insignificant p values for the test statistic indicate presence of heteroskedasticity.

2.2.12. Akaike Info Criterion (AiC)

The Akaike Information Criterion (AIC) is a statistical measure used for model selection (Bishop, 2006), particularly in the context of regression analysis and other parametric models. AIC is based on the principle of balancing model fit and complexity. It penalizes models for having more parameters, aiming to find the model that best explains the data while avoiding overfitting. The lower the AIC value, the better the model is considered in terms of balancing goodness of fit and model complexity (Bishop, 2006). The formula for AIC is:

$$AIC = -2\ln(L) + 2k \tag{8}$$

Where L represents the likelihood and k represents the number of explanatory variables included in the model (Bishop, 2006).

2.2.13. Pearson Correlation

Pearson's correlation coefficient quantifies the linear relationship between two continuous variables (Enders, 2004). It ranges from -1 to +1, with 1 representing perfect positive correlation, 0 representing no linear correlation, and -1 representing perfect negative correlation (Enders, 2004).

2.2.14. Variance Inflation Factor (VIF)

Variance Inflation Factor (VIF) is a statistical measure used to detect multicollinearity in regression analysis (Enders, 2004). VIF is calculated for each explanatory variable in the regression model and measures how much the variance of the estimated coefficient for a given variable is inflated due to multicollinearity. A VIF value of 1 indicates no multicollinearity, a value between 1 and 5 indicates moderate multicollinearity whereas a value of 10 or above indicates very high multicollinearity (Enders, 2004).

2.2.15. Bayesian Inference

Bayesian inference is an approach for inference, that utilizes prior beliefs of the researcher about a particular event or hypothesis and updates those beliefs using observed data to obtain 'posterior' probability (Gelman et al., 2004). The formula for the Bayes theorem is:

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)} \tag{9}$$

Where (Gelman et al., 2004):

- A is the event of interest (e.g., a hypothesis) and B is the observed evidence (e.g., data).
- P(A) is the prior probability of event A before observing any evidence (i.e., our initial belief).
- P(B) is the prior probability of event B before considering any evidence.
- P(A|B) is the posterior probability of event A given that event B has occurred (i.e., the updated belief after observing the evidence).
- P(B|A) is the likelihood of observing evidence B given that event A is true (Gelman et al., 2004).

The prior beliefs are expressed in the form of prior probabilities, which are then combined with the likelihood function, to obtain the posterior distribution (Gelman et al., 2004). The accuracy of the posterior distribution improves as more data or evidence become available and is used to update the prior beliefs. In order to carry out Bayesian Inference, simulation-based methods such as Markov Chain Monte Carlo can also be used, which are discussed below.

2.2.16. Markov Chain

Markov Chain is a statistical model which calculates the probability of a system transitioning from its current state into a new state (Gelman et al., 2004). Markov Chain is underpinned by the Markov property which is also known as 'memory lessness' because it does not take into account past history or how the system arrived at its current state. It calculates probabilities of transitioning into a future state only based on the current state of the system. Markov Chains create a transition matrix in which each element represents the transition probability of each state into another state.

2.2.17. Monte Carlo Simulation

Monte Carlo simulation technique is named after the Monte Carlo casino, which is known for its game of chance offerings and hence exhibiting a random experiment. Monte Carlo simulation creates a large number of samples by selecting random parameters for explanatory variables. This allows the researcher to explore the system behaviour under various conditions (Betancourt, 2017). Such explorations are particularly useful when there is an inherent uncertainty in the system or a lack of comprehensive data about the system's behaviour. Various sampling methods can be used as part of Monte Carlo simulations such as Stochastic Differential equations, Geometric Brownian Motion and for Bayesian inference, Markov Chains can be used.

2.2.18. Hamiltonian Markov Chain with Monte Carlo (HMCMC)

Markov Chain Monte Carlo (MCMC) is a method used in Bayesian inference, that combines the properties of Markov Chain with a Monte Carlo simulation in order to generate samples in the form of a posterior distribution that attempt to estimate the target property distribution (Gelman et al., 2004). It works by constantly adjusting the transition probabilities until the posterior distribution approaches the target distribution. One common MCMC algorithm is the 'Metropolis Hastings' algorithm while another type of MCMC algorithm is the Hamiltonian Monte Carlo.

In high dimensional space, it becomes challenging to compute expectations (Betancourt, 2017). This is because the expectations are computed using Integrals, which look for the area where the integrand or target function has the highest density and values. The challenge is that the areas with high density might not have much volume, and the area with the highest volume would be elsewhere. This could hinder the MCMC from converging. As the dimension space increases, this tension between volume and density grows and the regions where either of them makes significant contributions become narrower, and thus making it harder for them to be spotted in the search space. In such a situation, it makes sense to only take the integrand in the 'neighbourhood' between either extreme of volume and density. This neighbourhood is known as the 'typical set' is shown in Figure 1 below. Markov Chain Monte Carlo algorithms aim to map and explore this typical set using the target distribution. At first when the MCMC begins, it is initialized using some prior knowledge, hence the first few samples are biased.

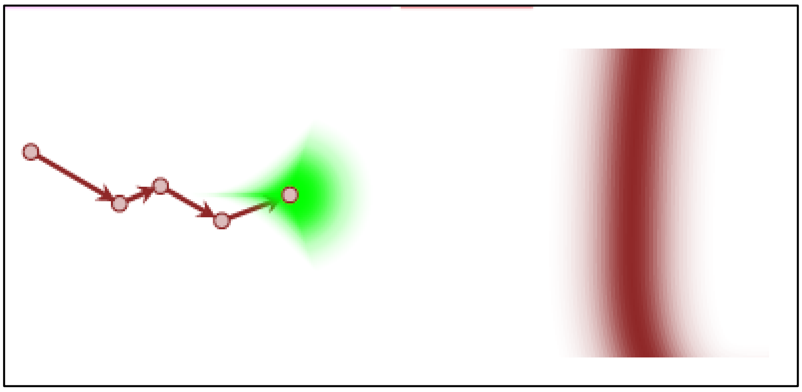

Figure 1. A Markov Chain Monte Carlo moving in the direction of a typical set (thick red line) (adopted from (Betancourt, 2017))

However, as the samples increase, it converges towards the typical set from which expectations can be estimated. The MCMC's movement in the search space is dictated by the Markov Transition, which is a conditional probability density that determines to which next point to jump to, from the current initial point.

The MCMC algorithm works in three steps (Betancourt, 2017). During the first step it finds the typical set, during the second step it maps the typical set and during the third step it explores the same typical set and refines its estimations. Moreover, in order to move in the direction of the target distribution, ‘proposals’ are made using the Metropolis Hastings algorithm, which are ‘corrected’ or rejected in case the next point is too far away from the target distribution.

The challenge with using the Metropolis Hastings MCMC algorithm is that if there is pathological behaviour in the target distribution, such as the one shown in Figure 2 below, there can be regions in the typical set that have strong curvature leading to incomplete exploration or biased MCMC estimations (Betancourt, 2017). They might also lead to the MCMC running for long periods, oscillating around the true expectations, without giving consistent expectations.

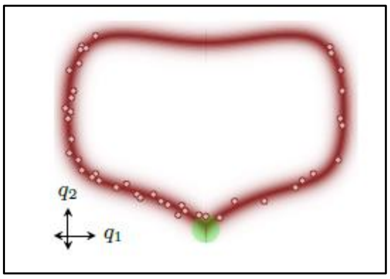

Figure 2. Pathological behaviour in the typical set (green point) where MCMC can get stuck (adopted from (Betancourt, 2017)).

Eventually this might lead to an overcorrection, in the form of a move in another direction, however, it is possible that this might take a long time to happen. Since infinite computational resources cannot be provided for running the MCMC, this process might end without convergence, or the computer might eventually crash.

Moreover, given the steps taken by Metropolis Hastings are too small, therefore they would also require a very long time for the MCMC algorithm to converge, especially in high dimensional space where the typical set is relatively narrow.

In order to overcome this weakness of the Metropolis Hastings (MCMC) algorithm, Hamiltonian Markov Chain Monte Carlo (HMCMC) (HMC) is used, since it can take large jumps away from the initial point in high dimensional space, even in the presence of some correlated variables (Barp et al., 2017)

The HMC method incorporates the concept of Hamiltonian dynamics from physics into the sampling process. In Hamiltonian dynamics, the state of a system is described by its position and momentum variables, and changes in the state are governed by Hamilton's equations of motion. In HMC, the target distribution that is to be sampled from is treated as a potential energy function, and the variables of interest are considered as positions (Hoffman & Gelman, 2011). The momentum variables are introduced as auxiliary variables that allow for efficient exploration of the distribution.

In simpler words, the incorporation of Hamiltonian Dynamics enables the MCMC to convert the one-dimensional parameter space that was being estimated earlier by the Metropolis Hastings, into a two dimensional ‘phase space’ (Barp et al., 2017). This is done by extracting more information from the target distribution using its gradient in order to convert it into a two-dimensional distribution, called the ‘canonical distribution’. Once the HMC maps the typical set in the first step, the phase space allows it to explore its volume and density faster, in large steps, in the second step, reducing the convergence time.

HMC algorithm’s performance is sensitive to the tuning of its degrees of freedom, since it can, under some constructs, engage in defining an unlimited number of transitions (Nishio & Arakawa, 2019). The first degree of freedom that can be tuned is the kinetic energy and the second one being the integration time. The tuning of these degrees of freedom and other parameters is automated, in frameworks such as STAN that use the No U-Turn Sampler (NUTS). NUTS has also been used for the implementation of HMC in this thesis.

2.2.19. Evaluating HMCMC Results

Convergence using R-hat

R-hat, also known as Gelman-Rubin statistic or potential scale reduction factor, is a convergence diagnostic used to assess whether multiple chains in the MCMC process have reached convergence and produced consistent results (Gelman et al., 2004). It compares the variance within each chain to the variance between chains and measures the potential scale reduction if the chains were to run indefinitely. R-hat close to 1 indicates that the chains have converged and produced consistent results (Gelman et al., 2004).

Effective Sample Size

Effective sample size (ESS) is a critical measure used to assess the efficiency (Gelman et al., 2004) and accuracy of the MCMC algorithm in exploring the target distribution and estimating parameters (Gelman et al., 2004). A higher ESS indicates that the MCMC chain is more efficient and produces more independent samples, reducing the uncertainty in the estimates of various statistics and model parameters. Conversely, a lower ESS indicates that the MCMC chain is less efficient, and more samples might be required to achieve accurate estimates.

Tree Depth

Tree depth is a diagnostic parameter used in Hamiltonian MCMC when sampled using the No-U-Turn Sampler (NUTS). It is a measure of how many leapfrog steps are taken during the NUTS simulation for each proposal in the MCMC chain (Gelman et al., 2004). Each leapfrog step is a partial integration of the Hamiltonian dynamics, and the tree depth effectively determines the total number of steps taken to explore the parameter space. Large tree depth in samples indicate that MCMC faced challenges in properly exploring the target distribution and converging.

Divergence

When the MCMC sampler encounters regions in the parameter space with highly curved or complex geometry, as shown in Figure 2, it may exhibit divergent transitions. These divergences occur when the Hamiltonian dynamics simulation becomes numerically unstable, leading to inaccurate or biased samples. Divergent transitions are typically characterized by large Tree Depths (Gelman et al., 2004).

2.2.20. Adstock Function for Lag Effects

Adstock functions are used as part of MMM models to quantify the lag effects described earlier in this section. Series of marketing cost or impressions data is transformed using the Adstock function based on the parameters specified in the function. Researchers (Broadbent & Fry, 1995) have used various Adstock functions based on linear decay, geometric decay and sigmoidal decay. In this thesis, the delayed Adstock function with geometric decay, adapted from (Jin et al., 2017) is used. The property of the delayed Adstock function is that it assumes that advertising impact is not noticeable immediately and that instead it accumulates and decays gradually over time. The equation of the delayed Adstock function is given below:

$$\mathrm{adstock}(x_{t-L+1,m}, \ldots, x_{t,m}; w_m, L) = \frac{\sum_{l=0}^{L-1} w_m(l) x_{t-l,m}}{\sum_{l=0}^{L-1} w_m(l)}, \tag{10}$$

Where L is the maximum level of the lag effect based on prior knowledge and w_m is the weight function. The weight function for the delayed Adstock that accounts for both the geometric decay as well as the peak effect theta is written as:

$$w_m^d(l; \alpha_m, \theta_m) = \alpha_m^{(l-\theta_m)^2}, \quad l = 0, \ldots, L-1, \quad 0 < \alpha_m < 1, \quad 0 \le \theta_m \le L-1, \tag{11}$$

Where α is the retention rate.

2.2.21. Hill Function for Shape Effects

MMM models use saturation functions to model the diminishing returns or shape effect of marketing spend, as described in this section earlier. Examples of saturation functions commonly used in marketing analytics include linear, sigmoidal, exponential and power law functions. In this thesis, Hill function adapted from (Jin et al., 2017) is used. The Hill function was built by biochemists to understand the maximum dose response relationships. It is formulated as:

$$\mathrm{Hill}(x_{t,m}; \mathcal{K}_m, \mathcal{S}_m) = \frac{1}{1 + (x_{t,m}/\mathcal{K}_m)^{-\mathcal{S}_m}}, \quad x_{t,m} \ge 0 \tag{12}$$

where S_m is the hill coefficient or the shape parameter which is also referred to as slope, and $K_m > 0$ is the half saturation point (Jin et al., 2017) – the point at which the system achieves half of its maximum response. The Hill function approaches the value of 1 as the value of x increases. Once the media variables are multiplied with the hill function, the equation can be written as:

$$\beta_m \mathrm{Hill}_m(x_{t,m}) = \beta_m - \frac{\mathcal{K}_m^{\mathcal{S}_m}\beta_m}{x_{t,m}^{\mathcal{S}_m} + \mathcal{K}_m^{\mathcal{S}_m}}. \tag{13}$$

2.2.22. Neural Network with a Single Hidden Layer

Neural networks are supervised learning models that are inspired by the structure and functioning of the human brain (Bishop, 2006). They consist of one or more user defined layers or artificial neurons. Each layer takes input data, assigns weights with an aim to reduce the model error, and produces an output. In order to assign these weights, the layers use an activation function, which is defined by the user. Amongst the commonly used activation functions is the Gaussian activation function, which is based on the normal distribution.

Each layer of the neural network consists of nodes, which represent the features of the data. The connections between nodes are represented by weighted edges, which determine the importance of the information flow between the nodes (Bishop, 2006). During the training process, a neural network learns by adjusting these weights of the connections between nodes. The learning process aims to find the optimal set of weights that minimize the overall error or maximize the network's performance. The final output layer produces the network's prediction or response based on the processed input (Bishop, 2006).

The challenge with using neural networks instead of statistical models is that although they can make better predictions than statistical models, it is not possible to understand how they reached those conclusions since they do not provide meaningful co-efficient weights for the explanatory variables (Nazir et al., 2021). In the context of MMM, this means that neural network based methods cannot be used to determine the most heavily contributing marketing channels.

A neural network with a single hidden layer however can be more interpretable as compared to (deep) neural networks with multiple layers. Since there is only one hidden layer, it ensures that the neural network reaches the given conclusions only based on input data without creating new variables. Moreover, the weights in a single-layer perceptron can provide insight into the relative importance of different input features. Larger weights indicate stronger influences, while smaller weights suggest weaker influences. This allows for some level of feature importance interpretation.

2.2.23. Fuzzy Logic

Although most measurements of intensity or degree are precise, there are certain types of measurements, such as linguistic ones, that can have multiple interpretations causing uncertainty, and hence be imprecise. For example, “very hot weather” in Finland might be between 30 degree Celsius till 36 degree Celsius, whereas only “hot weather” in Finland might be between 26 degree Celsius till 31 degree Celsius. Particularly notable is the overlap between the two verbal measurements, from 30 to 31 degree Celsius.

Fuzzy logic deals with such overlaps, and imprecise statements, by assigning degrees of truth in the form of membership degrees which can be between 0 and 1 (Jang et al., 1997). These membership degrees are calculated using Membership functions, using the pre-defined shape and characteristic of the fuzzy set. These shapes can be triangular, trapezoidal, or even gaussian shaped, similar to a normal distribution (gaussian).

Fuzzy sets represent the degrees of membership in a set, allowing for gradual transitions between different categories or states (Aly & Vrana, 2018). Fuzzy rules allow for the representation of uncertain or subjective knowledge, since they are logical statements that use fuzzy sets and condition-action pairs to model relationships or decision-making criteria. By using fuzzy logic, systems can handle and reason with incomplete or noisy information such as the one found in the marketing variables at times, and hence provide more robust solutions in complex and uncertain environments.

2.2.24. Meta Heuristics

Metaheuristics are optimization algorithms that look for approximate solutions, unlike traditional optimization algorithms that look for exact solutions (Agrawal et al., 2021). They trade off optimality for scalability in order to explore large search spaces in a short amount of time. They make use of heuristics derived from various phenomenon that occurs in nature such as in the animal kingdom, in order to guide the search process towards promising regions of the solution space. The algorithms start with a randomly generated solution that is improved iteratively based on the objective function. Their strength lies in their ability to balance exploration or diversification (volume) and exploitation (density) in order to escape local optima. Some popular metaheuristic algorithms include Genetic Algorithm, Ant Colony Optimization and Particle Swarm Optimization (PSO) (Agrawal et al., 2021).

2.2.25. Particle Swarm Optimization (PSO)

Particle Swarm Optimization is a population-based metaheuristic optimization algorithm (Shami et al., 2022) inspired by the social behaviour of bird flocking or fish schooling. It aims to find the optimal solution by simulating the collective intelligence and cooperation of potential solutions (also called particles) in a multidimensional search space. Each particle flies through the search space and adjusts its position based on its own experience and the experience of its neighbouring particles. The particles are guided by their individual best-known solution (personal best) and the best-known solution among all particles in the population (global best) (Shami et al., 2022).

The algorithm starts with the first step of randomly generating an initial population of particles within the search space (Shami et al., 2022). The parameters used for the generation of particles are their position (for density exploitation) and velocity (for exploration). In the second step, the particle’s current position is evaluated and if it’s the best that it has been so far given its previous positions, it is updated in the following step. The best position is determined based on the objective function. After that, each particle repeatedly adjusts its velocity and position based on its previous velocity, the distance to its personal best position, and the distance to the global best position amongst all the particles in the population (Shami et al., 2022). This adjustment is controlled by acceleration coefficients and random factors and continues for a prespecified number of iterations.

PSO is known for effectively balancing exploration and exploitation by allowing particles to explore the search space through their velocities while being influenced by the best solutions found so far, as in a bird flock searching for food together (Shami et al., 2022). The cooperation and information exchange among particles enable the algorithm to quickly converge towards promising regions of the search space. PSO is particularly suited for continuous optimization problems where gradient information (utilized by the HMC) is unavailable or not easily computable.

(Shami et al., 2022) still find that certain topologies of the standard Particle Swarm Optimizer can still get stuck in local optima, due to which they review a variety of techniques used by other researchers to further improve the performance of the PSO which are out of scope of the topic of this thesis.

2.2.26. Adaptive Neuro Fuzzy Inference System (ANFIS) trained using Particle Swarm Optimization (PSO)

ANFIS stands for Adaptive Neuro-Fuzzy Inference System. It is a hybrid computational model that combines the capability of artificial neural networks to learn from data and the capability of fuzzy logic to reason under uncertainty (Shihab & Pillai, 2018). It allows for the learning of the parameters of fuzzy inference systems automatically from data using neural network techniques.

The model typically involves five steps (Catalao et al., 2011b). In the first step, data is transformed into fuzzy sets using membership functions such as the Gaussian function. The membership functions determine the degree of membership for each data point in different fuzzy sets. In the second step, fuzzy rules and their associated initial parameters are defined (Catalao et al., 2011b). These rules use the fuzzy sets and condition-action pairs to model relationships. In the third step the fuzzy inference process computes the “firing strength” or activation level of each fuzzy rule based on the input data. This is done by evaluating the degree to which the input data matches the fuzzy sets defined in the rule base in the second step (Catalao et al., 2011b). After that, in the fourth step, the data is defuzzied. This is done through combining the outputs of the fuzzy rules to generate a “crisp” or numerical output. It typically involves aggregating the rule activations and calculating a weighted average or centroid of the outputs.

In the fifth and final step, the learning takes place, which is the power of neural networks. The learning algorithm used optimizes the parameters of the ANFIS model that are assigned earlier in the second step, repeatedly, if multiple iterations are to take place.

In (Shihab & Pillai, 2018)'s review paper they review research done in the past that combines neural networks with fuzzy logic. They then highlight that machine learning algorithms' learning ability is greatly improved due to fuzzy logic, enabling them to accurately model non-linear dynamic systems. They mention Tagaki-Sugeno-Kang (TSK) to be the most popular fuzzy inference system, whereas when it comes to fuzzy logic combined with neural networks the most popular model is the ANFIS model. In an ANFIS model, the researcher defines a set of premise parameters, which are modelled in order to get the consequent parameters. To perform this modelling, traditionally, neural networks used gradient descent techniques which had a key weakness that they could be stuck in the local minima. In order to deal with this weakness, researchers have used a variety of different techniques that have been discussed in the paper (Shihab & Pillai, 2018). These techniques include training the neuro fuzzy system using metaheuristic algorithms such as evolutionary algorithms, for e.g. genetic or differential algorithm, or otherwise population based algorithms such as ant colony optimization, or particle swarm optimization (PSO) that aims to mimic bird behaviour during migration, with various studies in the review paper proving that PSO tends to outperform our population based algorithms (Shihab & Pillai, 2018).

In a typical ANFIS model, gradient descent-based methods, such as backpropagation, are used (Jang et al., 1997) to update the parameters based on the difference between the predicted output and the target output. However, the challenge with such methods is that of interpretability as well as their limited ability to find the optimal solution. Due to this, other learning algorithms, such as the Particle Swarm Optimizer (PSO) defined earlier in this chapter are used to train the ANFIS model. The ANFIS model that is trained using PSO, is able to reach better solutions, through the training of multiple ANFIS models for user specified a number of iterations. Each model is built using a different set of parameters for the fuzzy inference system, without changing the shape of the fuzzy set, or adding additional layers in the neural network which can make the model very difficult to interpret (Jang et al., 1997).

(Shami et al., 2022) mention PSO as a preferred population-based algorithm for training ANFIS due to its simpler coding needs, and only 3 controlling parameters that need to be tuned i.e., inertia weight, cognitive weight and social ratio, and because the PSO has the ability to hybridize with other optimization algorithms. This hybridization ability has allowed for the creation of a variety of PSO variants that have been reviewed in the paper (Shami et al., 2022). However, in the simple form, the PSO uses a swarm or multiple concurrent attempts are made in the search space to seek a solution (or particle), and each solution has a current position where it is, and velocity with which it is moving, both of which are defined in multiple directions through the respective vector. The PSO starts randomly and at the end of each iteration, the positions of all solutions (or particles) in the 'swarm' are updated based on the best possible solution (particle) found yet, as well as the current position of the entire swarm. A PSO using STAR topology converges faster (Shami et al., 2022), whereas the PSO using RING topology takes longer to converge since its behaviour is more decentralized as each particle only considers the closest couple of particles as its neighbours to which it converges and updates its position after each iteration.

2.2.27. Evaluating ANFIS Results

Consequent Functions and it’s Coefficients

The coefficients of the consequent function in an ANFIS model represent the output values or consequences associated with each rule in the rule base. Each rule in ANFIS follows an

"IF-THEN" structure, where the antecedent part specifies the conditions for the rule, and the consequent part determines the action or output when the conditions are satisfied (Jang et al., 1997). The consequent function is a linear combination of the input variables' values (Jang et al., 1997), and the coefficients in this function determine the contribution or weight of each input variable in determining the final output of the ANFIS model for a specific rule. The coefficients of the consequent function in ANFIS represent the weights or contributions of the input variables to the rule's output and are learned during the training process to optimize the model's performance (Jang et al., 1997).

Variance of Mean in Consequent Function

The consequent function in ANFIS is modelled as a linear combination of the input variables (Jang et al., 1997), with each rule having its own set of parameters, including the mean and variance of the output consequence. The mean in the consequent function represents the central or average value of the output consequence, (Jang et al., 1997) while the variance represents the uncertainty or fuzziness associated with the output consequence of each rule in the rule base, enabling the model to account for fuzziness and uncertainty in the data. During the learning phase of an ANFIS with a Gaussian activation function, parameters including the mean and variance are adjusted using optimization techniques such as gradient descent or particle swarm optimization, to minimize the error between the model's actual output and the desired output (Jang et al., 1997).

3. Literature Review

This chapter reviews the MMM approaches adopted by various researchers in the past, particularly the ones that involve the use of machine learning or advanced statistics.

Marketing Mix Modelling is a term coined in the 1940s (Michael & John, 2011) that has been explored by researchers over several decades using new approaches. Broadly, the literature can be divided into frequentist techniques, Bayesian techniques and interpretable neural network techniques. Majority of the early literature focused on frequentist regression techniques, however over the last two decades most of the high-quality papers published used a Bayesian simulation technique such as HMCMC or a Bayesian hierarchical model, whereas recently only one highly cited paper was published utilizing a fuzzy neural network.

The databases used for the literature search included Google Scholar, EBSCO Host, LUT Primo, Springer, Science Direct, Research Gate and Sage. The queries used for researching included 'marketing mix modelling', 'marketing mix model', 'marketing mix attribution', 'advertising effectiveness' and 'Adstock'.

More than 100 publications were initially downloaded, including books, peer reviewed research articles, conference papers and articles published by private organisations. These were then shortlisted based on impact factor of the peer reviewed journal or reputation of the organisation, as well as based on the aims of the publication.

3.1. Table of Literature Reviewed

A number of peer reviewed research articles were reviewed as part of the Literature Review exercise. Table 1 below provides a high-level description of the data used, methodology and results reported in each article.

Table 1 – Summary of Literature Reviewed

Year Pub	No.	Document Title	Authors	Journal Name	Journal Rank Quartile (Scimago jr)	Journal Impact Factor (Google or Web of Science)	Focus of the Research
1975	1	**BRANDAID: A Marketing-Mix Model, Part 2: Implementation, Calibration & Case Study**	John D. C. Little, (John, 1975)	Operations Research	Q1	2.708 (2023)	The focus of this research is to create a judgemental model by regularly gathering verbal comments from sales and marketing staff, translating it into numerical numbers, and then calibrating it with a linear regression built using media spend, promo and category data. The results indicate that the model performed well over subsequent 20 months.
2000	2	**An Examination of Selected Marketing Mix Elements and Brand Equity**	Boonghee Yoo, Naveen Donthu, Sungho Lee (Yoo	Journal of the Academy of Marketing Science	Q1	17.99 (5 year Springer)	The authors aimed to model the impact of marketing mix variables such as spending and promotions, on brand equity. Data is collected using a survey of college students & measures developed and validated using Cronbach's alpha before a Structural Equation model is built. Results

			et al., 2000)				have limited generalizability due to the somewhat biased sample.
2004	3	**Response Modeling with Nonrandom Marketing-Mix Variables**	Manchanda, P., Rossi, P. E., & Chintagunta, P. K.	Journal of Marketing Research	Q1	8.63 (5 year)	A negative binomial distribution model is built for the pharmaceutical industry to model the impact of advertising expenditure, free samples, and sales calls by personnel, on the binomial event where the physician actually prescribes the case-company's drug. MCMC is also used for estimation. Results indicate a strong fit, although 'detailing' correlated with baseline.
2011	4	**No Customer Left Behind: A Distribution-Free Bayesian Approach to Accounting for Missing Xs in Marketing Models**	Yi Qian, Hui Xie	Marketing Science	Q1	5.411	The focus of this research paper is to build an MMM model using MCMC for a ketchup company. The author makes a point that MCMC distributions end up generating data for unobserved variables in order to converge to the typical set.
2011	5	**Marketing Mix Modeling for the Tourism Industry: A Best Practices Approach**	Michael J. Wolfe Sr. & John C. Crotts	International Journal of Tourism Sciences	2011	3.791	The focus of this research is to build an MMM model for a naval museum, which belongs to the tourism industry. Data is collected for over three years and a log linear model is built. Results indicate strong fit (96% R squared) and are validated in testing. This paper forms the basis of the first out of the three methods implemented in this thesis.
2014	6	**Media Mix Modeling – A Monte Carlo Simulation Study**	Liu, Yong et al.	Journal of Marketing Analytics	Q2	0.535	The focus of this research is to build an MMM using MCMC model, however the dataset is simulated & very limited as it models sales only as a function of spend and time.
2014	7	**Modeling advertising impact at campaign level: Empirical generalizations relative to long-term advertising profit contribution and its antecedents**	Philippe Aurier & Anne Broz-Giroux	Marketing Letters	Q1	3.426	A frequentist log-log regression is built to model profitability instead of sales. To include only positively contributing campaigns, a Tobit regression is used, while a simple Adstock function is added to the model. Data is collected on multiple TV campaigns from multiple brands. Results indicate a not-so-promising fit. while indicating that campaign carry-over is the biggest driver of sales.
2016	8	**Measuring Multichannel Advertising Response**	Zantedeschi, D., Feit, E.M., & Bradlow, E.T.	Management Science	Q1	7.77 (5 year)	The focus of this article is on building an MMM using MCMC and complex Adstock function. The data collection method is unique where randomized holdout groups of same customers are repeatedly targeted with only email and catalogue adverts over two years.
2016	9	**Cálculo del ROI de marketing en modelos de marketing mix, del ROMI, al valor creado del marketing para los accionistas EVAM**	Méndez Suárez, Mariano; Estevez, Macarena	Universia Business Review	Q4	0.138	The key focus in this research is on building an impulse response function using a vector AutoRegression model in order to give impulses to each media channel for determining the decay rate of Adstock scientifically. Then a linear regression is built which included the adstock parameters, resulting in an 86% R squared, while being trained on 100% of data. Results indicate that short term impact of advertising is negative while it is positive in the long run due to its contribution to brand awareness (baseline).

2017	10	**Bayesian Methods for Media Mix Modeling with Carryover and Shape Effects**	Yuxue Jin, Yueqing Wang, Yunting Sun, David Chan, Jim Koehler	Google Research (not a journal)	N/A	N/A	The focus of this study is to use Hamiltonian MCMC to estimate lag effects, saturation points, and media effectiveness. This article forms the basis of the second method that is implemented in this thesis.
2017	11	**Driving Online and Offline Sales: The Cross-**	Dinner, I. M., Heerde Van, H. J., & Neslin, S. A	Journal of Marketing Research	Q1	8.63 (5 year)	This research aims to quantify the impact of online advertising on the offline channel. Data is used from a high end clothing retailer, and media channels grouped into online and offline. The log-log two stage least squares implemented in this paper suggests that online spending positively impacts offline sales.
2018	12	**Fuzzy expert marketing-mix model**	Aly, Shady & Vrana, Ivan	Agricultural Economics (Czech Republic)	Q2	2.567 (2023)	The focus of this research is to use Fuzzy logic for building a marketing mix model. Marketers' expert knowledge is translated into IF-THEN-ELSE rules in matrices, on a 5-point scale, capturing the intensity of impact of each variable on another variable. Triangular fuzzification is used to fuzzify data, after which the crisp values are obtained. Results are applied to a confidential company however validation has not been provided.
2019	13	**A big data driven framework for demand-driven forecasting with effects of**	Ajay Kumar, Ravi Shankar, Naif Radi Aljohani	Industrial Marketing Management	Q1	8.89 (2023)	The focus of this paper is to create a predictive marketing mix model, that uses data on marketing, seasonality, control, and other variables to provide accurate forecasts. Data is used from a TV retailer for 7 years. The fuzzy neural network beats all other models and provides extremely accurate forecasts in testing and holdout. This paper forms the basis of the third method that is implemented in this thesis.
2021	14	**Hierarchical marketing mix models with sign constraints**	Hao Chen, Minguang Zhang, Lanshan Han & Alvin Lim	Journal of Applied Statistics	Q3	1.013 (2019)	The focus of this study is to build an MMM model using Hamiltonian Markov Chain. To deal with collinearity that can cause negative signs to be assigned by the model to coefficients of correlated variables (that are known to contribute positively), the authors introduce sign constraints in the objective function of the Hamiltonian.
2021	15	**Marketing Mix Modeling Using PLS-SEM, Bootstrapping the Model Coefficients**	Méndez-Suárez M	Mathematics	Q2	2.592 (2021)	The focus of this research is to build an MMM model for an omnichannel consumer electronics retailer, using a modified version of the partial least squares structural equation modelling method. The results indicate negative contributions from online advertising to offline sales.
2022	16	**Modelling Short- and Long-Term Marketing Effects in the Consumer Purchase Journey**	Cain, P.M	International journal of research in marketing	Q1	8.047 (2023)	For short term effects, the author builds a 'path to purchase' model i.e., a 'UCM panel model', which is similar to ARIMA, for decomposing time series. For long term effects, the decomposed time series as well as other variables are used in a vector autoregression and impulses are given to each media variable to determine Adstock parameters scientifically. Results indicated that social media does not drive sales for a car retailer.

2023	17	**Optimization of Media Strategy via Marketing Mix Modeling in Retailing**	Yana Fareniuk, Galyna Chornous	Ekonomika	N/A	0.59 (2021)	The focus of this research is to build an MMM model for a non-food retailer, that mainly uses OOH, TV and Radio for advertising. Simple Adstock function is integrated into a time series multivariate linear regression, resulting in a high fit.

3.2. Synthesis of Literature Review

(Chan & Perry, 2017) mention the objectives of marketing mix modelling (MMM) to be to better understand effectiveness of media spend and based on the understanding, better allocate budgets to the more effective channels. Typically, in an MMM, aggregated weekly or monthly time series data is used to model an outcome variable such as sales or registrations. The explanatory variables typically used, include media spend or impressions, control variables like weather, seasonality and competition (Chan & Perry, 2017), other marketing variables such as promotions, pricing, distribution channels or placement, and events or seasonal demand related variables. However, despite usage of such data, (Chan & Perry, 2017) mention that the results from MMM models cannot be used to infer causality, since machine learning models used to build MMMs typically only measure correlation. Hence, they suggest validating the results of MMM using causal experiments such as Randomized Control Trials.

Over the last few decades, various approaches have been adopted by researchers to measure advertising effectiveness. (John, 1975) used data from a grocery retailer for a period of 2 years, between 1966 and 1968, The retailer reportedly mostly used only one advertising channel. In order to build an MMM model for the retailer, data was sourced on variables such as marketing expenditure, sales and promotions, while data on performance of the product category in the wider market was purchased from Nielsen. The MMM model, called ‘Brandaid’ used both a judgemental response curve modelled by marketing staff of the grocery retailer, as well as a linear regression model. Developing the judgemental response curve involved regularly noting verbal comments from marketing and sales staff about relative performance of the marketing campaigns in a given period, and a numerical estimation by them about the evolution of sales given the market environment, advertising and other factors. Once the regression as well as the judgemental model were developed, the judgemental model was ‘calibrated’ after discussions with marketers, with the linear regression. By calibration, the author refers to modification of the coefficients of the judgemental model, although no formal method of calibration has been provided. The author also does refer to other researchers using nonlinear optimization algorithms to evaluate the calibrated model. The results from the calibrated model were then tracked for more than 20 months. (John, 1975) reported that apart from a couple of instances that were related to a ‘strike’ by workers, the calibrated model performed with high accuracy throughout the 20-month time period.

(Yoo et al., 2000) modelled brand equity using Structural Equation Modelling (SEM), concluding that promotional marketing contributed negatively to brand equity. The data used in the study was collected using a survey of college students. Based on the answers from the survey, various measures were developed which included ‘brand equity’, the validity of which was confirmed using Cronbach's alpha. These measures were then used as inputs to the SEM. The authors justified the survey-based approach to data collection, due to the high correlation that existed amongst their marketing variables. They claimed that the survey-based data collection method together with SEM provided causal results. However, given the small sample size (569) and the age group of college students, the conclusions from the study had limited generalizability.

Industry dynamics could also have an impact on the data collection and modelling method adopted for building the marketing mix model. (Manchanda et al., 2004) built an MMM model for a pharmaceutical company in the USA. A common practice in their industry was reported to be that pharmaceutical company's sales personnel approached doctors to sign them up for prescribing certain drugs to patients, in return for incentives. This activity was claimed to be the single largest marketing activity in the USA pharmaceutical industry (Manchanda et al., 2004) at the time, attracting three times more advertising expenditure than the second most used advertising method. In order to build an MMM model for the pharmaceutical company, the afore-mentioned researchers took a sample of data on the prescription behaviour of 1000 (General) physicians who had received at least two sales calls in the last 24 months, with regards to a medicine X. Data was also sourced on the free samples offered to them, advertising expenditure, as well as the calls made to them by sales personnel. Using the data collected, conditional Negative Binomial Distribution (NBD) and hierarchical NBD models were built at the physician level, with successful prescription of Drug X as the dependent variable. The models were also estimated using (Metropolis Hastings) Markov Chain Monte Carlo algorithms, running for up to 50,000 iterations to ensure convergence. The results of the model were validated using a holdout dataset of 2 months, while the model was trained on 21 months of data. The mean squared errors of the developed models were reported to be 38 and 43 respectively. The models' results indicated that sales personnel tended to approach doctors that had a large patient (customer) base. These same doctors however already contributed significantly to the company's baseline sales, resulting in a correlation between the baseline and the media variables. The study's weakness was identified by the authors to be its lack of data on competitor 'detailing' whereas it also had room for improvement in terms of insight gleaned for the marketer.

To deal with the challenge of lack of data and unobservable variables influencing marketing outcomes, (Qian & Xie, 2011) used a simulation-based Bayesian technique called the Markov Chain Monte Carlo (MCMC). They argued that without resolving for unobservable variables, there would exist a selection bias. They used datasets from a retail store as well as from a ketchup company and discovered that the results of the hybrid (Hamiltonian) MCMC model were better than a Discrete Choice model and a Hierarchical 'Probit' purchase incidence model. They pointed to the hybrid MCMC's ability to generate distributions based on a target posterior, as an estimation of the data on missing variables. Although they used consumer level dataset, they suggested that the MCMC can be applied also for other units of analysis at an aggregate level.

Depending on the nature of the business, and the strength of the signal and variation in data, MMM models that do not adopt complex techniques can also at times accurately model marketing effectiveness. (Michael & John, 2011) built an MMM model for a Naval Museum, whose ticket sales were strongly impacted by seasonality. They used 3 years of data and determined lag effects for media channels through trial and error, with various parameters of the Adstock function. Explanatory variables in their model included promotional data, seasonality, macroeconomic factors, weather, holidays and parking price. The entire dataset was transformed logarithmically to take advantage of multiplicative properties of log. A linear regression was built which achieved a 97% R squared. 6 weeks of data was held-out for validation, where the model achieved a 93% R squared. Results indicated that over 80% of sales were contributed by marketing activity, which was validated by the confirmation from the Naval Museum that most of their customers were new, which the marketing channels helped bring in. Most of the control variables and other marketing factors included in the model, apart from the media variables, demonstrated a negative contribution to sales. Whereas it was noted that Annual, weekly or daily seasonality were not modelled as part of this study.

(Liu et al., 2014) built onto the approach developed by (Qian & Xie, 2011) and applied the MCMC model after including an adstock function in it to model the 'time response' or lag effect. The functional form they chose for the ad stock was based on Gaussian convoluted exponential decay. (Liu et al., 2014) report that the results from their study led to a 60% improvement in revenues, once MCMC models' recommendations were adopted about marketing budget allocation. One particular weakness in this study was that a simulated dataset was used, where sales revenues were only a function of marketing spend and time and did not include any other marketing variables.

(Philippe & Broz-Giroux, 2013) built onto the multivariate linear regression approach and used a special type of OLS regression called 'Tobit', that allows for the partial modelling (or 'censuring') of variables if they exceed or breach a certain threshold. They used Tobit to model profitability instead of sales. Weekly data was used on 31 different brands across 6 different packaged goods categories, whereas budgets were mainly spent on TV advertising, across 254 campaigns during a four-year period. Data was transformed using natural log, to take advantage of the multiplicative property of log, which allows for the capturing of multiplicative relationships. The explanatory variables included five different types of promos, average price, and an Adstock function same as the one used by (Fareniuk & Chornous, 2023). To determine the decay rate, different parameters were repeatedly tested, representing decay rates from week 1 up to 20 weeks. For modelling profitability, which was the aim of their research, the results suggested that campaign carry over was the strongest driver of it, although the model only achieved an R squared of 47%. The use of 'Tobit' and the focus of the study on attempting to model positively contributing campaigns only, based on assumptions made about 'small brands' and 'big brands' using other literature, could be a methodological flaw.

(Zantedeschi et al, 2017) used an MCMC simulation that was initialized using Metropolis Hastings in a Bayesian Hierarchical model. The model was built using customer data and had a unique data collection method. Essentially, the same customers were targeted repeatedly over two years, with email and catalogue advertisements and each time, there were randomly selected holdout groups for each exposure, to capture more variance and avoid collinearity. Their approach was novel however it lacked value because it only used two marketing channels i.e. email and catalogue. Moreover, the usage of customer data for targeting and marketing measurement is becoming unfeasible due to the tightening privacy regulations that only allow usage of data on cohorts as opposed to individual customers. Additionally, no details have been provided by (Zantedeschi et al, 2017) about how they controlled the impact of targeted advertising to the same customers from other mediums (such as mobile apps) over the course of the two years, considering that the customers' emails were being read by the apps installed on their phone to show them more relevant unpaid content or paid advertisements. The results of the study indicated that the purchase incidence for catalogue advertisements shown in absence of emails, was twice as much as that of emails alone. Results indicated that the given formulation of MCMC model alongwith a discrete-time exponentially decaying ad-stock function had a large marginal likelihood and a negligible root mean squared error.

(Méndez-Suárez & Estevez, 2016) built an MMM model to assess return on investment on marketing in the long and short run. They used data from a fast-moving consumer good (FMCG) company. The aim of the research was to calculate the return on marketing investment (ROMI) and the calculation of economic value added of marketing (EVAM) as a metric of shareholder value creation (Méndez-Suárez & Estevez, 2016). The authors proposed a novel approach for estimating the decay rate in the Adstock model using impulse response functions as part of a vector autoregression (VAR) model. VAR models allow researchers to check the response of the system if an impulse is given to one of the variables. Such impulses were given to each media channel one by one, which allowed for the computation of the time it takes for the effect from the impulse, to decay to zero. This decay rate was then used in a simple additive Adstock function for media variables as used in studies done by (Fareniuk & Chornous, 2023) & (Philippe & Broz-Giroux, 2013) discussed earlier in this chapter. Along with other explanatory variables, the Adstock model was included in a linear regression, which achieved an R squared of 86%. Other Explanatory variables included offline and online marketing channels' activity measured using clicks as opposed to spend, brand awareness, price, covariance matrix of residuals of seasonal variables, competitor advertising data, promotional index and a weighted distribution to replace the constant term for measuring the baseline in the regression. The results from the study indicated that the return on advertising in the short term is negative and only becomes positive over the long run. The short-term return was measured using the co-efficient estimates for each media variable, whereas the long-term return was measured using the co-efficient estimate for brand awareness as well as the media variables.

(Cain, 2022) also did a similar study using a vector autoregression model to quantify the long-term impact of marketing on sales, and for dynamically determining the decay rate of Adstock using impulse responses. The model was developed in three steps. In the first step, a 'UCM panel model' was developed, that was similar to the ARIMAX model as it decomposed sales time series into trend (baseline), seasonality, cyclical and random components, while also allowing for other explanatory variables to be modelled. The dataset was converted to logarithm to take advantage of the multiplicative property of log, for building the UCM model. The coefficients from this model were used for quantifying short term advertising impact. For quantifying long term effects, the decomposed time series from UCM, in addition to other variables were used in a vector autoregression. Impulses were given to each media variable to determine Adstock decay parameter dynamically. It is notable that 'mindset metrics' i.e., brand awareness related metrics such as aided & unaided consideration were included in the VAR as a proxy for 'brand building'. In the third step, a cointegrated VAR, which included the cointegration between trend extracted from UCM and brand building metrics, was used to determine long term sales impact. This cointegrating mechanism for brand-building, linked mindset metrics directly to long-term base sales (Cain, 2022). The justification for this three-step approach was that not only does this address the dimensionality and interpretational problems of conventional VAR approaches, but it also incorporates a rigorous theoretical foundation for long-term brand building (Cain, 2022). This approach was applied to a global electronics manufacturer in USA and the results demonstrated accurate forecasts and residual diagnostics. The conclusion was that certain media channels were more geared towards brand building while other media channels were more geared towards short term sales. This study was done for the car industry, where the author found that the social media channels were more useful for word-of-mouth publicity, as opposed to sales. This is likely true for car industry as consumers don't buy cars via social media landing pages, however, for other industries such as apparel retail or gambling for example, social media has been observed to be as a key sales enabler in the short term.

(Mendez-Suarez, 2021) used a modified Partial least squares structural equations modeling (PLS-SEM) method for building an MMM model for an omnichannel Consumer Electronics retailer. The justification for the choice of PLS-SEM method was that its optimization algorithm maximizes the variance explained of the model's endogenous constructs (Mendez-Suarez, 2021), making it suitable to handle endogeneity that could exist in marketing data. However, when data are time series, as in marketing mix modelling, sampling bootstrapping shows inconsistencies (Mendez-Suarez, 2021) due to presence of autocorrelation caused by seasonal events, making the results unreliable. The alternative proposed in this research was the usage of maximum entropy bootstrapping (meboot) (Mendez-Suarez, 2021), a technique specifically designed for time series, which maintains the autocorrelation structure and preserves the occurrence over time of seasonal events or structural changes that occurred in the original series, in the bootstrapped series (Mendez-Suarez, 2021). The results from the study had 92% R squared for online sales, and 79% for store sales. In contrast to the results from omnichannel MMM done by (Dinner et al, 2017), the results from this study showed that for the consumer electronics retailer, online advertisements negatively impacted offline sales. This could be due to online-only promotions offered by the retailer to drive the online channel over the last decade, although no promos related data is included in this study.

Researchers have also used multivariate ordinary least squares regressions for modelling the marketing mix. (Fareniuk & Chornous, 2023) used the data from an omnichannel retailer to model online and offline sales. They decomposed the sales time series into a trend (baseline), additive seasonality, and introduced a simple additive Adstock function that added the decayed media spend from previous time period to current time period spend. How the decay rate was chosen has not been clearly specified, although it seems that multiple decay rates might have been specified, with the best fitting one chosen after trial and error. The resulting model fitted with an R squared exceeding 90%, and key media variables had highly significant likelihood (p) values. It was worth noting that the non-food retailer for whom this MMM model had been built, did not allocate a significant budget to online marketing channels, which are more targeted and have a different dynamic. The media spend was significant only for 'above the line' channels such as TV, radio, and outdoor, whereas data for competitors' spend on these channels had also not been integrated. The less dynamic marketing practices might explain the high fit of the model. The researchers also did not report the error metrics about the actual versus fitted values, and whether the regression met OLS assumptions.

(Dinner et al, 2014) also apply an OLS regression. albeit in a different setting and in two stages. They used data from a high end omnichannel retailer that advertised online as well as offline. The aim was to quantify the cross-channel sales impact i.e., how much of the online sales were driven by offline advertising and how much of the offline sales were driven by online advertising. Due to endogeneity, non-linearity, autocorrelation, competitor advertising effects and a multivariate dependent variable, the authors justify the use of a two-stage linear regression. The dataset was converted to logarithm before building the model in order to allow for the interpretation of coefficients of independent variables as elasticities. The results suggested that online advertising had a significant impact on the offline sales channels. However, the data used in this study was grouped into traditional media, online display and online paid search, hence the study did not provide specific insights about which particular media channels for this retailer were having a cross channel impact (on the offline sales channel), since online advertising channels tend to have their own unique audience. Moreover, the results from the study were not generalizable, since high end clothes are a considerate purchase and it might be that the cross-channel impact was significant more so because customers wanted to try on the high-end (i.e., expensive) clothes before purchasing, hence the usage of the offline sales channel was more often even if an advertisement was seen online.

(Jin et al. 2017) used weekly data from a Shampoo advertiser for 2.5 years, with an aim to model the lag effects, saturation points and media contributions using the Hamiltonian MCMC approach. Four different models were built using two modified adstock functions, a hill function and a reach transformation. The best model was then chosen using Bayesian Information Criterion (BIC), which preferred the most parsimonious model since the estimates from all the models were similar. This is highlighted as the strength of the MCMC's reliability, that regardless of the prior or starting point, the estimates are similar once the model has successfully converged. However, if the dataset is small, or if the dataset lacks signal, then prior distributions can bias the posteriors. It is also worth noting that the control variables (e.g., price, promotion, distribution) used in this study were highly correlated hence each one of them was regressed against price in a univariate regression as a first step. In the second step, only the residuals from the univariate regressions were included as predictors in the model, for the correlated control variables. How the distributions of such residuals that are estimated by MCMC, are interpreted for the marketer is not clear. Correlating variables however are a common problem in MMM models because usually all the marketing channel activity is increased or decreased by marketers at the same time.

To deal with this problem (Chen et al, 2021) also used the Hamiltonian MCMC for building an MMM for a clothing retailer, however, instead of running regressions on correlated variables and including residuals as predictors in HMCMC like (Jin et al., 2017) did, they imposed sign constraints in the objective function in order to deal with multicollinearity. Moreover, (Chen et al, 2021) accounted for the regional trends that can be observed in retail for e.g., TV advertisements might be more effective in one area of the country than the other. They accounted for such patterns by introducing regional sub models, in the form of a

Bayesian Hierarchical model. In the first step, they built three models on a simulated dataset, and evaluated their performance using the mean squared error, which was lowest for the Hamiltonian MCMC. In the second step, they built the HMCMC on a dataset consisting of media, marketing, control and other variables, and evaluated its performance using Conditional R squared, which stood at 78%. Hence the study concluded that a possible way to deal with multicollinearity is through assignment of sign constraints for the media variables in the objective function of the MCMC.

Other researchers have also attempted to use verbal judgements of marketers for building marketing mix models. (Aly & Vrana, 2018) attempted to use fuzzy logic for translating marketers' judgements into a marketing mix model. They created a rule-based decision support system, using a rating matrix which gathered conditional decision rules using IF-THEN-ELSE logic. The dataset that was used included media variables, sales, control and seasonality factors. Marketers were asked to create the rating matrices by giving a rating to each variable's intensity's impact on another variable's intensity, translating their expert knowledge into a rating scale with 5 different intensity values (Low, Medium, High, Very Low, Very High). These responses were then fuzzified using triangular fuzzification to model the marketer's expert knowledge. Each rule's consequents were determined after which they were defuzzified to get the crisp value. These crisp values were then taken as an input for determination of the remaining decision variables. The defuzzification was done using Centre of Largest Area (COA) method, that essentially allows for the usage of the largest membership value to be used for defuzzification. This model allowed for the handling of subjective inputs as well as inexact or imprecise inputs, however no validation of the results was provided, which was identified as a limitation of this study.

(Kumar et al., 2019) built a novel approach to marketing mix modelling, using a neuro fuzzy inference system called ANFIS (Adaptive neuro fuzzy inference system). The novelty of the method was that it not only provided accurate forecasts of sales using variables used in MMM models, but also provided fuzzy parameters for predictors that were used to achieve that forecast. (Kumar et al., 2019) used weekly aggregated data from a TV retailer, for 372 weeks or 7 years and the explanatory variables were divided into five categories by them. The first category being 'demand translation' that took into account the current known demand from a company's order books. The second category being media variables and promotional activities. The third category being 'demand sensing' which referred to control variables such as price, seasonality and sales in previous periods. The fourth category was called 'demand shaping', that used data on orders and shipping and inputs from other MMM models. The fifth category being an exercise as opposed to data, called 'demand orchestration' that referred to what if scenarios and application of advanced decision making after the predictive model was built. Results indicated that the ANFIS model performed better in forecasting than other models such as ARIMA, Random Forest, Support Vector Machine (SVM), Multiple Linear Regression and Artificial Neural Networks, that it was benchmarked against.

The challenge with the implementation of (Kumar et al., 2019)'s proposed model could be that it might only succeed in forecasting for companies with high data maturity, which gather tremendous amount of data at each step of the customer journey and supply chain. Moreover, the approach for feature engineering and selection has not been explicitly shared, and the researchers mention using an automated machine learning tool called 'Data Robot' for this purpose without further explanation. The authors recommend using marketing mix modelling methods in tandem with ANFIS and combining them together in a Multi-Tiered causal analysis (MTCA) that allows for the creation of a nested model linking the push/pull effects of supply chain using a linear regression (Kumar et al., 2019).

3.3. Summary of the Literature Review

From the literature review, it is evident that researchers have used a wide variety of approaches to understand the impact of various media channels, on online and offline sales. The key goals of these approaches included getting an insight into the carry over effect of advertising, in addition to accurately attributing sales to each marketing channel.

Some researchers used simple approaches such as a Linear Regression (John, 1975) (Philippe & Broz-Giroux, 2013) (Dinner et al., 2014) alongwith an adstock transformation (Michael & John, 2011) (Méndez-Suárez & Estevez, 2016) (Fareniuk & Chornous, 2023) using weekly aggregated data on sales, marketing, seasonal and control variables, that resulted in high R squared and provided valid coefficient weights for each media channel.

However, other researchers faced challenges, due to which different approaches were adopted to resolve them. The biggest challenges faced in understanding the impact of media channels on sales, stemmed either from unobserved variables, or from the correlation that often existed amongst the various marketing channels and control variables. To deal with these challenges, researchers leveraged both novel data collection and preprocessing techniques, as well as complex machine learning algorithms.

The novel data collection or preprocessing techniques included surveying consumers (Yoo et al., 2000), customer level datasets (Qian & Xie, 2011) instead of weekly aggregated datasets, surveys of sales and marketing professionals (John, 1975) (Manchanda et al., 2004) (Aly & Vrana, 2018) and inclusion of residuals from univariate regressions run on correlated variables as explanatory variables (Jin et al., 2017).

The complex machine learning algorithms used, included MCMC (Manchanda et al., 2004) (Qian & Xie, 2011) (Jin et al., 2017) (Zantedeschi et al., 2017) (Chen et al, 2021), SEM (Yoo et al., 2000) (Mendez-Suarez, 2021) and VAR (Liu et al., 2014) (Méndez-Suárez & Estevez, 2016) (Cain, 2022). These techniques also resulted in the calculation of adstock in a scientific way, in addition to providing insights into the contribution of media variables to sales, while demonstrating convergence and providing high values for co-efficients of determination.

A novel technique using explainable Ai which provided both coefficient weights for each media channel as well as accurate predictions was also implemented by (Kumar et al., 2019). It involved the development of a Fuzzy neural network with a single hidden layer.

In this thesis, three separate approaches have been adopted, based on literature reviewed above. These include a linear regression based on the study by (Michael & John, 2011), a Hamiltonian Markov Chain Monte Carlo based on the study by (Jin et al., 2017) and an explainable Ai forecasting method based on the study by (Kumar et al., 2019). The dataset used is, aggregated at a weekly level, on marketing, control and seasonal variables, based on the studies which have formed the basis of the approaches adopted in this thesis.

4. Data and Methodology

This section describes the data used and the methodological framework adopted. Literature review was conducted to understand the methodological approaches and data used by various researchers over the last few decades, the learnings from which underpinned the approach to Data and Methodology taken in this thesis.

Based on the literature review, it was understood that while results from survey-based data collection methods were claimed by researchers to be causal (Yoo et al., 2000) (Zantedeschi et al, 2017), it would have taken a significant amount of time and resources to collect responses that were a representative sample to derive valid and generalizable conclusions. Furthermore, the survey-based methods that performed best in the literature (Zantedeschi et al, 2017), collected longitudinal data over prolonged periods, which was not feasible given the limited time available for the development of this thesis. Moreover, with tightening privacy regulations regarding the collection of users' personally identifiable data to improve marketing outcomes, it was decided to adopt the weekly aggregated data approach (Jin et al., 2017) (Kumar et al., 2019) that did not rely on customer surveys or collect personally identifiable information, in order for the solution developed as part of this thesis to be productionized and used repeatedly by the case-company, even in the future.

With regards to the machine learning methods adopted, it was understood from literature that depending on the data, a simple approach such as a Linear Regression, that is highly interpretable, can also provide valid results (Michael & John, 2011) (Fareniuk & Chornous, 2023), hence it formed the basis of the first method implemented in this thesis. Moreover, it was noted that the most scientific and stable results for media contribution and adstock, were provided by HMCMC (Jin et al., 2017) (Chen et al, 2021), hence it was chosen as the second method implemented in this thesis. Study from (Jin et al., 2017) was used as the basis for implementation of HMCMC, because it provided HMCMC priors that were determined to be the best priors in context of marketing measurement, based on widespread research done by them at Google Inc., which can be found in the Appendix section of their study. Their approach also integrated a Hill Function to determine saturation points, unlike other studies reviewed, hence it was chosen.

ANFIS, trained using gradient descent, had proven to be the best explainable forecasting model in context of MMM by (Kumar et al., 2019), hence it formed the basis of the third approach implemented in this thesis. A review paper from (Shihab & Pillai, 2018) however suggested that an ANFIS model trained by PSO was more accurate and efficient than gradient descent, as the gradient descent could get stuck in local minima. Further evidence of the success of ANFIS trained by PSO was seen in various studies (Catalao et al., 2011a) (Ghasemi et al., 2016) (Kassa et al., 2020) (Hasanipanah et al., 2018) (Band et al., 2019) (Ghordoyee et al., 2021) (Jiang et al., 2015) (Noushabadi et al., 2020) (Mottahedi et al., 2018) (Mahdevari & Khodabakhshi, 2021) (Shamshirband et al., 2019), including one related to forecasting price (Catalao et al., 2011b), which is one of the MMM variables (Chan & Perry, 2017). Hence in this thesis, a third method that used an ANFIS model trained by PSO was implemented, which also formed the basis of the novelty of this thesis, since

ANFIS with PSO had not been implemented in context of MMM so far, at the time of the write up of this thesis.

Business understanding, data collection, data understanding and model development processes were carried out according to the Cross Industry Standard Process for Data Mining (Gomez et. al, 2021).

4.1.Data Collection and Pre-Processing

Since the aim of the research was to model sales using marketing mix and other variables, data was collected on variables that were likely to have a strong impact on the case-company's sales. These variables were selected based on knowledge from literature and based on interviews conducted with the marketing managers at the case-company. In total, data on 241 variables was initially sourced, to act as a proxy for measuring various factors impacting the company's sales. The same dataset was also sourced for provision to an external marketing consultancy for the MMM exercise.

The time horizon of the data used was 158 weeks, or 3 years. Data was aggregated at a weekly level in line with what was done by most researchers in the literature (Jin et al., 2017) (Hao Chen, 2021) (Cain, 2022) (Aurier & Broz-Giroux, 2014). The reason for working with weekly data was that daily data can have a low signal to noise ratio, in the marketing context.

In the data pre-processing phase, variables were dropped based on missing values. Moreover, since the case-company is a sports betting company, comprehensive effort was made to map the local and global sporting events that were of interest to the audience in the country for which data was sourced. Public holidays were also included in the model. Sports event calendar and holiday calendar however accounted for close to 70 dummy variables. Including such high number of variables would have sacrificed degrees of freedom and model accuracy as the dataset only had 158 rows. Due to this, dummy variables for sports events and public holidays were summed up into single variables (Sports_Events_Encoded and Calendar_Events_Total respectively) in order to make the model parsimonious.

Hence the total number of variables left was 72. Out of these 72, variables were dropped again based on duplicate measurement of the same phenomenon, that could lead to high correlations and multicollinearity, leaving behind only 27 variables. Table 2 provides data definitions of the 27 variables on which data was used for the development of this thesis.

Table 2 – Data Definitions and Data Sources for Replication of Study

Data Definitions and Data Sources for Replication of Study							
Variable Name	**Represents**	**Unit**	**Definition***	**Source**	**Collection Method**	**MMM Variable Type**	**Data Type**
Affiliate_Average_CPA_PerRegistration	Channel-Affiliate	Euro	Affiliates of the company are offered different deals that are performance based. This is an average of the Cost Per (new customer) registration paid to them.	Internal Company Financial Database	Invoice	Media Variable	Numerical

Calendar_Events_Total	Seasonality-Holidays	Count of Days	This column sums the total number of public holidays taking place in a given week.	Government Statistical Agency - Data on Public Holidays	Web Scraping	Seasonal Variable	Numerical
Calendar_Week_Tot4	Seasonality-WeekMonth	Week Number	This column provides the week number of the month for each row of data (*Customers order more at the beginning of each month after pay-day*)	Created using date column	Python or Excel Formula applied to date	Seasonal Variable	Numerical
Consumer_desire_to_save_in_current_economic_situation	Consumer Confidence (Inverted)	Index	Weekly consumer confidence survey question, quantifying consumers' desire (confidence) to save vs. spend.	Government Statistical Agency	Survey of consumer sentiment	Control Variable	Numerical
Covid19New_deaths	Baseline-Lockdown Impact	Count of Deaths	New deaths caused by Covid19 virus	Government Statistical Agency	Death Records	Control Variable	Numerical
date	Week	Date	Start date of the week used to group by and aggregate weekly observations	Created using date columns from multiple datasets	--	Seasonal Variable	Numerical
Deposits_Count (or Sales)	Target Variable-Sales	Count of Deposits	Total number of deposits that were made by new and existing customers of the company, in order to place a bet.	Internal Company Financial Database	Electronic Order Receipts	Control Variable	Numerical
Emails_Opened Total	Channel-Email	Count of Impressions	Number of emails which were opened by prospects	Provided by Email Sending software being used. E.g., MailChimp	Trackers (piece of code) installed by Email Provider that is triggered as soon as email is opened	Media Variable	Numerical
Emails_SentTotal	Channel-Email	Count of (Emails) Sent	Number of emails sent to prospects regardless of whether they were opened	Provided by Email Sending software being used. E.g., MailChimp	Email Campaign Records from Email service provider	Media Variable	Numerical
Financial_KPIs_TotalDiscount_Cost	Promotional-Discount	Euro	Total Discount Given to all new and existing customers	Internal Company Financial Database	Electronic Order Receipts	Other Marketing	Numerical
Google_Trends Brand1	Baseline-Google Trends	Index	Gives the relative popularity of search terms entered into the Google search engine in a given region. It is considered to be representative of the direct and indirect online searches on Google search engine about the given topic or business	On www.trends.google.com, search for 'Topic' by entering name of the brand for which MMM is being done. Download CSV	Exact Methodology not shared by Google, but it is broadly based on total searches for a 'topic'	Control Variable	Numerical

Net_Spend_euroAffiliate	Channel-Affiliate	Euro	Total expenditure by the brand on the Affiliates marketing channel in Euros	Internal Company Financial Database	Invoice	Media Variable	Numerical
Net_Spend_euroOnline_Display	Channel-Online Display	Euro	Total expenditure by the brand on the Online Display marketing channel in Euros	Internal Company Financial Database	Invoice	Media Variable	Numerical
Net_Spend_euroOnline_OLV	Channel-Online Video	Count of Euro	Total expenditure by the case-company on the Online Video marketing channel in Euros	Internal Company Financial Database	Invoice	Media Variable	Numerical
Net_Spend_euroOnline_PPC_TOTAL	Channel-Pay Per Click	Euro	Total expenditure by the brand on the Pay Per Click marketing channel in Euros	Internal Company Financial Database	Invoice	Media Variable	Numerical
Net_Spend_euroOnline_PS_Total	Channel-Facebook (Paid Social)	Euro	Total expenditure by the brand on the Facebook marketing channel in Euros	Internal Company Financial Database	Invoice	Media Variable	Numerical
Net_Spend_euroRadio	Channel-Radio	Euro	Total expenditure by the brand on the Radio marketing channel in Euros	Internal Company Financial Database	Invoice	Media Variable	Numerical
Net_Spend_euroTV_Total	Channel-TV	Euro	Total expenditure by the case-company on TV advertising in Euros	Internal Company Financial Database	Invoice	Media Variable	Numerical
Online_Display Impressions	Channel-Online Display	Count of Impressions	Number of times an advertisement was seen by users, on websites associated with the programmatic platform, regardless of whether they clicked on it	Programmatic Provider's API e.g. Betgenius API	Bots/Crawlers (piece of code) installed by Marketing Provider that is triggered as soon an impression happens and records it	Media Variable	Numerical
Online_OLVImpressions	Channel-Online Video	Count of Impressions	Number of times a video advertisement was seen by users, regardless of whether they clicked on it	Video website's API e.g. YouTube API		Media Variable	Numerical
Online_PPC_Totalimp	Channel-Pay Per Click	Count of Impressions	Number of times a Google Pay Per Click advertisement was seen by users, regardless of whether they clicked on it	Google API		Media Variable	Numerical
Online_PS_Totalimp	Channel-Facebook (Paid Social)	Count of Impressions	Number of times an advertisement was seen by users on Facebook regardless of whether they clicked on it	Facebook API or Facebook Business Manager		Media Variable	Numerical
Page_viewsDaily_Errors	Baseline-Website Downtime & Errors	Count of Errors	Number of errors due to downtime or other reasons that appeared on various pages of the brand's website in a given week	Google Analytics GA360	Piece of code from Google Analytics that is inserted on every page of the website, which continuously records various data and shares it with Google Analytics back-end database	Control Variable	Numerical

Radio_TRP30_A12_Brand1	Channel-Radio	Count of TRP	TRP is a measure for gauging the amount of audience that saw an advertisement. Calculated by multiplying the percentage of the target audience reached by the radio advertisement (known as the Reach) with the frequency at which the advertisement is aired.	Purchased for a fee from market research agency (KANTAR in Nordics)	Invoice	Media Variable	Integer
Sports_Event_Encoded	Seasonality due to Sport Events	Count of Events	This variable sums the total number of sport events taking place in a given week.	Data on dates of key sport events was fetched from sports news & broadcast media websites	Web Scraping	Seasonal Variable	Numerical
Sum_of_Sunshine_hours	Seasonality due to Weather	Count of Hours	Sum of hours of sunshine in a given week	Government Statistical Agency	Web Scraping	Seasonal Variable	Numerical
TV_TRP30_A12_NEWBrand1	Channel-TV	Count of TRP	TRP is a measure for gauging the amount of audience that saw an advertisement. Calculated by multiplying the percentage of the target audience reached by the TV advertisement (known as the Reach) with the frequency at which the advertisement is aired.	Purchased for a fee from market research agency (KANTAR in Nordics)	Invoice	Media Variable	Numerical

Data was sourced on target variables of interest to the case company that are generally used in marketing mix models. These variables included revenue generated, new customers acquired, new customers who placed a bet, and total deposits made for the purposes of betting by all customers of the company which was agreed to be the target variable for this MMM exercise.

Variables typically used in MMM models are explained in Section 2.1.6 with references to literature. Explanatory variables were divided into media variables, other marketing variables, control variables and seasonal variables (Chan & Perry, 2017). Media variables included media impressions for each marketing channel, sourced from the respective platform's API. Whereas media costs were sourced from internal company financial database. For marketing channels such as TV and Radio where impressions are not captured, data on target rating points (TRP) and media spend was sourced instead, from media agencies.

Since the particular brand that was modelled operated in North Europe, where the populace prefers to spend time outdoors in summers, data on sunshine hours was also sourced.

Based on discussions with marketers, and sales data exploration, it was revealed that Covid19 had an impact on company sales. Due to this data was sourced on various factors measuring Covid19. These included new deaths recorded per week, patients in ICU due to Covid, patients in the hospital due to Covid and weekly new cases.

Since the case-company only offers betting online, and does not have an offline business, hence data was also included on website downtime and errors from Google Analytics. Data was also sourced on economic factors such as the consumer confidence survey published by the government statistical agency.

In order to model the company's baseline sales dynamically, since the company is a digital native and only operates online, data was sourced from Google Trends (Choi & Varian, 2009) to represent baseline sales.

In the country where the case-company operates, strict regulation limits the company from offering most types of promotions in advertisements. Once customers sign up for the first time, they are however offered a 'bonus' or discount on the bets that they place. The aggregation of this discount on a weekly basis was used to measure the impact of promotions.

4.2. Methodology

Based on literature review, it became clear that researchers have used a wide variety of techniques to build marketing mix models in the past, each one of them, having their strengths and weaknesses. While log linear regression models built by (Michael & John, 2011), (Philippe & Broz-Giroux, 2013), (Dinner et al., 2014) and (Fareniuk & Chornous, 2013), provided scientifically valid results for businesses in industries such as tourism, high end clothing and non-food retailing, other researchers used simulation-based models such as MCMC. The use of simulation-based techniques was deemed necessary by the researchers (from the literature review) because of the existence of non-linear relationships between the marketing variables and sales, and to tackle selection bias that could exist due to unobserved variables.

Statistical or machine learning models, only measure correlations and not causation (Chan & Perry, 2017), however, the decisions that they contribute to as part of MMM, require causal inference. Unfortunately, causal methods such as Randomized Control Trials, are time consuming and costly, both in terms execution as well as missed revenues due to running tests as the tests require switching off certain marketing channels. Hence to answer the research questions in this thesis, it was logical to build multiple models using very different approaches and use their findings in conjunction altogether.

In the literature review, the most promising results for understanding marketing effectiveness of various channels as part of marketing mix modelling were noted to be from log linear regression as well as from the Markov Chain Monte Carlo method. Hence these two methods were chosen for application. While the strength of the log linear regression was in providing exact coefficients that can be interpreted as weights or impact of each marketing variable on sales, the strength of the MCMC method was in the scientific determination of carry-over effect of each marketing channel and modelling of their diminishing marginal returns or saturation points.

A third method was also developed based on a highly cited research paper from (Kumar et al., 2019), where they built an ANFIS model, that was recommended to be used for forecasting scenarios in conjunction with other backward looking MMM methods that aim to understand media effectiveness from the past. So far in the literature review, researchers had not explicitly used neural networks such as ANFIS to solve marketing mix problems. Since the results of the study were promising, developing the ANFIS model formed the basis of the third method implemented in this thesis.

The three methodologies together were the most applicable to the case-company, as the company needed reliable insights that were validated by the multiple methods, while also being able to accurately forecast sales using various media budget allocation scenarios.

A summary of the methods adopted in this thesis is give in Table 3 below:

Table 3 – Summary of Methods Adopted in this Thesis

Summary* of Methods Adopted in this Thesis			
Data Collection	Done based on (Chan & Perry, 2017) and 17 articles reviewed in literature (see Table 1). For Data Definitions see Table 2		
Chosen Methods	**Method 1 - Difference Log OLS Regression**	**Method 2 - Hamiltonian Markov Chain Monte Carlo**	**Method 3 - ANFIS with PSO**
Aim	Baseline model to Calculate each Media Channels' Contribution to Sales	Calculate each Media Channels' Lag Effects, Saturation Points and Contribution to Sales	Provide a forecast alongwith explanatory parameters that indicate each Media Channels' Contribution to Sales Forecast
Justification / Strengths	(1) Interpretability and low complexity (2) Widely used in literature	(1) Widely used in literature (2) Scientific approach to determine lag effects and saturation points (3) Bayesian simulation accounting for unobserved variables and non-linearities	(1) High accuracy in forecasting datasets with volatility & unlike deep learning or gradient boosting models, it is not a blackbox as it provides mean & variance estimates for the fuzzy distribution of the explanatory variables used to achieve that forecast (2) Explainable Ai (single hidden layer in the neural network)
Data Preprocessin g	First difference (week on week) of the natural logarithm of the data	Dataset transformed to natural logarithm	Week-on-Week Percentage Change
Feature Selection	Backward feature selection in a stepwise regression	Backward feature selection	Feature selection during ANFIS model training done by PSO particles
Model Training Algorithm	Ordinary Least Squares	No U-Turn Sampler (NUTS)	Particle Swarm Optimization (PSO)

Model Specification	Inclusion of intercept term in initial model to extend quantification of baseline	Define the Prior using STAN code provided in the Appendix in (Jin et al., 2017)	Gaussian Activation Function, single hidden layer for ANFIS
Model Selection	Based on AiC and Tests for OLS Assumptions	Automated using NUTS	Automated selection of Best Learner by PSO
Model Validation	(1) R squared in Training and Testing (2) RMSE, MAPE, Plot of Actual vs Predicted Values	(1) Rhat (Convergence) (2) Effective Sample Size (3) Divergence (4) Tree Depth (5) Reproducibility	(1) R squared in Training and Testing (2) RMSE, MAPE, Plot of Actual vs Predicted Values
Output	(1) Co-efficients for explanatory variables used to quantify media contributions	(1) Adstock parameters, (2) Hill function parameters (3) Posterior Distribution of explanatory variables that quantifies media contributions	(1) Forecast (2) Mean and Variance of the Consequent Function
Weaknesses	(1) Co-efficient estimates for media variables can have high uncertainty or standard errors due to weakness with capturing non-linearity, peaks/ troughs accurately (2) Forecast accuracy	(1) For media variables, a distribution is the output, which is not linearly interpretable as coefficient weight (*hence the ranking approach*) (2) Memoryless-ness i.e. the next state (e.g. forecast) only depends on the current state, hence challenges with out of sample forecasts (3) 3-4 hours of training time, even with high end specification laptop (4) Complex code and dependencies needed to implement it in STAN with NUTS	(1) For media variables, Mean and Variance of the Consequent Function is the output, which is not linearly interpretable as coefficient weight due to presence of fuzzy rules (*hence the ranking approach*)

4.2.1. Method 1 – OLS Log Linear Regression

When building econometric or machine learning models, a recommended practice is to start with a simple model, such as a linear regression (Bishop. 2006). The idea being that if a simple model could provide scientifically valid results and demonstrates a low level of error in testing and validation, then more complex models are not needed (Gujarati & Porter, 2009).

Based on literature review, various studies (Michael & John, 2011), (Philippe & Broz-Giroux, 2013), (Dinner et al., 2014) and (Fareniuk & Chornous, 2013) successfully used log linear regression to build MMM models that had promising results and provided interpretable insights free from multicollinearity.

Since linear regression as one of the most interpretable and easily explained algorithms (Davidson, 2000), it was chosen as the first method to answer the research questions in this thesis.

Linear regression along with its OLS assumptions was earlier introduced in the theoretical framework section of this. It assumes a linear relationship between explanatory variables and dependent variables (Gujarati & Porter, 2009). The goal is to find the line that minimizes the sum of the squared differences between the observed values of the dependent variables and the fitted values by the linear model (Gujarati & Porter, 2009).

In equation (1), ($\alpha_1,\ldots, \alpha_n$) represented the co-efficient weights for each explanatory variable ($x_1,\ldots,x_m$), while α_0 denoted the intercept. The optimal values for ($\alpha_1,\ldots, \alpha_n$) in equation (1) can be determined by minimizing the sum of squared residuals (SSR) in an ordinary least squares (OLS) regression. This can be formulated as:

$$RSS = \sum_{i=1}^{n} (y_i - f(x_i))^2 \tag{15}$$

Where y_i is the i^th^ real value of y, and $f(x_i)$ is the i^th^ predicted values of y, by the OLS regression (Gujarati & Porter, 2009).

The primary challenge in employing linear regression for our study arises from the nature of the target variable. Standard linear regression fails to consider multiplicative seasonality for the response variable while also assuming a linear additive functional form. Due to this, the log-log form of the regression is used, since it allows for leveraging the multiplicative property of the log which transforms multiplicative relationships into additive ones (Gujarati & Porter, 2009). Hence equation (1) can be written in the log-log form as:

$$\ln(y_i) = \gamma + \omega_1 \ln(x_1) + \omega_2 \ln(x_2) + \cdots . + \omega_n \ln(x_n) \tag{16}$$

Where ($\omega_1,\ldots, \omega_n$) represent the co-efficient weights for each explanatory variable, calculated after the log transformation of the series.

Despite using log transformations, explanatory variables can demonstrate heteroskedasticity, which was the case for our study. The presence of a trend component in the target variable could lead to such heteroscedasticity, resulting in the variability of the residuals to change over time. In order to de-trend and remove heteroscedasticity, first differences of the weekly time series were applied to enable the linear modeling of relationships by stabilizing the variance. The log-differenced linear regression equation can then be written as (Gujarati & Porter, 2009):

$$\ln(y_t) - \ln(y_{t-1}) = \varphi + \mu_1(\ln(x_{1t}) - \ln(x_{1t-1})) + \mu_2(\ln(x_{2t}) - \ln(x_{2t-1})) + \cdots . + \mu_n(\ln(x_{nt}) - \ln(x_{nt-1})) \tag{17}$$

With log differencing, interpretation of the results of a linear regression improves, in context of MMM, since the coefficients from a log differenced linear regression represent the percentage change in sales associated with a one-unit change in the explanatory variable (Gujarati & Porter, 2009). Hence equation (10) can also be written as:

$$\Delta y_t = \varphi + \mu_1 \Delta x_{1t} + \mu_2 \Delta x_{2t} + \cdots. + \mu_n \Delta x_{nt} \quad \textbf{(18)}$$

4.2.2. Method 2 – Hamiltonian Markov Chain Monte Carlo Approach with Adstock and Hill Function

Amongst the MMM approaches used in literature, the most commonly used approach in high impact factor journals was based on Markov Chain Monte Carlo (MCMC) method (Jin et al., 2017) (Manchanda et al., 2004) (Qian & Xie, 2011) (Liu et al., 2014) (Zantedeschi et al., 2016) (Chen et al., 2021).

(Manchanda et al., 2004) also highlighted the ability of MCMC models to generate new data during the convergence process that led to not only modelling of variables in higher dimensions and thus avoiding multicollinearity, but also the removal of any possible selection bias that could be caused by including only select variables in the model based on prior beliefs.

Hamiltonian MCMC was explained earlier in detail in the theoretical framework section of this thesis. HMCMC was chosen over other MCMC methods, such as random-walk Metropolis-Hastings, because it explores the target distribution more efficiently by utilizing information from the gradient of the log probability density function (Betancourt, 2017). This allows for longer and more informative proposals, resulting in faster convergence and less correlation between samples. By incorporating these gradient-based techniques, HMC offers an improved and more effective approach for sampling from complex and multi-modal distributions.

(Honkela, 2020) formulates HMCMC by establishing a dynamical system in which $-\log \pi(\theta)$ denotes the potential energy. This system is then simulated using Hamiltonian dynamics. Hamiltonian dynamics have the remarkable property of conserving the energy and volume of the system and when combined with an appropriate (symplectic) numerical solver (Honkela, 2020), this conservation property extends to the numerical solution. As a result, it enables the simulation of exceptionally long trajectories with a notably high acceptance probability.

In HMCMC, each regular "position" variable θ_i is augmented with a corresponding "momentum" variable $r_i \sim N(0, 1)$ (Honkela, 2020). The potential energy of the system, defined by the target (Honkela, 2020), is denoted as:

$$U(\theta) = -\log \pi(\theta) \quad \textbf{(19)}$$

and the momentum variables have a kinetic energy:

$$K(r) = \sum_{i=1}^{d} \left(\frac{1}{2}\right) r_i^2 \quad \textbf{(20)}$$

resulting in the total energy:

$$H(\theta, r) = U(\theta) + K(r) \tag{21}$$

The Hamiltonian dynamics governing the system are thus expressed (Honkela, 2020) (Subbery & Haugland, 2008) as follows:

$$\frac{d\theta_i}{dt} = \frac{\partial H(\theta,r)}{\partial r_i} \tag{22}$$

$$\frac{dr_i}{dt} = \frac{-\partial H(\theta,r)}{\partial \theta_i} \tag{23}$$

Given our choice of H(θ, r), these equations become:

$$\frac{d\theta i}{dt} = r \tag{24}$$

$$\frac{dr_i}{dt} = \frac{-\partial U(\theta)}{\partial \theta_i} = \frac{\partial \log \pi(\theta)}{\partial \theta_i} \tag{25}$$

These Hamiltonian dynamics are implemented using the No U-Turn Sampler (NUTS) as part of the HMCMC simulation. NUTS operates by progressively extending the simulated trajectory until it detects a U-turn, where the path reverses direction on itself (Hoffman & Gelman, 2011). To ensure reversibility, the path is extended both forwards and backwards in virtual time. Unlike a single accept/reject decision, the algorithm samples the next point from a pool of candidates generated along the trajectory. Additionally, the algorithm incorporates a heuristic for tuning its parameters.

The HMCMC approach that was implemented for the case company as part of this thesis was used NUTS as the sampler, based on a widely cited paper from marketing scientists (Jin et al. 2017) at Google, and a peer reviewed journal (Hao Chen et al., 2021) article from marketing scientists at Nielsen.

4.2.3. Method 3 – Fuzzy Neural Network Method (ANFIS with PSO)

The purpose of developing a third method was to build a forecasting model, to forecast sales over a considerable period accurately, using media variables, control variables, seasonal variables and other marketing variables. An ANFIS model with a gaussian membership function and a single hidden layer was chosen for this purpose, due to its success with forecasting sales using marketing mix variables in a study by (Kumar et al., 2019). Moreover, the model was trained using Particle Swarm Optimization (PSO), as various studies (Catalao et al., 2011a) (Catalao et al., 2011b) (Ghasemi et al., 2016) (Kassa et al., 2020) (Hasanipanah et al., 2018) (Band et al., 2019) demonstrated the superior performance of ANFIS trained by PSO in forecasting. The advantage of ANFIS with PSO over deep learning and gradient boosting methods was that it was not a blackbox as it provided mean & variance estimates for the fuzzy distribution of the explanatory variables used to achieve that forecast.

ANFIS and its hybridization with PSO was explained earlier in the Theoretical Framework section of this thesis. An ANFIS model has 5 layers. In Layer 1, every node ‘i’ is represented by a node function (Catalao et al., 2011b) that can be formulated as:

$$O_i^1 = \mu A_i(x), \quad i = 1, 2 \tag{26}$$

Or

$$O_i^1 = \mu B_{i-2}(y), \quad i = 3, 4 \tag{27}$$

Where. x (or y) represent the input to the *i*th node (Catalao et al., 2011b), and A_i (or B_{i-2}) represent the linguistic label associated with this node. Consequently, O^1_i stands for the membership grade of a fuzzy set A (=A1, A2, B1, or B2), indicating the extent to which the input x (or y) fulfils the quantifier A (Catalao et al., 2011b). The membership functions for A and B are specified using gaussian functions (Okokpujie & Sinebe, 2023) which can be formulated as:

$$\frac{-(x-c)^2}{e^{2\sigma^2}} \tag{28}$$

Where the parameter set {c, e, σ}, known as premise parameters, govern the gaussian-shaped function (Okokpujie & Sinebe, 2023). When these parameters are modified, the function's shape changes, leading to different variations of membership functions for linguistic label A_i.

In layer 2, every node Q performs signal multiplication with its incoming signals and subsequently transmits the resulting product. As a consequence, the output of each node represents the firing strength of a rule. This can be formulated (Catalao et al., 2011b) as:

$$O_i^2 = w_i = \mu A_i(x) \mu B_i(y), \quad i = 1, 2 \tag{29}$$

In layer 3, every node N calculates the ratio of the ith rule's firing strength to the sum of all rules' firing strengths (Catalao et al., 2011b). The outputs generated by this layer are referred to as normalized firing strengths. They can be formulated (Catalao et al., 2011b) as:

$$O_i^3 = \bar{w}_i = \frac{w_i}{w_1 + w_2}, \quad i = 1, 2 \tag{30}$$

During layer 4, each node calculates the contribution of the ith rule to the overall output (Catalao et al., 2011b). This calculation involves using the output w_i from layer 3 and a set of parameters {ai, bi, ci}, which are specific to this layer and termed consequent parameters. They can be formulated as (Catalao et al., 2011b):

$$O_i^4 = \bar{w}_i z_i = \bar{w}_i (a_i x + b_i y + c_i), \quad i = 1, 2 \tag{31}$$

Layer 5 consists of a single node P responsible for computing the final output, achieved by summing all incoming signals. Consequently, an adaptive network operates in a manner that is functionally equivalent to a Sugenotype fuzzy inference system (Catalao et al., 2011b).

$$O_i^5 = \sum_i \bar{w}_i z_i = \frac{\sum_i w_i z_i}{\sum_i w_i} \quad \textbf{(32)}$$

In this thesis, ANFIS utilizes the PSO method to adjust the parameters of the membership functions. The PSO technique offers the advantage of being computationally efficient (Shami et al., 2022). It has already been explained earlier in the Theoretical Framework section. To formulate the PSO equations (Catalao et al., 2011b), consider an optimization problem with D variables. Initially, a swarm of N particles is created, and each particle (model / potential solution) is randomly positioned in a D-dimensional hyperspace, representing a candidate solution for the optimization problem. x is used to denote a particle's position (coordinate) and v to represent its flight velocity in the solution space. Each particle's position x is evaluated using a scoring function, determining its fitness value, which reflects how well it solves the problem (Catalao et al., 2011b).

Among the particles, there is a notion of personal best position (P_{best}) (Catalao et al., 2011b), representing the best position a particle has attained so far, and a global best position (G_{best}), which is the best position among all particles in the swarm. Each particle keeps track of its P_{best} and also knows the G_{best} position found by the entire swarm. To reach the global optimal position, the particles iteratively update their positions according to specific rules. The process continues until the best possible solution is found. The velocity and position of each particle are updated using both stochastic and deterministic update rules. These can be formulated (Catalao et al., 2011b) as:

$$v_i(t) = \omega v_i(t-1) + \rho_1(x_{Pbest_i} - x_i(t)) + \rho_2(x_{Gbest} - x_i(t)) \quad \textbf{(33)}$$

$$x_i(t) = x_i(t-1) + v_i(t) \quad \textbf{(34)}$$

Where x represents an inertia weight (Catalao et al., 2011b), and p1, p2 are random variables. These random variables are defined as p1 = j1 * k1 and p2 = j2 * k2, where r1 and r2 are uniformly distributed random numbers in the range (0, 1), and k1 and k2 are positive acceleration constants (Catalao et al., 2011b). This process happens repeatedly until the set number of iterations is completed. Figure 3 below demonstrates the workflow and the repeated optimization of ANFIS parameters by PSO:

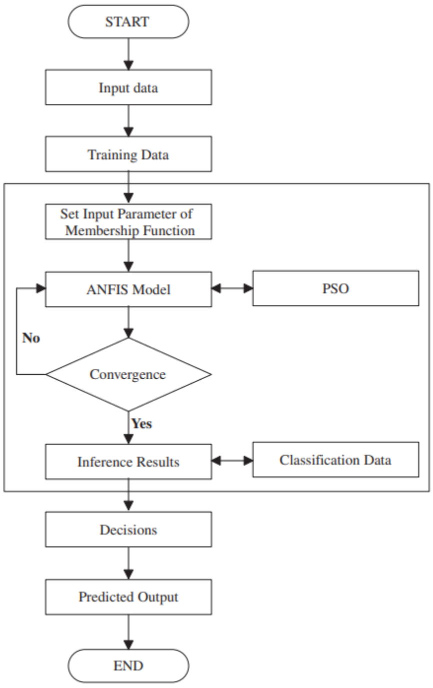

Figure 3: ANFIS PSO Workflow (Adapted from (Band et al., 2019))

4.3. Validation Method

(Chan & Perry, 2017) mention that marketing mix models that use advanced statistical or machine learning techniques, are measuring correlations, whereas the domain in which they are used in i.e., allocating marketing budgets to make a return on the investment, requires causal insights. Hence correlation should not be confused with causation.

Validation of results from Marketing Mix Models can also be challenging since the input data is time series (Jin et al., 2017). Marketing patterns change over time, there can be structural breaks and unobserved variables, all of which deteriorate the model's performance on the test or validation dataset.

The validation strategy for the three different methods used in this thesis for marketing mix modelling was defined based on the nature of the method.

The log differenced OLS regression was first validated through the R squared statistic (Kasuya, 2018) and the F statistic (Gujarati & Porter, 2009). These concepts have been defined in the Theoretical Framework section of this thesis. It was also validated by testing for the six OLS assumptions explained in the theoretical framework section of this thesis. (1) Autocorrelation was tested using Breusch-Godfrey Test, and Durban-Watson statistic – a value close 2 for which indicates that little or no autocorrelation exists (Gujarati & Porter, 2009). (2) Normality of residuals was visualized using Q-Q plot and also tested using the Jarque Beta test. (3) Heteroskedasticity was tested for by using the Breusch-Pagan Test and Goldfeld-Quandt test (Davidson, 2000) (4) Zero mean and constant variance of residuals were investigated using the plot of residuals. (5) Whereas the correlation between explanatory variables (Yamashita et al., 2007) (6) and the correlation between explanatory variables and the residuals, was quantified using the Pearson correlation co-efficient (Benesty et al., 2009).

The linear regression was further validated by training on 80% of data and testing on 20% of data. The model's performance metrics in testing, such as Mean Absolute Percentage Error (MAPE) (Chai & Draxler, 2014) and Root Mean Squared Error (RMSE) were used for measuring performance in testing.

The Hamiltonian Markov Chain Monte Carlo method was validated using four different measures. Firstly, the R-hat statistic, also known as the Gelman-Rubin statistic (Gelman & Rubin, 1992) was used to determine whether the HMCMC had converged and delivered valid results. Secondly, the Effective samples sizes were used as an indicator of chain mixing, absence of autocorrelation and reliability of estimates (Ballnus et al., 2017). Thirdly, Tree Depth was also used for validation of results since a Tree Depth of 10 or beyond indicates issues with the exploration of the parameter space by the Markov Chain (Monnahan et al., 2017). Fourthly, Divergences during HMCMC iterations using NUTS were also used for validity of results. Divergences are specifically associated with Hamiltonian MCMC and the No-U-Turn Sampler (NUTS). They indicate that the sampler was having difficulties in exploring certain regions efficiently (Monnahan et al., 2017) and hence the results might be biased. Therefore, they were also used for validation. For further validation, the HMCMC was run with increased number of iterations, to ensure that the results are stable and do not change with increasing iterations from 50,000 to 75,000.

The ANFIS model, since it is a forecasting model, was validated using MAPE and RMSE. The model was trained on 80% of data and tested on 20% of data (Kumar et al., 2019).

5. Steps for Replicating Methods Implemented

5.1.1. Method 1 – OLS Log Differenced Regression

Table 4 – Detailed Steps for Replicating OLS Regression developed in this thesis

Detailed Steps for Replicating OLS Regression developed in this thesis				
Process	**Description**	**Application of Process in this Thesis**	**Grounding in Literature**	**Python Libraries for Implementation**
Step 1 - Set Target Variable	Discuss with business stakeholders the target variable that they attempt to influence using their marketing campaigns	Deposit Counts, which represents 'total number of sales orders' for the case-company, was agreed to be the target variable	(Chan & Perry, 2017), (Chen et al., 2021), (Kumar et al., 2019)	--
Step 2 - Data Sourcing	Source data based on Literature Review and based on discussions with business stakeholders about factors impacting the targeted marketing outcome in their industry. It is recommended to use weekly data for at least last 3 years or 156 weeks.	Weekly data was sourced for last 3 years on a number of variables as mentioned in Table 2. alongwith their definitions. 'Impressions' were used as a proxy for representing media performance, where applicable, instead of costs.	(Chan & Perry, 2017), (Think with Google, 2022), (Chen et al., 2021), (Kumar et al., 2019), (Jin et al., 2017), (Michael & John, 2011)	Pandas
Step 3 - Descriptive Analysis	Descriptive analysis of the dataset to be carried out by calculating the column wise mean, variance, median, mode, outliers, null values, maximum value, minimum value. Histograms of variables can be plotted to check whether they have any class imbalance or non-normal distribution	There were no null or missing values in the dataset. The dataset also did not include outliers. Although the variance of the media variables and sales were high due to weekly seasonality each month (due to pay day) & sport event driven seasonality. The distributions of variables in the raw dataset were also non normal.	(Ramachandran & Tsokos, 2009)	Pandas (pandas.DataFrame.describe)
Step 4 - Feature Engineering	Transform any categorical variables using One Hot Encoding or another suitable method. Carry out other transformations and feature engineering as necessary. In MMM models, dimensionality reduction is not applicable since the aim is to quantify individual contribution of explanatory variables	There were no categorical variables in this dataset. Sport events were quantified into a single variable representing the number of sport events taking place in a given week. Public holidays were also quantified into a single variable representing their weekly total.	(Zheng & Casari, 2018)	Pandas

Step 5 - Multicollinearity checks	1) Calculate Pearson correlation of each variable against another one. (1a) Note the variables with correlations close to or above 70%. 2) Calculate the Variance Inflation Factor (VIF). Note variables with high VIF values of 5 or above, as they could induce multicollinearity in the OLS regression	1) High correlation of 71% was only observed between TV TRPs and Online Display impressions 2) VIF values for all variables were close to 1 after log differencing the data. They were also below 5 for raw data however they were above 10 once logs were taken	(Christopher et al., 2017), (Edward & Billy, 1982) (Kim, 2019)	(a) Pandas *(pandas.DataFrame.corr)* *(b) Statsmodels (from statsmodels.stats.outliers_influence import variance_inflation_factor)*
Step 6 - Train/Test Split	Split Data into Training and Testing Set.	80% of data was used for training and 20% for testing	(Kumar et al., 2019), (Chan & Perry, 2017), (Xu, 2018), (Medar et al. 2017)	Scikit Learn (*from sklearn.model_selection import train_test_split*)
Step 7 - First Iteration of the Regression with Stepwise Backward feature selection	Run the first iteration of the linear regression with raw data that has not been transformed. Perform sequential backward feature selection and observe the tests in (next) steps below to check if the results improve. *Note that if the model assigns a negative sign to, or suggests to drop, a media channel that is known to contribute positively through other causal studies such as RCTs then such a channel cannot be dropped, and the model needs to be improved in other ways based on findings from tests*	A first iteration of the model was run with raw data. Results did not improve with sequential backward stepwise feature selection that was applied, where variables were systemically dropped based on the following criteria: (1) VIF values higher than 5 (2) Highest p value that is above 0.1 (one by one) (3) Incremental improvement in Adjusted R squared	(Desboulets, 2018), (Ding et al., 2018), (Yamashita et al., 2007), (Jennrich & Sampson, 1968), (Heinze et al., 2018), (Fareniuk& Chornous , 2023), (John et al., 1975)	Statsmodels (*from statsmodels.formula.api import ols*)
Step 7a - Goodness of Fit Test	Investigate the R squared and Adjusted R square statistic for the regression. An R squared above 70% suggests a good enough model	R squared from the initial model was close to 64%. Adjusted R squared demonstrated poorer result of 60%	(Kasuya, 2018), (Colin & Windmeijer, 1997), (Chicco et al., 2021)	Statsmodels, Numpy
Step 7b - Joint Hypothesis	Investigate the F statistic of the model. A p value below 0.05 indicates that variables in the model are together significantly better at modelling the relationship to the target	The p value of the F statistic was significant	(Gujarati & Porter, 2009)	Statsmodels, Numpy

	variable than a model with only an intercept			
Step 7c - Model Accuracy	Investigate the MAPE and RMSE of the model during testing	MAPE was 60% and RMSE 0.39	(Chai & Draxler, 2014), (Chicco et al., 2021)	Statsmodels, Numpy
Step 7d - OLS Assumption 1 - Mean of Residuals	Calculate the Mean of Residuals. OLS assumption is that it should be zero (or close to zero)	The mean of the residuals was nonzero.	(Gujarati & Porter, 2009)	Numpy, Matplotlib
Step 7e - OLS Assumption 2 - Serial Correlation	(1) Investigate the Durban Watson Statistic. A value close to '2' indicates no autocorrelation. (2) Investigate the results of the Breusch-Godfrey Test. P-value above significance level of 0.05 indicates that there is serial correlation in the errors.	(1) Durban Watson Statistic was close to 0.8, indicating presence autocorrelation (2) Breusch-Godfrey Test indicated serial correlation	(Gujarati & Porter, 2009), (Davidson, 2000)	Statsmodels, Numpy
Step 7f - OLS Assumption 3 - Heteroskedasticity	(1) Investigate the results of the Goldfield Quandt Test. (2) Investigate the results of the Breusch-Pagan Test P-value below significance level 0.05 for either of tests indicates that there is heteroscedasticity in the model.	(1) the Goldfeld-Quandt indicated a p value was above 0.1, indicating presence of heteroskedasticity (2) Breusch-Pagan Test indicated presence of heteroskedasticity as p value was above 0.1	(Gujarati & Porter, 2009), (Davidson, 2000)	Statsmodels, Numpy *(Statsmodels.stats.diagnostic.het_goldfeldquandt)//(statsmodels.stats.api.het_breuschpagan)*
Step 7g - OLS Assumption 4 -Normally Distributed Residuals	(1) Investigate the results of the Jarque Bera Test. P-value above significance level 0.05 indicates that residuals are not normally distributed. (2) Investigate the Q-Q Plot. If the points fall along a straight line, then the data set is normally distributed	(1) p value from Jarque Bera test was insignificant (2) Q-Q plot showed repeated divergence from the straight line	(Gujarati & Porter, 2009), (Davidson, 2000), (Kumar, 2005)	Statsmodels, Numpy
Step 7h - OLS Assumption 5 -Explanatory variables are not linearly dependent	Measure (linear) dependence i.e. Pearson correlation amongst explanatory variables	Linear dependence of 71% existed between TV TRP and Online Display Impressions. Since MMM's goal is to understand impact of each media variable, the media variables are not dropped during model iteration.	(Yamashita et al., 2007)	Statsmodels, Numpy
Step 7i - OLS Assumption 6 - no relationship between the	Measure (linear) dependence i.e., Pearson correlation of residuals vs. explanatory variables	There was evidence of relationship between residuals and explanatory variables.	(Benesty et al., 2009)	Statsmodels, Numpy

residuals and each of the explanatory variables				
Step 7j - Model Quality	Note the AiC value for the model	The AiC value for the model was 0.2944	(Ding et al., 2018), (Burnham & Anderson, 2004), (Wagenmakers , 2004)	Statsmodels, Numpy
Step 8 - Diagnosis to guide next steps	Based on diagnosis from tests in previous steps, select appropriate technique(s) to improve the model.	Residuals were non normally distributed. Also, Descriptive analysis revealed that the distributions were non normal, hence log transformation was applied.	(Gujarati & Porter, 2009), (Enders, 2004), (Curran, 2018), (Hongyue et al., 2013), (Keene, 1995), (Davidson, 2000)	
Step 9 - Normalize data	Proceed with normalization of the data. If the data is not normally distributed, log transformations are preferred over min max and z score normalizations. If the data series has strong trend and seasonality components, the data would also have to be differenced	Log transformation is applied to the entire dataset. Natural log is used.		Numpy, Pandas
Step 10 - Repeat step 7 for the log linear model	Repeat previous step 7 and tests from 7a till 7j	Although the R squared improved considerably (85%), the Durban Watson statistic stayed around 0.7, demonstrating autocorrelation	(Dinner et al., 2014), (Philippe & Broz-Giroux, 2013)	
Step 11 - Diagnosis to guide next steps	Based on diagnosis from previous tests, select appropriate technique(s) to improve the model.	After the application of log, the distributions of variables remained non normally distributed, likely due to the strong weekly seasonality and sport event driven seasonality components. There was also evidence of serial correlation.	(Gujarati & Porter, 2009), (Davidson, 2000), (Zanias, 1994), (Myers, 1989), (Zhong-guo, 1997)	
Step 12 - Normalize data	Based on diagnosis from previous tests, proceed with normalization of the data. If the series has a strong trend and seasonality component, taking first differences is preferred to normalize it	First difference (week on week) of the natural logarithm of the data		Numpy, Pandas
Step 13 - Repeat step 7 for the log-	Check Goodness of fit and Test for OLS assumptions	R squared improved to 89% and OLS assumptions were satisfied		

differenced model				
Step 14 - Best Model Selection	Based on Aic, Tests for OLS Assumptions and RMSE & MAPE	All model diagnostics for log differenced model, were better than raw model and log linear model. AiC dropped to -120.3 and tests for OLS assumptions were satisfactory. RMSE was 0.16 and MAPE 134%.	(Ding et al., 2018), (Burnham & Anderson, 2004), (Wagenmakers , 2004)	
Step 15 - Interpret co-efficients of regression to deliver insight	The coefficient indicates the percentage change in response variable (sales), with an additional 1% of an explanatory variable	Coefficients for media channels in descending order	(Gujarati & Porter, 2009), (Davidson, 2000)	

5.1.2. Method 2 – Hamiltonian MCMC

Table 5 – Detailed Steps for Replicating Hamiltonian MCMC developed in this thesis

Detailed Steps for Replicating Hamiltonian MCMC developed in this thesis				
Process	**Description**	**Application of Process in this Thesis**	**Grounding in Literature**	**Python Libraries for Implementation**
Step 1 - Set Target Variable	Discuss with business stakeholders the target variable that attempt to influence using their marketing campaigns	Deposit Counts, which represents 'total number of sales orders' for the case-company, was agreed to be the target variable	(Jin et al., 2017), (Chen et al., 2021), (Chan & Perry, 2017), (Think with Google, 2022)	--
Step 2 - Data Sourcing	Source data based on Literature Review and based on discussions with business stakeholders about factors impacting the targeted outcome in their industry. It is recommended to use weekly data for at least last 3 years or 156 weeks.	Weekly data was sourced for last 3 years on a number of variables as mentioned in Table 2., alongwith their definitions. Due to the structure of this model, 'impressions' were used as a proxy for representing media performance, whereas both 'impressions' and 'cost' were used to calculate the saturation points for each media channel.	(Jin et al., 2017), (Chen et al., 2021), (Chan & Perry, 2017), (Think with Google, 2022)	Pandas

Step 3 - Descriptive Analysis	Descriptive analysis of the dataset to be carried out by calculating the column wise mean, variance, median, mode, outliers, null values, maximum value, minimum value. Histograms of variables can be plotted to check whether they have any class imbalance or non-normal distribution	There were no null or missing values in the dataset. The dataset also did not include outliers. Although the variance of the media variables and sales were high due to weekly seasonality each month (due to pay day) & sport event driven seasonality. The distributions of variables in the raw dataset were also non normal.	(Ramachandran & Tsokos, 2009)	Pandas (pandas.DataFrame.describe)
Step 4 - Feature Engineering	Transform any categorical variables using One Hot Encoding or another suitable method. Carry out other transformations and feature engineering as necessary. In MMM models dimensionality reduction is not applicable since the aim is to quantify individual contribution of explanatory variables	There were no categorical variables in this dataset. Sport events were quantified into a single variable representing the number of sport events taking place in a given week. Public holidays were also quantified into a single variable representing their weekly total.	(Zheng & Casari, 2018)	Pandas
Step 5 - Normalize data	Normalize the data based on the modelling objective and distributions	Descriptive analysis revealed that the distributions were non normal, hence log transformation was applied as in (Jin et al., 2017)	(Jin et al., 2017), (Gujarati & Porter, 2009)	Numpy, Pandas
Step 6 - Model Training Algorithm	No U Turn Sampler (NUTS) would be used for all MCMC simulations performed in this thesis. Implementation of NUTS in STAN programming language is adopted.		(Carpenter et al., 2017), (Hoffman & Gelman, 2011), (Nishio & Arakawa, 2019), (Robert et al., 2018)	pySTAN, Numpy
Step 7 - Model Specification	Define the Priors for the simulation in STAN.	Use STAN code from Appendix in (Jin et al., 2017) to define the priors. These priors are defined in STAN programming language (stored as a string in Python) and passed to pySTAN to sample using NUTS. The priors include: 1) Adstock function transformation and prior belief about maximum no. of days of lag 2) Hill function transformation 3) Expected output parameters	(Jin et al., 2017)	pySTAN, Numpy, Example of implementation of this model is available on Github (He, 2020)

		4) Prior distributions for each output parameter determined based on wide research by Google Inc.		
Step 8 - Model Development	Run the MCMC with a prespecified number of iterations and the number of chains.	No. iterations was set to 50,000 No. of chains was set to 6	(Jin et al., 2017), (Manchanda et al., 2004), (Qian & Xie, 2011), (Liu et al., 2014), (Zantedeschi et al., 2016), (Chen et al., 2021),	pySTAN, Numpy
Step 9- Model Evaluation	(1) Gelman-Rubin potential scale reduction factor Rˆ (Rhat) equals to '1' indicates convergence (2) Effective Sample Size (ESS). Higher ESS indicates better mixing and more reliable estimates. (3) Divergence (4) Tree Depth (5) Reproducibility	(1) Rhat was equal to 1 demonstrating convergence (2) Effective samples sizes were very high, indicating chain mixing, no autocorrelation and reliable estimates (3) Divergence 0.0 of 50000 iterations ended with a divergence (0.0%) (4) Tree Depth: 0 of 50000 iterations saturated the maximum tree depth of 10 (0%) (5) Reproducibility: Running the model with higher number of iterations did not impact the calculated posterior distributions for media variables and saturation points or adstock parameters	(Jin et al., 2017), (Gelman & Rubin, 1992), (Monnahan et al., 2017), (Ballnus et al., 2017), (Barp et al., 2017)	pySTAN, Numpy
Step 10 - Plot the Adstock Charts	The STAN model results include the posterior adstock parameters such as decay rate, and peak.	Given that this is weekly dataset, the peak parameter is multiplied by 7, in order to get the day of the peak. Lag effect during subsequent days is determined using decay rate.	(Jin et al., 2017)	Matplotlib, Seaborn, Numpy
Step 11 - Calculate Saturation Points	The STAN model results include the posterior hill function parameters such as inflection or half saturation point ('ec') (Km), steepness ('slope')(Sm), alongwith the range of the spend values taken by the function (x), in addition to betas (i.e., mean of posterior distribution) of each media variable.	The half saturation point 'ec' is used alongwith betas and real spend values for each media channel, for determination of media saturation points.	(Jin et al., 2017)	Matplotlib, Seaborn, Numpy

Step 12 - Rank the media variables based on mean and variance of the posterior distribution of each variable	MCMC models allow for the calculation posterior distributions of each input variable, that quantify their contribution to the target distribution. Use the lower (2.5%) and upper (97.5%) bound on the uncertainty of each parameter estimate to summarize the posterior distribution of each media variable	Rank the media variables from highest to lowest contribution to target distribution	(Nandy et al., 2021), (Jin et al., 2017), (Zhou et al., 2023)	Numpy, Pandas

5.1.3. Method 3 – ANFIS with PSO

Table 6 – Detailed Steps for Replicating ANFIS model developed in this thesis

Detailed Steps for Replicating ANFIS with PSO forecasting model developed in this thesis				
Process	**Description**	**Application of Process in this Thesis**	**Grounding in Literature**	**Python Libraries for Implementation**
Step 1 - Set Target Variable	Discuss with business stakeholders the target variable that they attempt to influence using their marketing campaigns	Deposit Counts, which represents 'total number of sales orders' for the case-company, was agreed to be the target variable	(Chan & Perry, 2017), (Chen et al., 2021), (Kumar et al., 2019)	--
Step 2 - Data Sourcing	Source data based on Literature Review and based on discussions with business stakeholders about factors impacting the targeted marketing outcome in their industry. It is recommended to use weekly data for at least last 3 years or 156 weeks.	Weekly data was sourced for last 3 years on a number of variables as mentioned in Table 2, alongwith their definitions. 'Cost' was used as a proxy for representing media performance, instead of impressions.	(Chan & Perry, 2017), (Chen et al., 2021), (Kumar et al., 2019), (Jin et al., 2017), (Michael & John, 2011)	Pandas
Step 3 - Descriptive Analysis	Descriptive analysis of the dataset to be carried out by calculating the column wise mean, variance, median, mode, outliers, null values, maximum value, minimum value. Histograms of variables can be plotted to check whether they have any class imbalance or non-normal distribution	There were no null or missing values in the dataset. The dataset also did not include outliers. Although the variance of the media variables and sales were high due to weekly seasonality each month (due to pay day) & sport event driven seasonality. The distributions of variables in the raw dataset were also non normal.	(Ramachandran & Tsokos, 2009)	Pandas (pandas.DataFrame.describe)

Step 4 - Feature Engineering	Transform any categorical variables using One Hot Encoding or another suitable method. Carry out other transformations and feature engineering as necessary. In MMM models dimensionality reduction is not applicable since the aim is to quantify individual contribution of explanatory variables	There were no categorical variables in this dataset. Sport events were quantified into a single variable representing the number of sport events taking place in a given week. Public holidays were also quantified into a single variable representing their weekly total.	(Zheng & Casari, 2018)	Pandas
Step 5 - Normalize data	Normalize the data based on the modelling objective and distributions	Since the aim of the model is to develop a sales forecasting tool for the marketers, all variables were transformed into week-on-week percentage change	(Stevenson & Porter, 2009), (Darzi et al., 2015),	Numpy, Pandas
Step 6 - Train/Test Split	Split Data into Training and Testing Set.	80% of data was used for training and 20% for testing	(Kumar et al., 2019), (Chan & Perry, 2017), (Xu, 2018) (Medar et al. 2017)	Scikit Learn (*from sklearn.model_selection import train_test_split*)
Step 7 - Model Training Algorithm	Particle Swarm Optimization was used to train ANFIS and to determine optimal hyperparameters for ANFIS		(Catalao et al., 2011), (Catalao et al., 2011), (Ghasemi et al., 2016), (Kassa et al., 2020), (Hasanipanah et al., 2018), (Band et al., 2019), (Ghordoyee et al., 2021), (Jiang et al., 2015), (Noushabadi et al., 2020), (Mottahedi et al., 2018), (Mahdevari & Khodabakhshi, 2021), (Shamshirband et al., 2019)	Python code for implementing ANFIS with PSO is available in (Gilardi et al., 2020) Github repo. Python libraries used include numpy, anfis, itertools, matplotlib
Step 8 - Model Specification	ANFIS parameters: 1) Specify the Membership function 2) Specify the number of hidden layers for the Neural Network 3) Specify the 'starting parameters' of the membership function to be optimized by PSO PSO parameters: 1) Number of particles: Specify the number of models (i.e. particles) that would be trained 2) Number of iterations: Specify the number of iterations they would be trained for 3) Cognitive acceleration 4) Social acceleration 5) Initial inertia weight 6) Final inertia weight	ANFIS parameters: 1) Membership function was chosen to be Gaussian 2) Number of hidden layers was chosen to be none 3) The starting point was chosen to be '1' for each feature Specify PSO starting parameters based on best practice (Catalao et al., 2011b): 1) Number of particles 100 2) Number of iterations 10000 3) Cognitive acceleration k1 2.0 4) Social acceleration k2 2.0 5) Initial inertia weight xmin 0.9 6) Final inertia weight xmax 0.4		
Step 9- Model Accuracy	Investigate the MAPE and RMSE of the model. Plot the actual and predicted values.	MAPE was 7% and RMSE 0.10 Actual and predicted values show high accuracy	(Chai & Draxler, 2014), (Chicco et al., 2021)	Statsmodels, Numpy

Step 10 - Rank the media variables based on the best learner parameters found by PSO for ANFIS	The consequent parameters indicate the average weekly percentage change in response variable (sales) and the variance of that change around the mean of the gaussian membership function, if an additional 1% of input is added to an explanatory variable.	Mean & Variance for gaussian membership function distribution of media channels in descending order (Mean, Variance)	(Catalao et al., 2011)	Numpy, Pandas, Anfis

6. Results

6.1.Exploratory Analysis

Exploratory analysis was conducted on the data collected, in order to explore the behaviour of the data, underlying distributions and significant correlations.

A first look at the target variable demonstrated a strong trend and seasonal pattern in the sales data, as demonstrated by Figure 4 below. These seasonal patterns were specifically due to sport events, since customers placed high number of bets during weeks in which sport events were to take place, followed by weeks of lower activity.

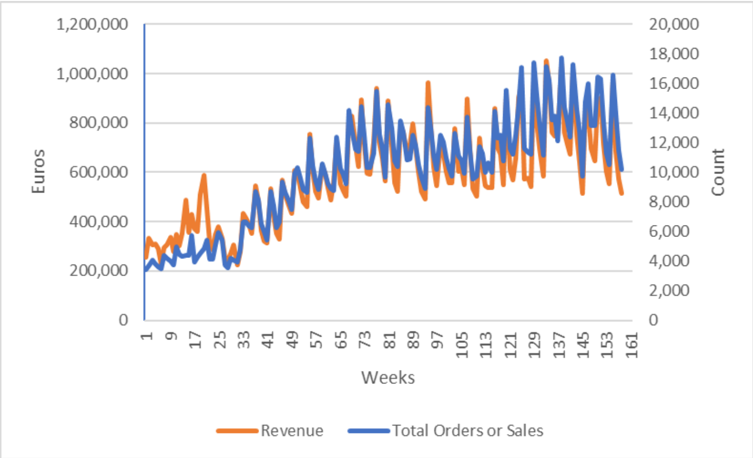

Figure 4: Plot of Sales and Revenue

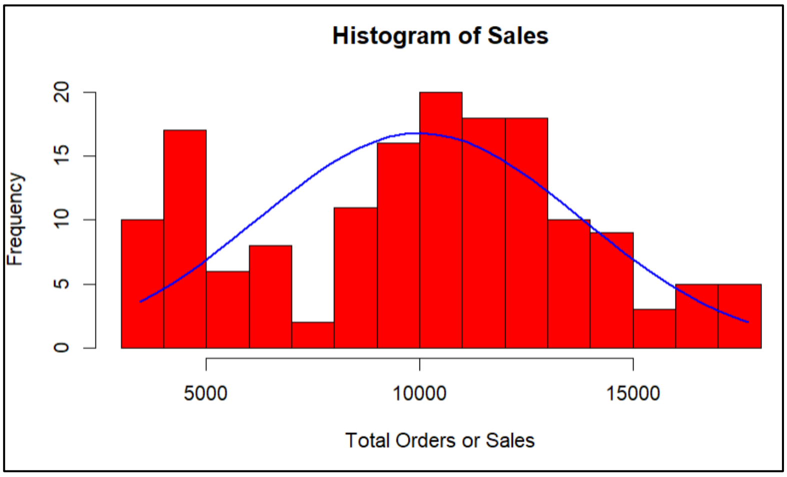

Figure 5: Plot of the distribution of sales variable

The plot of the distribution of the sales variable (in Figure 5) demonstrated a mean of around 10,000 and very fat tails demonstrating volatility in the sales values. This was also confirmed by the descriptive analysis presented in Table 7 below:

Table 7 – Descriptive Statistics of Media Variables and Sales (Note: Data definitions and collection method of the variables in Table 7 above have been provided in Table 2)

Variable Name	count	mean	std	min	25%	50%	75%	max
Sales (Deposits_Count)	158	10,002	3,753	3,432	6,663	10,423	12,542	17,711
Revenue	158	585,201	197,129	223,086	454,373	571,419	715,647	1,053,311
Net_Spend_euroTV_Total	158	33,920	30,707	0	10,738	28,334	47,090	122,822
Net_Spend_euroRadio	6	18,611	10,146	6,526	11,002	17,362	27,849	30,187
Net_Spend_euroOnline_Display	158	2,225	1,346	553	1,226	1,704	3,268	7,779
Net_Spend_euroOnline_PS_Total	158	2,125	2,236	0	539	1,524	2,911	9,837
Net_Spend_euroOnline_OLV	34	2,651	1,672	0	1,360	3,350	3,859	4,801
Net_Spend_euroOnline_PPC_TOTAL	158	3,985	2,761	96	2,073	3,344	6,041	13,081
Net_Spend_euroAffiliate	158	6,570	2,854	2,649	4,644	6,587	7,395	15,535
Total_EmailsSent	158	3,912	3,413	0	1,542	3,068	5,842	21,253
TV_TRP30_A12_NEWBrand1	158	64	58	0	20	51	89	232
Radio_TRP30_A12_Brand1	16	20	20	3	10	14	15	73
Online_DisplayImpressions	158	1,143,725	656,398	315,326	676,846	876,193	1,412,927	3,574,630
Online_PS_Totalimp	158	279,202	388,570	0	75,983	176,066	285,911	2,088,965
Online_OLVImpressions	25	174,618	69,438	87,935	135,800	172,675	211,954	410,541
Online_PPC_Totalimp	158	96,361	69,784	1,374	17,302	100,975	159,103	228,231
Emails_OpenedTotal	158	1,307	1,125	0	558	1,026	1,763	6,375

From Table 7 it can be observed that the case company spent budget on Radio advertising only for 6 weeks out of 158 weeks for which data was sourced. The mean value of the spend was also considerably high for Radio, compared to other channels, demonstrating that the spend was only very high for a few weeks, after which the channel was almost never used. Also notable was that the highest average weekly spend was on TV. In fact, the company spent two thirds of the budget on TV, during the past 3 years.

Table 8 – Share of Media Expenditure (Note: Data definitions and collection method of the variables in Table 8 have been provided in Table 2)

Media Variable Name	Total Spend	Rank	% of Total	% of Total Excluding TV	% of Total Excluding TV and Affiliates
mdsp_TVTRP30A12NEWBrand1	5,359,386	1	67.70%		
mdsp_RadioTRP30A12Brand1	111,663	6	1.41%	4.37%	7.35%
mdsp_OnlineDisplay	351,552	4	4.44%	13.75%	23.15%
mdsp_OnlinePSTotal	335,699	5	4.24%	13.13%	22.10%
mdsp_OnlineOLV	90,141	7	1.14%	3.53%	5.94%
mdsp_OnlinePPCTotal	629,564	3	7.95%	24.62%	41.45%
mdsp_NetSpendeuroAffiliate	1,038,024	2	13.11%	40.60%	
mdsp_EmailsOpenedTotal	100	8	0.00%		
Total Spend During Entire Period	7,916,129				

From Table 8, it can be observed that Email was the most cost-efficient channel since the marginal cost of sending an additional email is close to zero. Excluding TV and Affiliates, the largest expenditure was on the Pay Per Click (PPC) channel, however, the largest number of impressions were instead generated by Online Display channel, as demonstrated by Table 9. Of course, spending more on a channel, or having more impressions from it, does not necessarily mean that the channel brings in more sales. However, these numbers demonstrate the need for scaling the data before modelling.

Table 9 – Share of Media and TRP

Media Variable Name	Total Impressions or TRP	Rank
mdip_TVTRP30A12NEWBrand1	10,114	
mdip_RadioTRP30A12Brand1	320	
mdip_OnlineDisplay	180,708,519	1
mdip_OnlinePSTotal	44,113,848	2
mdip_OnlineOLV	4,365,455	4
mdip_OnlinePPCTotal	15,225,088	3
mdip_EmailsOpenedTotal	206,449	

Table 10 given below provides descriptive statistics for the control variables and seasonal variables used for modelling. Out of 158 weeks for which data was sourced, 138 weeks had sports events, specific to which campaigns were developed by the case-company to take advantage of the seasonal patterns. Whereas there were public holidays or seasonal festivals during 65 out of the 158 weeks for which data was sourced.

Table 10 – Descriptive Statistics for Control Variables (Note: Data definitions and collection method of the variables in Table 10 above have been provided in Table 2)

Variable Name	count	mean	std	min	25%	50%	75%	max
Calendar_Events_Total	158	2	3	0	0	0	2	10
Calendar_Week_Tot4	158	3	1	1	2	3	4	4
Consumer_desire_to_save_in_current_economic_situation	158	58	2	53	56	58	60	65
Covid19New_deaths	124	51	69	0	5	17	75	296
Financial_KPIs_TotalDiscount_Cost	158	19,451	9,829	1,939	12,358	17,637	25,921	51,126
Google_TrendsBrand1	158	54	17	12	42	53	64	100
Page_viewsDaily_Errors	146	1,261	1,087	11	669	1,106	1,525	8,767
Sports_Event_Encoded	158	2	1	0	1	2	3	5
Sum_of_Sunshine_hours	158	34	23	1	14	31	49	88
Week	158	27	15	1	14	27	40	53
Year	158	2,020	1	2,019	2,020	2,020	2,021	2,022

The correlation plot of the media variables along with sales demonstrated mild correlations between media variables and sales variables. It was notable that the TV TRP had a high correlation of 71% with Online Display impressions. It was also notable that the TV TRP as well as Online Display impressions both had a negligible correlation with Affiliates Cost. Affiliates is a media channel that quantifies the sales brought in by social media influencers or ‘aggregator websites’. At the case-company, these Affiliates were paid a fixed fee regardless of the number of sales they generated, in addition to a revenue share ‘deal’ for the level of sales they generated. In some companies, marketers complain that these Affiliates cannibalize sales from other media channels, however, for our case company the Affiliates channel showed low correlations with media impressions of other channels. Even though it's correlation with the Online Video (OLV) spend was at 78%, that correlation dropped to -38% when OLV impressions were considered, as demonstrated in Figure 6 below.

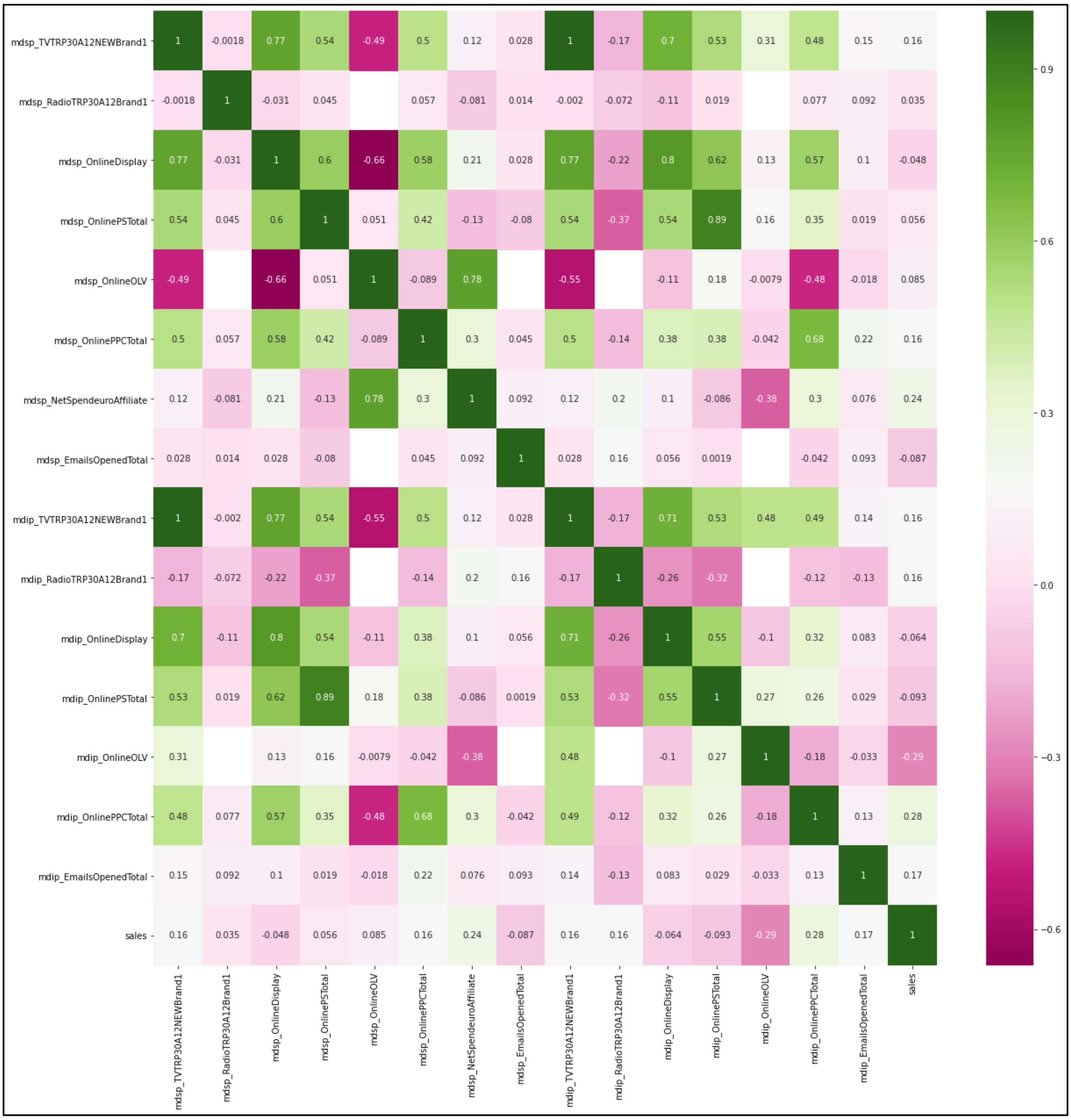

Figure 6: Correlation Matrix of Media Variables and Sales

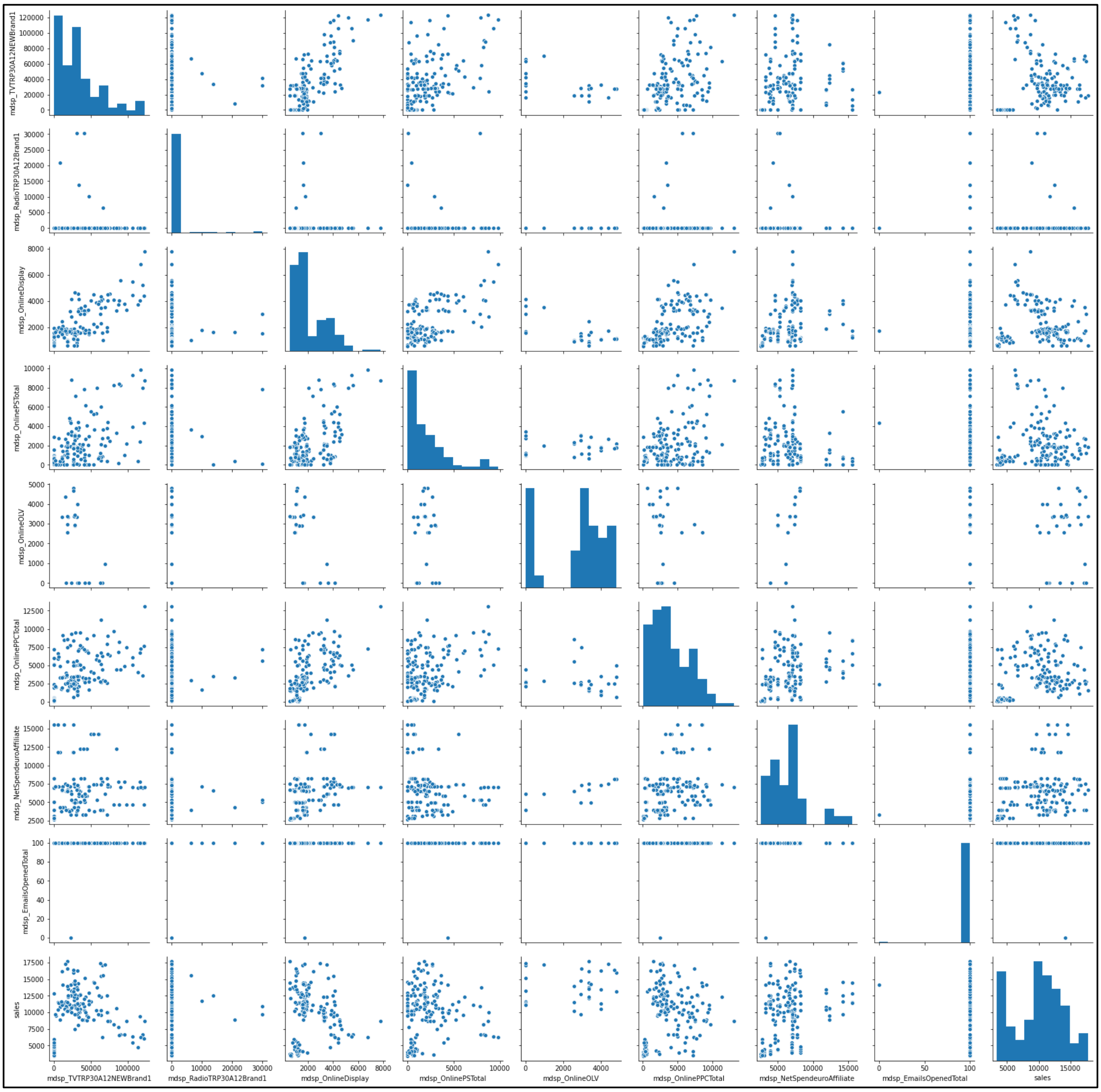

Figure 7: Scatter and Distribution plot of media spend variables and sales

The behaviour of the correlation among media impressions and sales can also be visually inspected from the scatter plot in Figure 7 above. The plot of distributions in the diagonal of Figure 7, indicated that the distributions of raw data were not normal and hence the likes of log transformation would be needed to make them normal. For example, the distributions of the media spend variables was highly skewed and far from normal. This meant that the data needed transformations in order for the linear regression model to have met normality assumption of Ordinary Least squares. Moreover, scatter plots in the last row in Figure 7 showed that increasing spend (x axis) on a marketing channel could result in a variety of sales outcomes (y axis), i.e., an increase in spend was not directly proportional to an increase in sales. Only the OLV channel seemed to have had a relatively consistent positive relationship to sales, although it was noted that there were fewer data points for that channel since it was activated in the last year only.

Comparably, a similar plot of media impressions and sales in Figure 8 below showed that the relationship of OLV with sales deteriorated when measured using impressions.

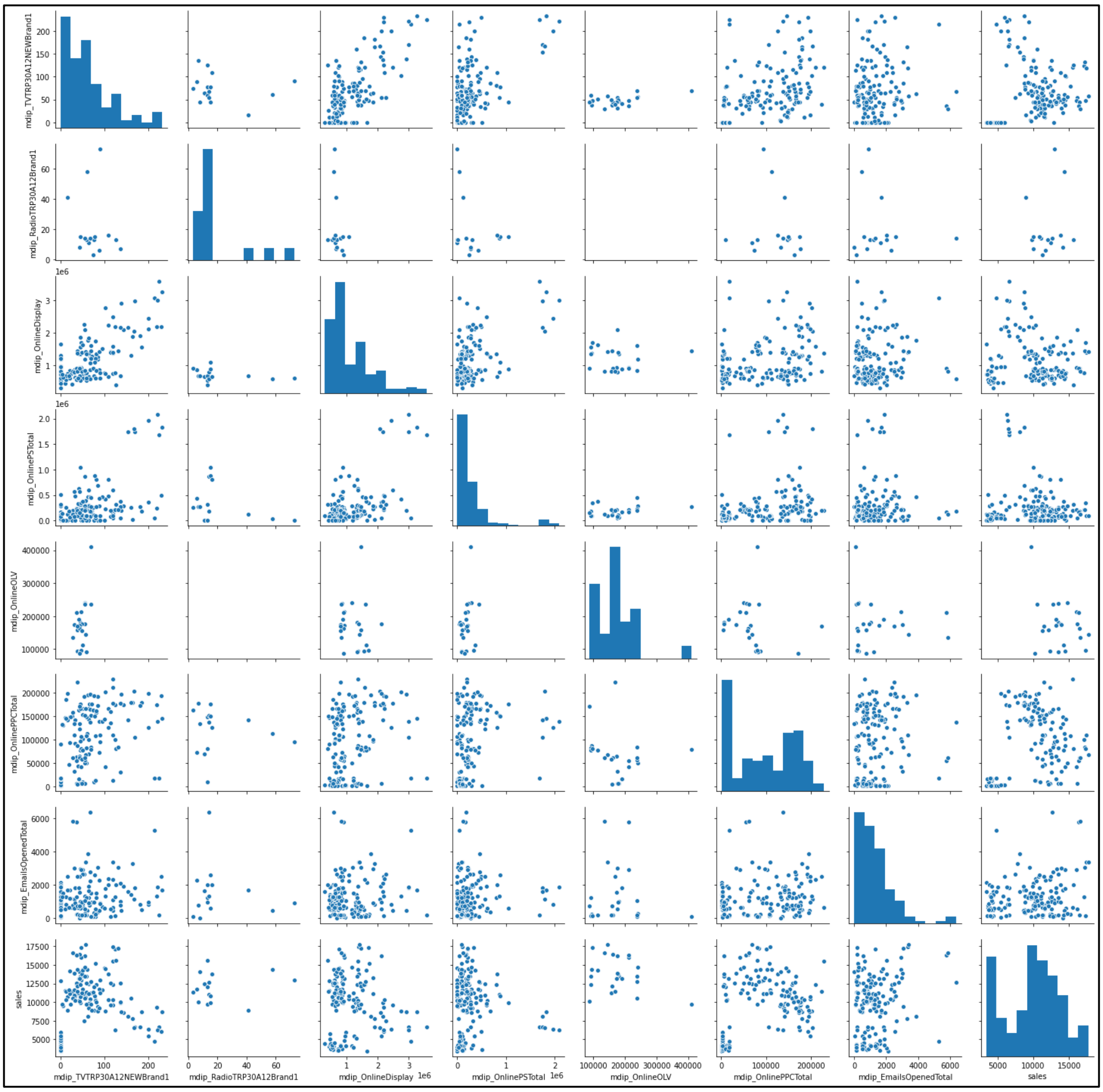

Figure 8: Scatter and Distribution plot of media impression variables and sales

Moreover, the last row in Figure 8 shows that the variance in sales outcomes (y-axis) based on the same level of media impressions (x-axis) can be very high, demonstrating that there is a strong impact from other variables that go beyond the media mix, such as control variables.

6.2. Model Performance – Log Differenced Linear Regression

Table 4 provides the step-by-step process used for the development of the log differenced linear regression model, whereas Table 2 provides the definitions of the variables used.

Model development involved iterative removal of variables using step wise (Yamashita et al., 2007) backward feature selection based on highest p values above 0.1 (Heinze et al., 2018), VIF values above 5 (Desboulets, 2018), and incremental improvements observed in the Adjusted R squared (Jennrich & Sampson, 1968).

This allowed for the achievement of a strong goodness of fit value of 89.7% and a significant F statistic as demonstrated by Table 11:

Table 11: Log Differenced OLS Regression Performance

Measure	Value
R-squared	89.70%
Adjusted R-squared	88.50%
F-statistic	76.6
Probability (F-statistic)	3.41E-61
AIC	-120.3

In order to test for OLS assumptions, various tests were used, and the model did not violate the OLS assumptions. The first OLS assumption was that the residuals had constant variance and zero mean. The plot of residuals provided below in Figure 9 confirms this assumption.

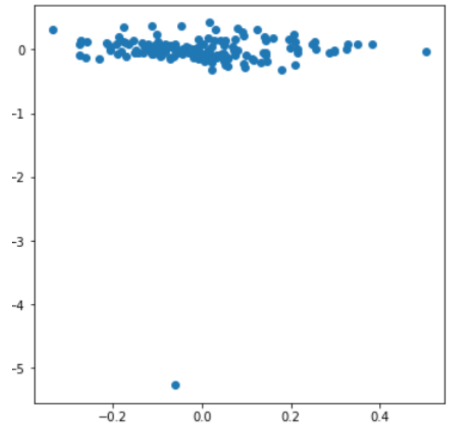

Figure 9: Plot of residuals of for Log Differenced OLS regression

The second assumption of OLS was that the residuals did not have autocorrelation. This was measured using the Durban Watson statistic, the value for which was 2.4, indicating only mild to no first order autocorrelation. The Breusch-Godfrey test that regresses for residuals also did not reject the hypothesis of no autocorrelation in the nth order. P value for Breusch-Godfrey Test was below significance level, indicating that the null hypothesis of no autocorrelation cannot be rejected.

Moreover, the plot of residuals in Figure 9 also did not exhibit any heteroskedasticity, since the residuals hovered around the mean like white noise throughout the observation period.

The assumption of the presence of heteroskedasticity was also tested using the Goldfeld-Quandt test and the Breusch-Pagan Test, both of which had significant p values for the null hypothesis of the presence of homoskedasticity.

The fourth assumption of OLS was that the residuals were normally distributed. In order to confirm this, a Q-Q plot is provided in Figure 11 below, which confirms that the residuals were normally distributed.

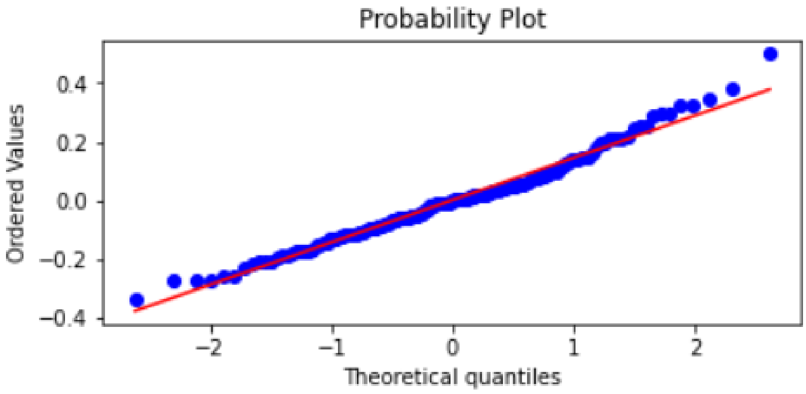

Figure 10: Q-Q plot for Log Differenced OLS regression

The Jarque Bera Test (Thorsten & Herbert, 2007) however revealed a moderate score 0f 4, with a p value of 0.1, indicating that the residuals were not perfectly normal. This is likely due to a few outliers as can also be observed in the Q-Q plot in Figure 10. Removal of observations that are outliers however risks the loss of useful information to marketers due to which additional models are built.

The fifth assumption of OLS was that variables should not be linearly dependent on one another. Figure 6, 7 and 8 from the exploratory analysis section demonstrated that the Online Video (OLV) impressions had a high correlation of 71% with TV TRP. Due to the nature of marketing, media variables could exhibit such linear dependence on one another due to which more advanced simulation-based techniques were used to make conclusions about their behaviour based on their distribution. The correlation between explanatory variables and residuals was also insignificant, hence satisfying the sixth OLS assumption.

The model's results were also validated when trained on 80% of data and tested on 20% of data. The model had a MAPE of 116% and RMSE of 0.15. The plot of actual and fitted values is given in Figure 11 below:

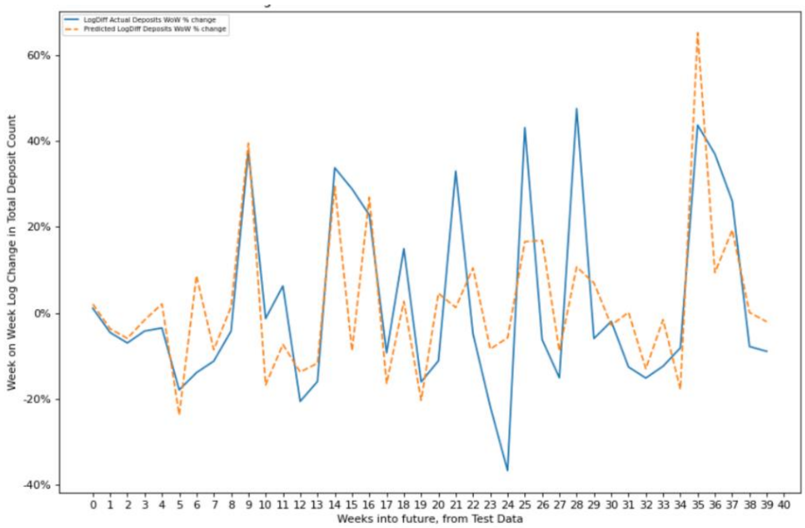

Figure 11: Plot of actual and fitted values from OLS Log Differenced regression

Figure 11 above demonstrates that the model managed to directionally predict the sales correctly on most occasions, during the testing period. However, the errors were exacerbated due to some of the peaks and troughs.

The results of this OLS Log Differenced linear regression model were validated with marketers at the case-company. Sunshine hours and website downtime were reported to contribute negatively to sales, as people were less likely to stay at home and place bets if the weather was sunny in the Nordic country where the brand operates. It was also confirmed that customers usually placed more bets during the first week of the month when they got salaries, and the number of bets placed dropped significantly during the last week of the month. The variable named ‘Calendar_Week_Tot4’ successfully captured this negative relationship.

Based on Table 12 given below, the equation (18) from Methodology section can be rewritten as:

$$\Delta Sales_t \ = 0.0028\varphi \ + \ 0.1756\Delta Financial_K PIs_T otalDiscount_C ost \ +$$

$$0.0901\Delta TVTRP \ - 0.0172\Delta Radio \ + 0.0031\Delta PS \ + 0.3156\Delta OnlineDisplay \ +$$

$$,,,,, + 1.97\Delta SportsEvents \ + \ 0.32\Delta CalendarEvents \ - \ 2.14\Delta Calendar\ Week \ -$$

$$\ldots .0.0220\Delta Covid19Deaths \qquad \textbf{(35)}$$

Table 12: Log Differenced OLS Regression Co-efficients (Note: Data definitions and collection method of the variables in Table 12 above have been provided in Table 2)

Variable Name	Coefficient
Intercept	0.0028
FinancialKPI_TotalDiscount	0.1756
TV_TRP30_A12_NEWBrand1	0.0901

Radio_TRP30_A12_Brand1	-0.0172
Online_PS_Totalimp	0.0031
Online_DisplayImpressions	0.3156
Online_PPC_Totalimp	0.0584
Online_OLVImpressions	0.0451
Google_TrendsBrand1	0.1438
Affiliate_Average_CPA_PerRegistration	0.1656
Emails_OpenedTotal	0.0064
Sports_Event_Encoded	1.9704
Calendar_Events_Total	0.3237
Calendar_Week_Tot4	-2.1453
Sum_of_Sunshine_hours	-0.1278
Page_viewsDaily_Errors	-0.0205
Covid19New_deaths	-0.0220

6.3. Model Performance – Hamiltonian Markov Chain Monte Carlo

The Hamiltonian Markov Chain Monte Carlo (HMCMC) method was implemented using STAN programming language and Python. Table 5 provides the step-by-step process used for the development of the HMCMC model, whereas Table 2 provides the definitions of the variables used. The results of all simulations reported full convergence with Rhat values of 1. As mentioned earlier, HMCMC is a sampling method, that provides a posterior distribution as opposed to coefficients for each explanatory variable, based on the target variable's posterior distribution. The summary statistics of the posterior distribution such as the mean, standard deviation and percentiles for each simulation are provided below.

6.3.1. Simulation Results for Baseline Estimation

In the first step, the HMCMC model estimated the baseline sales i.e., sales that the business would have achieved in absence of media spend. The variables used in this model included Google Trends (beta1(1)), Sports Calendar (beta1(2)), Promotional Discounts (beta3(3)) and Promotional Spend (beta4(4)). These variables were constrained to only hold a positive sign, due to their positive relationship with sales.

Eight other variables were also modelled, which were allowed to hold both positive and negative values. These included Year (beta2(1)), Week (beta2(2)), Sunshine hours (beta2(3)), Weekly seasonality in a given month (beta2(4)), Website Downtime and Errors (beta2(5)), Covid19 New Deaths (beta2(6)), Consumers' Desire to Save given the economic situation (beta2(7)), and Calendar of public holidays or seasonal festivals (beta2(8))

The results of the simulation are provided below in Figure 12. Each row represents a different parameter or variable of interest, and the columns provide information about the posterior distribution these variables obtained from the MCMC sampling.

```
Inference for Stan model: anon_model_a396b59aabedfaa132f3a814776a219f.
6 chains, each with iter=25000; warmup=12500; thin=1;
post-warmup draws per chain=12500, total post-warmup draws=75000.

            mean se_mean     sd   2.5%    25%    50%    75%  97.5%  n_eff   Rhat
beta1[1]    0.27  3.8e-4   0.08   0.11   0.22   0.27   0.33   0.44  49481    1.0
beta1[2]    0.11  1.9e-4   0.04   0.03   0.09   0.11   0.14    0.2  47236    1.0
beta1[3]    0.05  1.5e-4   0.04 2.6e-3   0.02   0.05   0.08   0.14  65217    1.0
beta1[4]  3.0e-3  9.0e-6 2.8e-3 8.1e-5 9.3e-4 2.2e-3 4.2e-3   0.01  93135    1.0
beta2[1] -9.8e-5  1.0e-6 2.8e-4-6.4e-4-2.8e-4-9.8e-5 8.9e-5 4.4e-4  70079    1.0
beta2[2]  4.1e-3  7.0e-6 1.9e-3 3.9e-4 2.8e-3 4.1e-3 5.3e-3 7.8e-3  73124    1.0
beta2[3]  6.6e-4  3.9e-6 1.2e-3-1.6e-3-1.2e-4 6.6e-4 1.5e-3 3.0e-3  90508    1.0
beta2[4]   -0.06  7.7e-5   0.02   -0.1  -0.07  -0.06  -0.04  -0.02  80211    1.0
beta2[5]    0.08  1.2e-4   0.03   0.02   0.06   0.08    0.1   0.14  64811    1.0
beta2[6]    0.09  7.7e-5   0.02   0.06   0.08   0.09    0.1   0.13  57515    1.0
beta2[7]    0.46  2.1e-3   0.54  -0.59    0.1   0.46   0.82   1.51  67595    1.0
beta2[8]   -0.01  3.3e-5 8.9e-3  -0.03  -0.02  -0.01-5.8e-3 5.7e-3  74644    1.0
alpha       0.17  3.2e-4    0.1 8.4e-3   0.08   0.17   0.26   0.33  95038    1.0
noise_var   0.09  4.2e-5   0.01   0.07   0.09   0.09    0.1   0.12  68726    1.0
lp__       92.47    0.02   2.93  85.73  90.75  92.84  94.58   97.1  24897    1.0

Samples were drawn using NUTS at Sat Jun 17 20:16:40 2023.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).
```

Figure 12: MCMC Results (I)

The column named mean calculates the estimated mean value of the parameter based on the MCMC samples. The mean value of the posterior distribution of Google Trends (beta1(1)) is significantly positive at 0.27 whereas the standard deviation is 0.08 (or 30%), indicating the uncertainty or variability in the estimated parameter. The column named se_mean represents the standard error of the estimated mean.

The columns 2.5%, 25%, 50%, 75%, 97.5% represent the percentiles of the posterior samples, indicating the values below which the respective 2.5%, 25%, 50%, 75% and 97.5% percentile of the posterior distribution lies. The 50% value equals to the median of the posterior distribution. These percentiles also describe the shape of the posterior distribution and can be used to summarize it, for drawing meaningful conclusions, especially when the posterior distribution demonstrates a strong skew (Nandy et al., 2021).

The column named N_eff in Figure 12 indicates the effective sample size, which measures the effective number of independent samples obtained from the MCMC algorithm. The results from the simulation demonstrated very high effective sample sizes confirming chain mixing, in addition to full convergence (Rhat = 1), reproducibility and samples without divergences or high tree depth.

The model had a high MAPE of 32% and RMSE of 2933 during this estimation, which is understandable since it is not possible to build a highly accurate model in absence of media variables. The fitted values of this model were then inputted into the second part of the HMCMC model with adstocked media impressions, to model sales.

6.3.2. Simulation Results for Adstock Parameters and Media Contribution

The second part of the HMCMC model estimated the adstock parameters and the contribution of media variables' posterior distribution to the posterior distribution of sales. The priors were specified earlier in STAN and adopted from (Jin et al., 2017). The simulation results from the model are reported in Figure 13 below:

```
Inference for Stan model: anon_model_c6677ffefdee0513f144508ee1783d0c.
12 chains, each with iter=75000; warmup=37500; thin=1;
post-warmup draws per chain=37500, total post-warmup draws=450000.

                          mean se_mean     sd   2.5%    25%    50%    75%  97.5%  n_eff   Rhat
noise_var                 0.01  2.6e-6 1.8e-3   0.01   0.01   0.01   0.02   0.02 473209    1.0
tau                      -0.07  1.3e-4   0.07  -0.21  -0.12  -0.07  -0.03   0.06 278628    1.0
beta[1]                   0.05  4.7e-5   0.03 4.5e-3   0.03   0.05   0.07   0.11 327716    1.0
beta[2]                   0.05  2.4e-5   0.01   0.02   0.04   0.05   0.06   0.08 373497    1.0
beta[3]                 9.6e-3  1.2e-5 9.5e-3 2.4e-4 2.8e-3 6.7e-3   0.01   0.03 586402    1.0
beta[4]                 6.6e-3  8.1e-6 6.3e-3 1.7e-4 2.0e-3 4.7e-3 9.2e-3   0.02 597450    1.0
beta[5]                   0.09  2.4e-5   0.02   0.07   0.08   0.09    0.1   0.12 392985    1.0
beta[6]                   0.16  5.5e-5   0.03    0.1   0.14   0.16   0.18   0.23 354675    1.0
beta[7]                   0.06  5.8e-5   0.03 7.4e-3   0.04   0.06   0.09   0.13 317026    1.0
beta[8]                   0.76  1.8e-4    0.1   0.57    0.7   0.76   0.83   0.96 287195    1.0
decay[1]                   0.5  2.5e-4   0.19   0.15   0.36    0.5   0.64   0.85 551651    1.0
decay[2]                  0.51  2.5e-4   0.19   0.16   0.37   0.51   0.65   0.86 553831    1.0
decay[3]                   0.5  2.5e-4   0.19   0.15   0.36    0.5   0.64   0.85 560780    1.0
decay[4]                   0.5  2.5e-4   0.19   0.15   0.36    0.5   0.64   0.85 560609    1.0
decay[5]                  0.51  2.5e-4   0.19   0.15   0.37   0.51   0.64   0.86 555878    1.0
decay[6]                  0.51  2.5e-4   0.19   0.16   0.38   0.51   0.65   0.86 545708    1.0
decay[7]                   0.5  2.6e-4   0.19   0.15   0.36    0.5   0.64   0.85 539705    1.0
peak[1]                   0.47  3.8e-4   0.29   0.02   0.22   0.46   0.71   0.97 578432    1.0
peak[2]                   0.52  3.7e-4   0.28   0.03   0.29   0.53   0.76   0.98 575879    1.0
peak[3]                    0.5  3.8e-4   0.29   0.03   0.25    0.5   0.75   0.97 574611    1.0
peak[4]                    0.5  3.8e-4   0.29   0.02   0.25    0.5   0.75   0.98 583785    1.0
peak[5]                   0.49  3.7e-4   0.28   0.02   0.24   0.48   0.73   0.97 577887    1.0
peak[6]                   0.56  3.7e-4   0.27   0.04   0.35   0.59   0.79   0.98 547750    1.0
peak[7]                   0.41  3.8e-4   0.27   0.02   0.18   0.37   0.62   0.95 518747    1.0
cum_effect              1559.5     0.3 197.34 1149.7 1436.5 1567.1 1684.2 1948.5 437266    1.0
X_media_adstocked[1,1]     0.0     nan    0.0    0.0    0.0    0.0    0.0    0.0    nan    nan
X_media_adstocked[2,1]     0.0     nan    0.0    0.0    0.0    0.0    0.0    0.0    nan    nan
X_media_adstocked[3,1]     0.0     nan    0.0    0.0    0.0    0.0    0.0    0.0    nan    nan
X_media_adstocked[4,1]     0.0     nan    0.0    0.0    0.0    0.0    0.0    0.0    nan    nan
X_media_adstocked[5,1]     0.0     nan    0.0    0.0    0.0    0.0    0.0    0.0    nan    nan
```

Figure 13: MCMC Results (II)

The model converged for media variables and failed to find carryover effect beyond the first week, due to which the 'nan' values appear for X_media_adstocked in the table above.

The variables for which posterior distributions were estimated, as per Figure 13, included fitted sales values from previous simulation to represent control variables (beta1), TV TRP (beta2), Radio TRP (beta3), Online Display Impressions (beta4), Online Paid Social Impressions (beta5), Online Video Impressions (beta6), Pay Per Click Impressions (beta7) and Emails Opened (beta8).

Adstock parameters such as decay and peak factors were estimated for each media channel, which represented the posterior 'truth' discovered by the model. The prior belief specified for lag effects was 3 weeks, based on business knowledge of the marketers at the case company. Repeated experiments with different priors for lag effects, including the recommended 13 weeks from Google (Jin et., 2017) yielded almost exactly similar estimates, although the time to convergence increased, which can be expected if the specified priors are farther from the posterior.

For this part of the model, 75000 iterations were specified to ensure convergence, although the results were almost the same when the sampling was repeated with 25000 iterations, confirming reproducibility. The model had a MAPE of 4.8% and RMSE of 0.14.

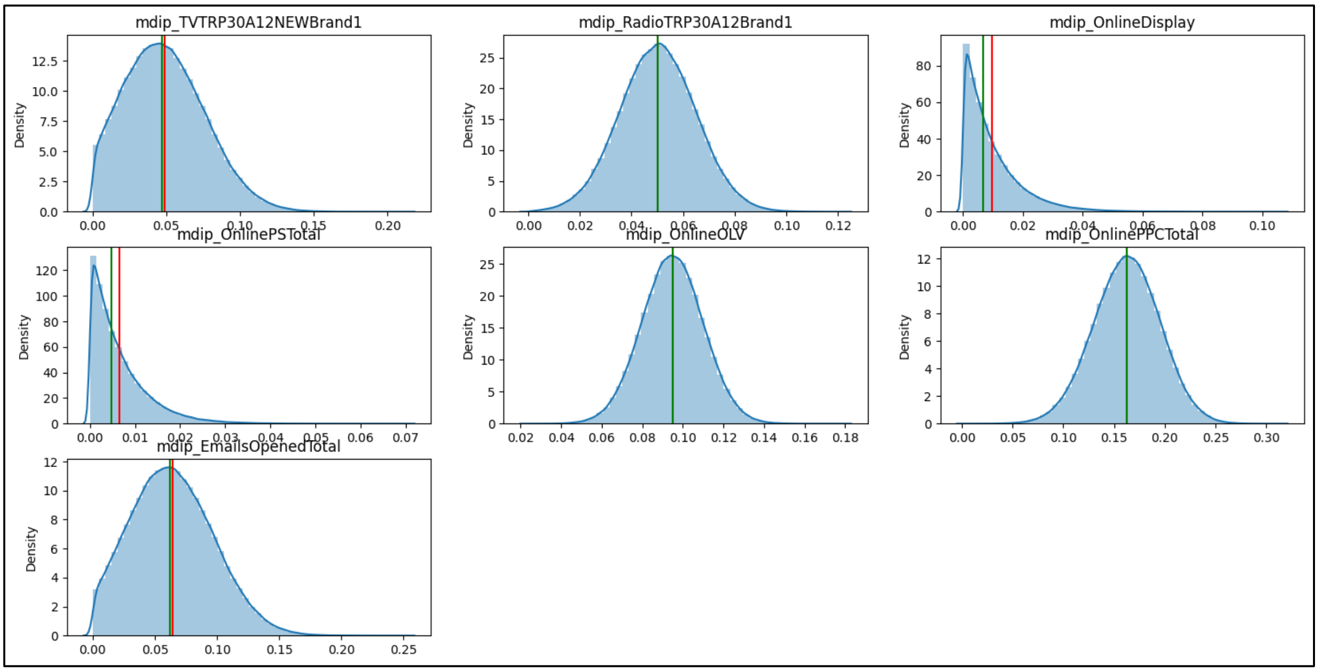

Figure 14: Distribution of Return on Adspend computed using Posterior distribution of media variables

Figure 14 demonstrates a plot of the distribution of return on advertising spend computed using posterior distribution of media variables. The distributions demonstrate a range of behaviours or typical set of each media channel developed using repeated iterations from the MCMC simulation.

6.3.3. Simulation Results for Posterior Estimation of Saturation Points for Each Media Channel

The previous steps of the HMCMC model allowed for the fitting of baseline sales using control variables in the first part, and fitting of the media mix to understand the contribution of various media channels in the second part, in addition to the estimation of posterior estimates for Adstock parameters such as peak and decay.

The media contribution calculated in the second part of the model, alongwith posterior Adstock parameters were then used in the third and final part of the model. Media spend on each channel was transformed using the posterior Adstock parameters. The priors for this model were specified in STAN programming language, and were adapted from (Jin et al., 2017). These priors included ‘beta hill’, which harboured the equation of the Hill Function that was specified earlier in the Theoretical Framework section of this thesis. The priors also included a vector of means of the posterior distribution of the media variables calculated in the previous step, in addition to the ad-stocked media spend. Sales were modelled once again using the parameters specified in the priors, with an aim to obtain ‘ec’ and ‘slope’ parameters of the Hill Function that allowed the calculation of the saturation points for each media channel. The results from each simulation demonstrated full convergence (Rhat = 1), very high effective sample sizes confirming chain mixing, reproducibility and samples without diverges or high tree depth.

6.4.Model Performance – ANFIS with PSO

In this thesis, 100 ANFIS models were trained for 10000 iterations using the PSO. While the earlier OLS regression and HMCMC methods measured media variables using impressions or GRP, for this forecasting model, media spend was used as the proxy for media performance, in order to give the marketer, the most confidence in their budget allocation scenarios. This is because media impressions are an ‘after the fact’ variable whereas ‘media spend’ is known ‘before the fact’, or before the marketing campaign is initiated. Hence using media spend instead of media impressions was logical for forecasting purposes.

The model specification was initially done using feature selection techniques such as XGBoost Regressor, Random Forest and Decision Trees, using forward feature selection, however, it resulted in dropping of key media variables such as TV expenditure. These media variables were earlier considered significant not only by OLS Regression, but also by the HMCMC model, since they had concentrated distributions indicating high contribution to the outcome variable. Hence feature selection was abandoned and knowledge accumulated from the OLS Regression and Hamiltonian Markov Chain Monte Carlo distributions was used to specify the model. The model performed with high accuracy, capturing peaks and troughs, when tested on 20 weeks or almost 5 months of unseen data, as shown in Figure 15 below:

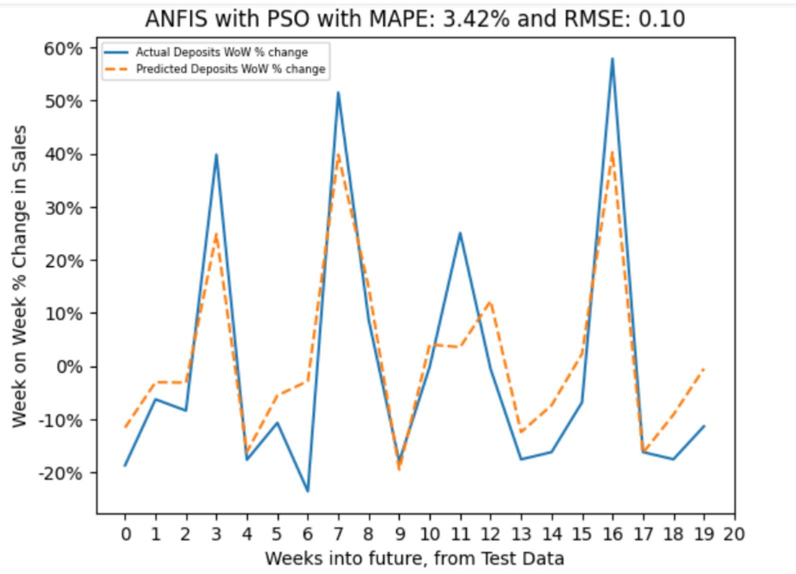

Figure 15: Forecast of sales from ANFIS PSO

6.5.Findings

In this sub-section, the findings from the model development exercises are presented, with an aim to answer the research questions. These findings were based on the outputs of the three approaches developed.

6.5.1. Contribution of Media Channels to Sales

The main goal of marketing mix modelling is to understand effectiveness of marketing channels, and use the insight generated to better allocate or optimize media expenditure, especially in cases where budgets are over or under allocated to channels that could potentially perform well. Three interpretable models were developed albeit under different statistical paradigms and hence interpreted differently. Despite the different interpretations the common finding was that TV, Online Display and Affiliates were amongst the top 3 contributing channels to sales. These results were however measuring correlation and should not be confused with causation.

The first model was a log differenced OLS Regression model, which can be interpreted as percent change. The model found that the highest percentage change in sales, can be achieved by increasing the Online Display Impressions (31.6%), however the HMCMC saturation points explained later in Table 17, indicated a low saturation point for this channel.

Table 13: Log Differenced OLS Regression Co-efficients Note: Data definitions and collection method of the variables in Table 13 above have been provided in Table 2

Variable	Coefficient	Rank	Standard Error
TV_TRP30_A12_NEWBrand1	9.0%	3	4.9%

Online_DisplayImpressions	31.6%	1	3.4%
Affiliate_Average_CPA_PerRegistration	16.6%	2	6.7%
Online_PS_Totalimp	0.3%	7	0.6%
Online_OLVImpressions	4.5%	5	2.1%
Online_PPC_Totalimp	5.8%	4	1.9%
Emails_OpenedTotal	0.6%	6	1.1%
Radio_TRP30_A12_Brand1	-1.7%	8	13.8%

It should be noted that the Online Display Impressions had a high correlation of 71% with TV TRPs, as was noted earlier in the Exploratory Analysis sub-section of the Results section in the thesis. This high correlation would have likely induced multicollinearity between the two variables due to which some of the contribution from the TV TRPs had likely been misallocated to the Online Display Impressions by the OLS regression model. This is confirmed by both MCMC posterior distributions (Table 14) as well as coefficients from ANFIS (Table 15), both of which are more suitable models to deal with multicollinearity in high dimensional space.

The OLS regression assigned the second heaviest coefficient to variable named ‘Sports_Events_Encoded’ which measured the number of sport events happening in a given week. This was confirmed by the marketers as the single strongest driver of their business, since they operated in the sports betting industry. Marketers validated the signs of the co-efficients of all variables in the OLS regression model, apart from Radio, which had a negative sign, indicating the presence of some multicollinearity in the OLS model. However, it is notable that the case-company only used the radio channel for 6 weeks throughout the 3-year period and hence a low number of observations could have hindered the performance of the OLS regression with regards to it. It is also notable that Radio spend was at a significantly high level of Euro 111,663 during the course of the 6 week period. While it is practically not possible for a media channel to contribute negatively, it can be concluded that the channel had a mild impact when considering the entire 3 year period of data that was modelled.

The second model was a Markov Chain Monte Carlo model using Hamiltonian Dynamics. This method allowed for the integration of both costs and impressions or TRP data in order to calculate saturation points. The output of this model with regards to media mix variables was in the form of a posterior distribution as opposed to a single coefficient, as shown in Table 14 below. Since the posterior distributions were non normal, and instead skewed, considering only the ‘mean’ of the distribution and using it as the coefficient of the variable would’ve not fully informed the marketer about the true behaviour of the explanatory variable (Nandy et al., 2021). Due to this, the mean was considered in conjunction with other metrics such as the 25%, 50%, 75% 95% percentiles, since they inform about the range of values that the variable can take on. Table 14 below provides a summary of the posterior estimates of each media variable:

Table 14: Hamiltonian MCMC model's estimation of variable distributions (Note: Data definitions and collection method of the variables in Table 14 have been provided in Table 2)

Dataset Variable Name	STAN Model Output Name	mean	sd	Credible Interval 50%	Credible Interval 2.5%	Credible Interval 97.5%	Proportion of Range of Parameter (97.5% / 2.5%)	Proportion of Range * Mean (Accounting for Mean Value)	Rank
TV_TRP30_A12_NEWBrand1	beta[1]	0.0500	0.0300	0.0500	0.0045	0.1100	24.4	1.2	1
Online_DisplayImpressions	beta[3]	0.0096	0.0095	0.0067	0.0002	0.0300	125.0	1.2	2
Affiliate impressions (no data)	--	--	--	--	--	--	--	--	--
Online_PS_Totalimp	beta[4]	0.0066	0.0063	0.0047	0.0002	0.0200	117.6	0.8	5
Online_OLVImpressions	beta[5]	0.0900	0.0200	0.0900	0.0700	0.1200	1.7	0.2	8
Online_PPC_Totalimp	beta[6]	0.1600	0.0300	0.1600	0.1000	0.2300	2.3	0.4	6
Emails_OpenedTotal	beta[7]	0.0600	0.0300	0.0600	0.0074	0.1300	17.6	1.1	4
Radio_TRP30_A12_Brand1	beta[2]	0.0500	0.0100	0.0500	0.0200	0.0800	4.0	0.2	7

The larger the proportional difference between the lower and upper bounds of the percentiles of the posterior distribution, the greater the variability associated with the variable. Moreover, visually inspecting the plots of distributions did not provide enough confidence for ranking the media channels. Hence a ranking was created by multiplying the mean of the distribution of each variable, with the proportional difference between the maximum and minimum percentile. This technique enabled insights into the overall variability or uncertainty associated with each variable (Nandy et al., 2021).

From the linear model it was understood that there existed some multicollinearity between the TV TRP and Online Display impressions. From the HMCMC model, it was noted that the mean for TV (0.05) was proportionally much higher than the one for Online Display impressions (0.0096), This indicated that on average, the TV channel contributed five times more than the Online Display channel. However, the distribution of the Online Display Channel showed a much higher variance than TV, as evident from the difference between the percentiles. Accounting for this variance indicated that the Online Display Channel could have had as high positive contribution as TV, if the conditions were right. However, on average the channel's impact on sales was five times lower on TV. This finding was also validated with the marketers as they confirmed the occasional viral campaigns from the Online Display channel that led to exceptional impact, however, those instances were fewer. For the remaining channels, the insights provided by HMCMC about the media channel's behaviour complemented the already estimated coefficient estimates from the Linear model.

Using the mean and the percentiles together hence informed more about the nature of the distribution of the media variable, and its behaviour, which was the primary purpose of running simulation models. Considering this, a ranking of the media channels was created. The ranking from the HMCMC model validated the ranking for the top three contributing channels, from the OLS Regression built earlier. The key difference being that the HMCMC model accounted for the uncertainty in parameters by modelling the media channels' posterior distributions and sampling from it.

It is important to note that the HMCMC model adapted from (Jin et al., 2021) followed a structure that modelled adstock parameters and media impressions simultaneously for each media channel, followed by modelling of the spike in sales caused by media impressions along with the media cost using the Hill function. Hence data was required on both media costs and media impressions to get insights from this model, however, such data on impressions was unavailable for Affiliates. Affiliates are typically partners of a company, such as social media influencers, which post their discount codes on personal social media profiles to entice customers to purchase. They are paid both a fixed fee and a performance fee, at the case-company. However, the industry-wide challenge with this channel is its measurement, since data is only available on costs and not on activity (such as impressions or clicks). Hence it was not possible to include the Affiliates channel in the second and third stage of the HMCMC model and calculate its lag effects, contributions or saturation points. Instead, its inclusion was tested in the first stage of the model alongwith control variables which led to a 10 percentage point improvement in the MAPE values from the model.

The third model that was built was the ANFIS model. While its primary purpose was to provide accurate forecasts, it also provided with ‘Consequent Coefficient’ values for each variable, which have been defined earlier in the Theoretical Framework section of this thesis. Like the HMCMC, the ANFIS model also provided a (fuzzy) distribution with mean and variance values, as opposed to single point coefficient estimates provided by the OLS Regression model. Even though the ANFIS model is interpretable due to the usage of gaussian and a single hidden layer, it is important to consider that it provides fuzzy values for mean and variance. Mathematical operations on fuzzy data are carried out differently to regular math. Hence it is not possible to get exact coefficients without further research on defuzzification of values obtained from an ANFIS model. Approaches such as COA exist however due to time and scope limitations they have not been carried out as part of this thesis. Given below in Table 15, are the parameter estimates obtained from the ANFIS model:

Table 15 – ANFIS model’s estimation of variable distributions

Variable	CONSEQUENT_Coefficient_Values 'A'	mu Variance of Mean in Consequent Function	Rank based on Mean
Net_Spend_euroTV_Total	19.3	26.1	1
Net_Spend_euroOnline_Display	18.1	-0.6	2 *(Note Negative Variance)*
Net_Spend_euroAffiliate	13.8	1.9	3
Net_Spend_euroOnline_PS_Total	5.8	-4.3	4 *(Note Negative Variance)*
Total_EmailsSent	5.6	25.2	5
Net_Spend_euroOnline_PPC_TOTAL	4.7	5.8	6
Net_Spend_euroOnline_OLV	3.7	2.8	7
Net_Spend_euroRadio	8.6	-0.1	

The mean of the Consequent function determined by the ANFIS model validates the finding of the OLS regression model that the top 3 contributing channels are TV, Online Display and Affiliates. The same finding was also confirmed by HMCMC as discussed above.

Since the variance of mean in consequent function of ANFIS, for the Online Video variable, was negative and below 1, it demonstrated that the distribution was skewed to the positive side, and experienced spikes further to the positive side, which could have been due to viral campaigns, as also indicated by the HMCMC model. Similarly, for the Radio channel, the distribution was also skewed to the positive side and also experienced a positive spike. Here it is notable that while techniques such as OLS regression and HMCMC would consider the Radio channel an abnormality due to only 6 weeks of data available on it (out of 158 weeks), the neuro fuzzy inference system in ANFIS instead first checks the impact of each variable one by one against the target variable before evaluating them jointly and calculating consequent coefficients. The insight from this process for the case-company is that the radio channel did contribute positively, and during the 6 weeks while it was switched on, it had greater impact than other digital channels apart from Online Display.

The variance of mean in consequent function of ANFIS, for the Paid Social channel was -4.3, i.e., it was negative and greater than 1. This demonstrated that the distribution was mildly positive and experienced strong negative variance, indicating some positive performance followed by periods of underperformance that reduced its overall impact on sales. This finding about the low impact of Paid Social was also confirmed from the HMCMC and the OLS regression, as well as by the marketers.

Further analysis of the positive and negative behaviour of the distributions obtained from the HMCMC model and the ANFIS model can be conducted in order to better understand situations when the media channels make the higher contributions, so that those situations can be replicated.

6.5.2. Carry Over or Lag Effects

The second stage of the HMCMC model allowed for the computation of the lag effects. A Bayesian prior of three weeks of Adstock was set, which was then corrected by the model, to output posterior parameters of ‘peak’ and ‘decay’, as shown in Table 16. Due to the nature of the business, which is sport event driven, as well as the timing of campaigns by the case company, the lag effects decay within a given week.

Table 16 – Adstock Parameters Estimated by MCMC

Media Channel	L (Prior in weeks)	P (Posterior Peak in Weeks)	D (Decay Rate
TVTRP30A12NEWBrand1	3	0.472767272	0.502345716
RadioTRP30A12Brand1	3	0.522670395	0.508759631
OnlineDisplay	3	0.498605185	0.499729298
OnlinePSTotal	3	0.498191937	0.499557455

OnlineOLV	3	0.487747561	0.505786603
OnlinePPCTotal	3	0.560904403	0.512128724
EmailsOpenedTotal	3	0.410477651	0.501398196

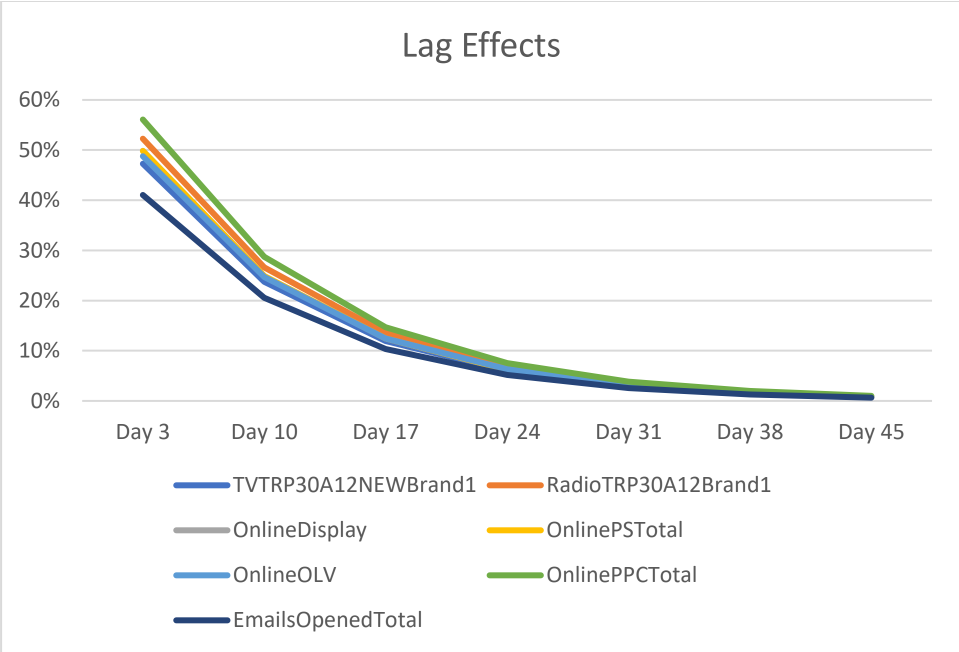

Figure 16. Lag effects for each media channel

Figure 16 below demonstrates that the lag effects for media channels peak on third day of the spend and then decay fast. This has been discussed with the marketers at the case company and is attributed to the timing of sport event specific campaigns launched only a few days prior to the event, with the marketing impact of the campaign decaying to zero at the end of the given sport event. Table 16 above provides the actual adstock parameters calculated using the HMCMC.

6.5.3. Saturation Points for each media channel

The third stage of the HMCMC model allowed for the calculation of saturation points for each media channel using the Hill function. TV had the highest saturation point at a weekly spend of 52,638 Euros, as shown in Table 17. Although the contribution of Online Display could sometimes be very high, it had a low weekly saturation point which might justify the lower expenditure on the channel. This was also confirmed by the marketers that they had tried raising the spend for Online Display channel in the past, however, the results from the channel did not justify the additional spend, due to which a relatively lower level of spend was maintained for the channel.

Here it is important to note that the Hill Function only finds the local optima based on maximum expenditure on each media channel so far, based on the history of the data series. This means that there can be more local optima at higher spend levels so far unobserved in the data, which could be observed, if spending levels are raised by the case-company to significantly higher new levels. Hence the outputs from the Hill function posterior estimates need to be considered in light of the marketing investment strategy.

Table 17 – Saturation Points estimated by MCMC

Brand Name	Maximum Weekly Spend from Dataset	Maximum Range of Normalized Outputs from STAN	Saturation Point of the Hill Curve	Saturation point fitted to Cost in Real Range
Net_Spend_(euro)TV	EUR 122,822	3.50	1.50	52,638.00
Net_Spend_(euro)Radio	EUR 30,187	18.50	2.50	4,079.30
Online_PS_Cost_total(Euro)	EUR 7,814	4.50	2.00	3,472.80
Net_Spend_(euro)Online_PPC	EUR 13,081	3.00	0.50	2,179.08
Net_Spend_(euro)Online_Display	EUR 7,779	3.50	1.50	3,333.80
Online_OLVSpend_cost_(euro)	EUR 4,801	8.50	3.50	1,976.80
Email Total	6,375.00	0.50	0.40	5,100.00

Table 17 above shows that there is room to increase the spend for TV, given that the current spend is almost 30% below the saturation point. Whereas Online Display and Paid Social channels' spend is already close to the saturation point, determined based on marketers' own experiences so far with these channels and validated by the Hill Function. It was also notable that the saturation point for the PPC channel was almost 50% below the current spend levels. This was confirmed by the marketers that they had experienced PPC to be a more expensive channel for the results it brings in, which could be due to the spending beyond the saturation points.

Figures 17 and 18 below provide plots of the saturation points. The fewer data points for both radio and OLV are noticeable.

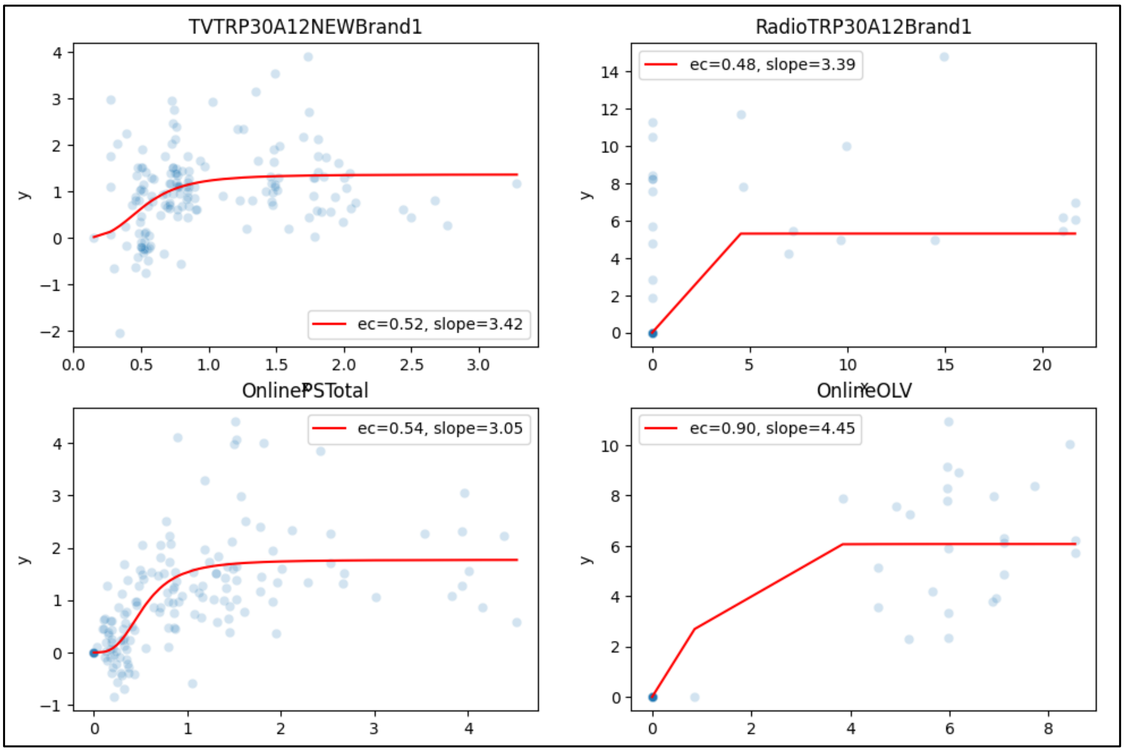

Figure 17 – Plot of Saturation Points (I)

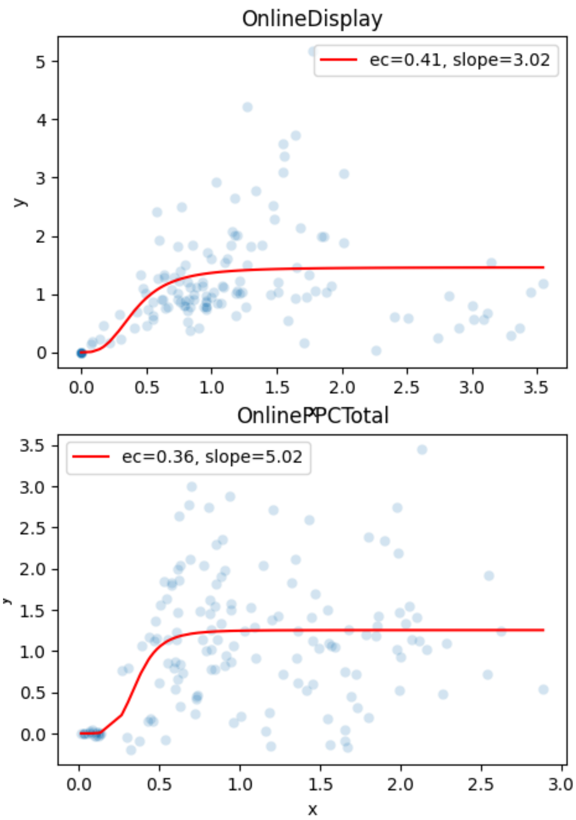

Figure 18 – Plot of Saturation Points (II)

6.5.4. Forecast

The ANFIS model trained using PSO provided a highly accurate forecast for weekly change sales, using media spend, control and seasonal variables. MAPE was at 3% and RMSE at 0.10 . The plot of actual and predicted values is given in Figure 19 below:

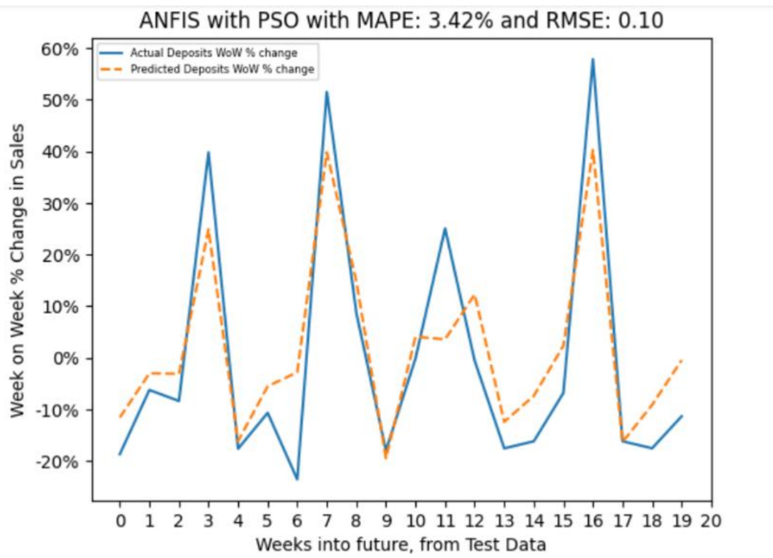

Figure 19: Forecast of sales from ANFIS PSO

6.6. Results Summary

In summary, HMCMC, ANFIS, and OLS regression were identified as state-of-the-art methods due to their publication in high-impact factor journals and papers that were highly cited. Table 18 provides a summary of the performance of these methods:

Table 18 – Model Performance Summary

Performance Metric	Log Differenced OLS Regression	ANFIS with PSO	HMCMC
R Squared	89.70%	--	--
MAPE	116%	3.40%	4.80%
RMSE	0.15	0.1	0.14
Rhat	--	--	1
N_eff	--	--	High
Tree Depth	--	--	0
Divergence	--	--	0
Reproducibility	--	--	TRUE

Table 19 shows a summary of the outputs from the three models developed, quantifying the impact of media channels on sales. The models broadly validated the findings from each other, and their findings were also validated by the marketers at the case company.

Table 19 – Media Variables' Impact on Sales

Channel	OLS Regression	ANFIS with PSO		HMCMC
	Coefficients	Mean of Fuzzy Distribution	Variance of Fuzzy Distribution	Proportion of Range of Posterior Distribution * Mean of Posterior
TV	0.09	19.3	26.1	1.2
Online Display	0.316	18.1	-0.6	1.2
Affiliates	0.166	13.8	1.9	--
Pay Per Click	0.058	4.7	5.8	0.4
Online Video	0.045	3.7	2.8	0.8
Email	0.006	5.6	25.2	1.1
Paid Social (Facebook)	0.003	5.8	-4.3	0.8
Radio	-0.017	8.6	-0.1	0.2

Based on the findings, TV advertising had the most significant impact on the case-company's sales, followed by Online Display and Affiliates channels. Despite Online Display at times showing a similar impact per input unit as compared to TV, it reached saturation at a much lower spending level. The findings indicate that case-company has the opportunity to increase its TV ad spending as there is still potential for growth in that channel. However, other media channels are already near saturation levels, except for PPC, where spending is already exceeding saturation by 50%, which can be reduced and re-allocated to TV. The ANFIS model suggested that Radio channel, during the 6 weeks when it was switched on, was more effective than Paid Social, PPC and OLV. Hence another key insight for the case-company is to re-start Radio advertising, the saturation point for which has been calculated to be at Euro 4,079 per week.

There have been no research papers published so far, that perform marketing mix modelling for the iGaming or sports betting industry and share findings. Studies for other industries however exist. (Fareniuk & Chornous, 2023) found TV advertising to be the largest contributor to online and offline sales of a non-food retailer in Ukraine, after seasonal and control factors, which was also the case in this thesis for the case-company. Another study from (Carlos et al., 2020) that used the HMCMC and OLS regression, also found the impact of discounts and other control variables on sales to be highly significant, which was also the case in this thesis. For the case-company, the Paid Social channel, where the company mainly invested in Facebook advertising, was not a key contributor. Comparatively, (Carlos et al., 2020) found Facebook to be the main driver of sales for a clothing retailer in the United Arab Emirates. In contrast, the low performance of Facebook for the case company can be attributed to industry dynamics and regional consumer behaviour, while the Facebook platform itself is also known to regulate gambling companies' advertising in order to meet regulatory requirements related to targeted advertising by gambling companies, which could be a reason for its underperformance.

7. Conclusions and Discussion

In this thesis, three different marketing mix modelling approaches were applied to the past 3 years of data of a digital native iGaming company. The objective of the application was to quantify the impact of media channels on sales and to also provide a highly accurate sales forecasting model, that used marketing variables as inputs, amongst other variables. Literature was reviewed to determine the cutting-edge marketing mix modelling methods, and three of the most suitable methods were applied to the case-company’s data. Earlier research in the realm of marketing mix modelling, since the 1970s, focused on the usage of statistical methods, such as linear regression. Application of MCMC methods became more prevalent since the 2000s, likely due to significant improvement in available computational power to run such computationally expensive methods. However, challenges with tuning MCMC methods meant that other approaches such as SEM and VAR were also used for MMM, while log-log regressions continued to stay in fashion. These challenges with tuning MCMC were solved by the automated tuning provided by the NUTS sampler implemented in STAN. It was also notable that to this day, only a scarce amount of literature is available on the application of AI and deep learning methods to answer questions posed by MMM.

In this thesis, three methods were implemented to perform MMM modeling, on 3 years of weekly data, sourced from an iGaming company. These methods included log differenced OLS regression, Hamiltonian MCMC simulation, and an ANFIS model for forecasting. OLS regression provided strong goodness of fit, however its performance with test data was moderate. There was also some evidence of multicollinearity and divergence of its residuals from normality. On the other hand, the MCMC simulations fully converged, with no samples having high tree depth or divergences, all chains successfully mixing and high effective sample sizes. The ANFIS model also provided better than expected forecasting accuracy, along with mean and average values for the consequent function of each media variable, that validated the ranking of media channels’ contributions to sales provided by MCMC posterior distributions as well as by the coefficients of the linear regression.

The findings indicated that TV was ranked by all three models as the largest contributor to the case-company’s sales, followed by Online Display and Affiliates channel. It was also notable that one unit of input to the Online Display channel, resulted in a similar impact on sales, as compared to TV, however the saturation point of the Online Display channel was much lower than TV. This finding was validated by the marketing and product managers at the case-company as something they had experienced and were already aware of. The saturation points also indicated that there was room for the case-company to increase spending on TV, whereas spending was close to saturation already for other media channels, apart from PPC, where the current level of spending was 50% above saturation. These insights were also validated by other research studies commissioned by the case-company, including an MMM model that was developed on the exact same dataset by a marketing analytics consultancy.

7.1. Answering the Research Questions

RQ1: What are the state-of-the-art machine learning or statistical analysis based Marketing Mix Modelling methods according to the literature and what are their key application areas?

Table 1 provides a summary of 17 research articles that apply state-of-the-art machine learning or statistical analysis-based Marketing Mix Modelling methods. These research articles were shortlisted based on the impact factor of the peer reviewed journals that they were published in. The methods applied in these articles, include log-log regression, Vector Autoregression (VAR), Tobit regression, Structural Equation Modelling (SEM), MCMC and ANFIS. These methods were applied in a wide variety of industries. In situations where interpretability of the model was of utmost importance (Nazir et al., 2021), linear regression models were preferred over other models. However, for datasets with high level of nonlinear trends or multicollinearity, SEM and MCMC models were preferred, since a linear model would likely fail to provide valid results in such cases. Moreover, MCMC methods can generate data through simulation, that allows to account for unobserved variables while also allowing for the determination of both lag and shape effects scientifically, hence it was also recommended by Google (Jin et al., 2017) as well as by a global marketing measurement consultancy called Nielsen (Chen et al., 2021). VAR model’s application area was in determining lag effects scientifically through the usage of the impulse response function. ANFIS model’s application area was in providing explainable forecast (Kumar et al., 2019).

RQ2: What are the cutting-edge methodologies that are best suited to case-company and provide a high degree of confidence in the results?

The decision-making process in marketing, especially when allocating capital to different media channels, demands interpretable models. This is because it is a high stakes decision (Gosiewska, 2021) and marketers are required to justify their decisions. The objective of the marketing mix modelling exercise for the case-company was to provide such interpretable results, that allow for understanding of the impact of each media variable on sales, while also allowing the marketer to forecast various scenarios.

OLS Log Differenced regression was the first most suitable method for the case-company because it provided the most interpretable and easily explainable insights. However, linear regressions provide point estimates, in the form of co-efficients, which do not always sketch the complete picture, especially in cases where there is multicollinearity amongst variables, seasonal volatility or continuously evolving underlying trends. Given the volatility in case-company's sales, and evolving marketing trends post covid lockdowns and re-opening, it was important to understand the behaviour of each media variable across the time period through its posterior distribution. Hence Hamiltonian MCMC was the second most suitable method for the case-company because it allowed for the modelling of the posterior distribution of the media variables. Moreover, in the case-company's data, TV TRP and Online Display impressions demonstrated a high Pearson correlation of 71%, leading to the possibility of a degree of multicollinearity between the two channels. Hence it was important to build a second method that specialized in dealing with multicollinearity. Through literature review, MCMC was identified to be the most reliable and cutting-edge method for such modelling. An added benefit of the MCMC was that it allowed for the scientific modelling of carryover and shape effects, while also reducing any possible selection bias. Hence it was implemented.

To validate the results and bolster marketers' confidence in the insights from the two measurement models (OLS Log Differenced regression and MCMC), an explainable Ai model (ANFIS with PSO) was also developed. A widely cited research paper by (Kumar et al., 2019) had already proven the effectiveness of ANFIS in forecasting sales using marketing and other data. It was the most suitable method for forecasting because the objective was not just to forecast accurately, but also to get an insight into the contribution of the variables towards the forecast. This contribution was then compared with the distributions provided by MCMC and coefficients provided by OLS Regression for validation, in order to give maximum confidence to the marketer in the results of the modelling exercise.

7.2.Future Research

Due to the absence of a database for betting odds offered to customers, future research could include such data, as (Sam & Chatwin, 2019) highlight that "time to withdrawal" and the odds offered to customers could impact gamblers' choices about the selection of an online betting company.

The integration of product data scraped from competitor websites and case-company website could result in enhancement of the marketing results. Since the quality (Suardika & Sari, 2021) and assortment of product offered can also influence marketing outcomes such as sales.

An important question often left unaddressed by marketing mix modelers is about how to link the high-level strategic recommendations from the MMM model to tactical optimizations on each media platform? Additional sub-models for each media platform could be developed to answer this question and derive maximum value out of MMM results. Moreover, differences in targeting strategies, and other unobserved variables can cause the outcomes of marketing campaigns to be different from the one suggested by the MMM models. Hence such variables can also be included in the sub-models.

In the case-company's data, high amount of weekly seasonality was observed each month, in addition to sports event driven seasonality. Models that specialize in modelling marketing seasonality, such as Prophet (Taylor & Letham, 2017) which uses Fourier series, can be built to model the seasonality. The seasonality output from the Prophet model could then be used in another model such as the MCMC as a representation of seasonality (Carlos et al., 2020).

The ANFIS model implemented in this thesis, used the Tagaki-Sugeno-Kang (TSK) inference system. (Shami et al., 2022) mention that interpretability wise, Tagaki-Sugeno-Kang inference systems are weaker, because regardless of whether they are trained using gradient descent or population-based algorithms such as PSO, they end up having high overlap area between the fuzzy partitions, resulting in a behaviour that is significantly different from system output. The alternative to TSK is mentioned to be Mamdani inference system (Shami et al., 2022), which is geared towards interpretability, however, suffers from weaknesses similar to that of gradient descent i.e., being stuck in local optimum, and hence not performing as well as TSK in forecasting.

Further research can also be done to find out if the TSK based neuro fuzzy system, through its high overlap ratio, ends up capturing the cross effects & interactions of media channels that have been difficult to quantify otherwise, due to which it manages to provide accurate forecasts.

Further analysis of the positive and negative behaviour of the skewed distributions obtained from the HMCMC model and the ANFIS model can also be conducted in order to better understand situations when the media channels made maximum contributions, so that those situations could be replicated.

8. References

Aly, S. & Vrana, I. (2018). Fuzzy expert marketing-mix model. Agricultural Economics (Zemědělská ekonomika). 51. 69-79. 10.17221/5079-AGRICECON.

Agrawal, P., Abutarboush, H.F., Ganesh, T., & Mohamed, A.W. (2021). Metaheuristic Algorithms on Feature Selection: A Survey of One Decade of Research (2009-2019). IEEE Access, 9, 26766-26791.

Ballnus, B., Hug, S., Hatz, K. (2017). Comprehensive benchmarking of Markov chain Monte Carlo methods for dynamical systems. BMC Syst Biol 11, 63. https://doi.org/10.1186/s12918-017-0433-1

Band, S. & Hadipoor, M. & Baghban, A. & Mosavi, A. & Bukor, J. & Varkonyi-Koczy, A. (2019). Developing ANFIS-PSO Model to Predict Mercury Emissions in Combustion Flue Gases. 10.20944/preprints201905.0124.v3.

Barp, Alessandro & Briol, Francois-Xavier & Kennedy, Anthony & Girolami, Mark. (2017). Geometry and Dynamics for Markov Chain Monte Carlo. Annual Review of Statistics and Its Application. 5. 10.1146/annurev-statistics-031017-100141.

Benesty, J. & Chen, J. & Huang, Y. & Cohen, I. (2009). Pearson Correlation Coefficient. 10.1007/978-3-642-00296-0_5.

Betancourt, M. (2017). A Conceptual Introduction to Hamiltonian Monte Carlo. Link: https://arxiv.org/abs/1701.02434

Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.

Blum, A. & Hopcroft, J. & Kannan,R. (2018). Foundations of. Data Science. Cambridge Universuty Press. Link: https://www.cs.cornell.edu/jeh/book.pdf

Broadbent, S., & Fry, T. (1995). Adstock Modelling for the Long Term. Market Research Society. Journal., 37(4), 1–18. https://doi.org/10.1177/147078539503700404

Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. Sociological Methods & Research, 33(2), 261–304. https://doi.org/10.1177/0049124104268644

Cain, P.M. (2010). Marketing Mix Modeling and Return on Investment. In: Kitchen, P.J. (eds) Integrated Brand Marketing and Measuring Returns. Palgrave Macmillan, London. https://doi.org/10.1057/9780230297340_4

Cain, P.M. (2014). Brand management and the marketing mix model. Journal of Marketing Analytics. 2. 10.1057/jma.2014.4

Cain, P.M. (2022). Modeling Short-and Long-Term Marketing Effects in the Consumer Purchase Journey. International journal of research in marketing 39.1 : 96–116. Web. Link: https://thearf-org-unifiedadmin.s3.amazonaws.com/ARF%20Member%20Only%20Events/2021.09.20_AUDIENCExSCIENCE/modeling-short-and-long-wp-.pdf

Carlos, R. U., Miguel, C. S., Anwer, M. & Davide, F. (2020). Exploring Marketing Mix Modeling (MMM) and Conversion Lift Experiment (CLE) blending. Deloitte Inc. Link: https://www2.deloitte.com/es/es/pages/strategy-operations/articles/combinacion-marketing-mix-modeling-y-conversion-lift-experiment-roas.html

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., & Riddell, A. (2017). Stan: A Probabilistic Programming Language. Journal of statistical software, 76, 1. https://doi.org/10.18637/jss.v076.i01

Catalão, J. & Pousinho, H.M.I. & Mendes, V.M.F. (2011a). Hybrid Wavelet-PSO-ANFIS approach for short-term electricity prices forecasting. Power Systems, IEEE Transactions on. 26. 137 - 144. 10.1109/TPWRS.2010.2049385.

Catalão, J. & Pousinho, H.M.I. & Mendes, V.M.F. (2011b). Short-term electricity prices forecasting in a competitive market by a hybrid PSO-ANFIS approach. Energy Conversion and Management. 52. 1061–1065. 10.1016/j.enconman.2010.08.035.

Chai, T. & Draxler, R.R.. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?– Arguments against avoiding RMSE in the literature. Geoscientific Model Development. 7. 1247-1250. 10.5194/gmd-7-1247-2014.

Chan, D., & Perry, M. (2017). Challenges and Opportunities in Media Mix Modeling. Google Inc. Link: https://storage.googleapis.com/pub-tools-public-publication-data/pdf/2d0395bc7d4d13ddedef54d744ba7748e8ba8dd1.pdf

Chen, A. & Au, T. (2019). Robust Causal Inference for Incremental Return on Ad Spend with Randomized Geo Experiments.

Chen, H. & Zhang, M. & Han, L. & Lim, A. (2021). Hierarchical marketing mix models with sign constraints, Journal of Applied Statistics, 48:13-15, 2944-2960, DOI: 10.1080/02664763.2021.1946020

Chenavaz, R/ & Escobar, O. & Rousset, X. (2019). An analytical framework for retailer price and advertising decisions for products with temperature-sensitive demand. Applied Economics. 51. 1-11. 10.1080/00036846.2019.1617398.

Chicco, D. & Warrens, M. & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science. 7. e623. 10.7717/peerj-cs.623.

Christopher, G. T. & Aloe, M. & Betsy, J. B. (2017), Extracting the Variance Inflation Factor and Other Multicollinearity Diagnostics from Typical Regression Results, Basic and Applied Social Psychology, 39:2, 81-90, DOI: 10.1080/01973533.2016.1277529

Colin, C. A. & Windmeijer, F. A. G. (1997). An R-squared measure of goodness of fit for some common nonlinear regression models. Journal of Econometrics, Elsevier, vol. 77(2), pages 329-342, April.

Cowling, D. & Modayil, M. & Stevens, C. (2010). Assessing the relationship between ad volume and awareness of a tobacco education media campaign. Tobacco control. 19 Suppl 1. i37-42. 10.1136/tc.2009.030692.

Curran-Everett, Douglas. (2018). Explorations in statistics: the log transformation. Advances in physiology education. 42. 343-347. 10.1152/advan.00018.2018.

Darzi, F. H., Khoshnazar N., & Eslami, M. (2015). Fuzzy time series forecasting Comparing Markov-Chain and percentage change models. 4th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS), Zahedan, Iran, 2015, pp. 1-4, doi: 10.1109/CFIS.2015.7391636.

Davidson, J. (2000). Econometric Theory, Blackwell Publishing Ltd, Oxford.

Desboulets, L. (2018). A Review on Variable Selection in Regression Analysis. Econometrics, 6(4), 45. https://doi.org/10.3390/econometrics6040045

Ding, J., Tarokh, V., & Yang, Y. (2018). Model selection techniques -- an overview. St. Louis: Federal Reserve Bank of St Louis. Retrieved from https://ezproxy.cc.lut.fi/working-papers/model-selection-techniques-overview/docview/2126056781/se-2

Dinner, I. M., Heerde Van, H. J., & Neslin, S. A. (2014). Driving Online and Offline Sales: The Cross-Channel Effects of Traditional, Online Display, and Paid Search Advertising. Journal of Marketing Research, 51(5), 527–545. https://doi.org/10.1509/jmr.11.0466

Douglas, B. & Hubert, G. (2010). Market Response and Marketing Mix Models: Trends and Research Opportunities. Foundations and Trends® in Marketing: Vol. 4: No. 3, pp 129-207. http://dx.doi.org/10.1561/1700000015

Drossos, D., Giaglis, G.M. (2005). Factors That Influence the Effectiveness of Mobile Advertising: The Case of SMS. In: Bozanis, P., Houstis, E.N. (eds) Advances in Informatics. PCI 2005. Lecture Notes in Computer Science, vol 3746. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11573036_26

Edward R. M. & Billy P. H. (1982). Detecting Multicollinearity. The American Statistician. 36:3a, 158-160, DOI: 10.1080/00031305.1982.10482818

Enders, W. (2004). Applied Econometric Time Series, 2nd edition, John Wiley & Sons: Hoboken.

Fareniuk Y. & Chornous, G. (2023). Optimization of Media Strategy via Marketing Mix Modeling in Retailing. Ekonomika, 102(1), pp. 6-25. doi: 10.15388/Ekon.2023.102.1.1.

Farris, P. & Hanssens, D. & Lenskold, J. & Reibstein, D. (2015). Marketing Return on Investment: Seeking Clarity for Concept and Measurement. Applied Marketing Analytics. 1. 267-282.

Filzmoser, P. & Nordhausen, K. (2020). Robust linear regression for high-dimensional data: An overview. Wiley Interdisciplinary Reviews: Computational Statistics. 13. 10.1002/wics.1524.

Freedman, D. A. (2009). On regression adjustments in experiments with several treatments. The Annals of Applied Statistics, 3(1), 166-189.

Gaur, J. & Bharti, K. (2020). ATTRIBUTION MODELING IN MARKETING: LITERATURE REVIEW AND RESEARCH AGENDA. Academy of Marketing Studies Journal. 24. 1-2

Gelman, A. & Rubin, B. (1992). Inference from Iterative Simulation Using Multiple Sequences. Statist. Sci. 7 (4) 457 - 472. https://doi.org/10.1214/ss/1177011136

Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B. (2004). Bayesian Data Analysis. Chapman and Hall/CRC. Link: https://users.aalto.fi/~ave/BDA3.pdf

Ghasemi, E., Kalhori, H. & Bagherpour, R. (2016) A new hybrid ANFIS–PSO model for prediction of peak particle velocity due to bench blasting. Engineering with Computers 32, 607–614. https://doi.org/10.1007/s00366-016-0438-1

Ghordoyee, M. S. & Roozbahani, A. & Arya, A. N. & Javadi, S. (2021). Development of Adaptive Neuro Fuzzy Inference System –Evolutionary Algorithms Hybrid Models (ANFIS-EA) for Prediction of Optimal Groundwater Exploitation. Journal of Hydrology. 598. 126258. 10.1016/j.jhydrol.2021.126258.

Gilardi, G. (2020). ANFIS(code_python). GitHub repository. Link: https://github.com/gabrielegilardi/ANFIS/tree/master

Gilardi, G., Haslam, E., Bundhoo, V., & Park, E. (2010). A shape memory alloy based tendon-driven actuation system for biomimetic artificial fingers, part II: Modelling and control. Robotica, 28(5), 675-687. doi:10.1017/S0263574709990324

Goodfellow, I., Bengio, Y., Courville, A. (2016). Deep Learning. MIT Press.

Gosiewska, A. & Kozak, A. & Biecek, P.. (2021). Simpler is better: Lifting interpretability-performance trade-off via automated feature engineering. Decision Support Systems. 150. 113556. 10.1016/j.dss.2021.113556.

Gujarati, D. N., & Porter, D. C. (2009). Basic Econometrics (5th Edition). McGraw-Hill Education.

Hasanipanah, M. & Amnieh, H. & Arab, H. & Zamzam, M. (2018). Feasibility of PSO–ANFIS model to estimate rock fragmentation produced by mine blasting. Neural Computing and Applications. 30. 10.1007/s00521-016-2746-1.

Hawkins, D. (2004). The Problem of Overfitting. Journal of chemical information and computer sciences. 44. 1-12. 10.1021/ci0342472.

He, S. (2020). Python/STAN Implementation of Multiplicative Marketing Mix Model (code_python). GitHub repository. Link: https://github.com/sibylhe/mmm_stan

Heinze, G. & Schilhart-Wallisch, C. & Dunkler, D. (2018). Variable selection - A review and recommendations for the practicing statistician. Biometrical Journal. 60. 10.1002/bimj.201700067.

Hoffman, M. & Gelman, A. (2011). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research. 15.

Hongyue, W. & Lu, N. & Tu, X. (2013). Log transformation: Application and interpretation in biomedical research. Statistics in medicine. 32. 10.1002/sim.5486.

Honkela, A. (2020). Computational Statistics I. University of Helsinki. Link: https://www.cs.helsinki.fi/u/ahonkela/teaching/compstats1/book/hamiltonian-monte-carlo-hmc.html

Hubert, G. (1993). Chapter 15 Marketing-mix models. Handbooks in Operations Research and Management Science.

Jang, J.S.R., Sun, C.T. and Mizutani, E. (1997). A computational approach to learning and machine intelligence. Neuro-Fuzzy and Soft Computing, Upper Saddle River, NJ: Prentice-Hall.

Jennrich, R. I. & Sampson, P. F. (1968). Application of Stepwise Regression to Non-Linear Estimation, Technometrics, 10:1, 63-72, DOI: 10.1080/00401706.1968.10490535

Jiang, H. & Kwong, C.K. & Siu, K. W. M. & Liu, Y. (2015). Rough set and PSO-based ANFIS approaches to modeling customer satisfaction for affective product design. Advanced Engineering Informatics. 29. 10.1016/j.aei.2015.07.005.

Jin, Y., Yueqing, W., Sun, Y., Chan, D., & Koehler, J. (2017). Bayesian Methods for Media Mix Modeling with Carryover and Shape Effects. Google Research. Link: https://research.google/pubs/pub46001/

John D. C. L. (1975). BRANDAID: A Marketing-Mix Model, Part 1: Structure. Operations Research, vol. 23, no. 4, pp. 628–55. JSTOR, http://www.jstor.org/stable/169847. Accessed 29 May 2023.

John D. C. L. (1975). BRANDAID: A Marketing-Mix Model, Part 2: Implementation, Calibration, and Case Study. Operations Research, vol. 23, no. 4, 1975, JSTOR, http://www.jstor.org/stable/169848. Accessed 29 May 2023.

Kappe, E. & Stadler B., A. & Desarbo, W. (2018). A random coefficients mixture hidden Markov model for marketing research. International Journal of Research in Marketing. 35. 10.1016/j.ijresmar.2018.07.002.

Kassa, Y. & Zhang, J. & Zheng, D. (2020). EMD-PSO-ANFIS based hybrid approach for short-term load forecasting in microgrids. IET Generation Transmission & Distribution. 14. 470 - 475. 10.1049/iet-gtd.2019.0869.

Kasuya, E. (2018). On the use of r and r squared in correlation and regression. Ecological Research. 34. 10.1111/1440-1703.1011.

Keene, O.N. (1995). The log transformation is special. Statistics in medicine, 14 8, 811-9 .

Kim, J. (2019). Multicollinearity and misleading statistical results. Korean Journal of Anesthesiology. 72. 10.4097/kja.19087.

Koenig-Lewis, N. & Bischoff, E. (2005). Seasonality research: The state of the art. International Journal of Tourism Research - INT J TOUR RES. 7. 10.1002/jtr.531.

Kumar, A. & Shankar, R. & Alijohani, N. (2019). A big data driven framework for demand-driven forecasting with effects of marketing-mix variables. Industrial Marketing Management. 90. 10.1016/j.indmarman.2019.05.003.

Kumar, U. (2005). Comparison of neural networks and regression analysis: A new insight. Expert Systems with Applications. 29. 424-430. 10.1016/j.eswa.2005.04.034.

Liu, Y., Laguna, J. & Wright, M. (2014). Media Mix Modeling – A Monte Carlo Simulation Study. Journal of marketing analytics 2.3 : 173–186. Web. Link: https://link.springer.com/content/pdf/10.1057/jma.2014.3.pdf

Shami, T. & El-Saleh, A. & Alswaitti, M. & Al-Tashi, Q. & Summakieh, A. & Mirjalili, S. (2022). Particle Swarm Optimization: A Comprehensive Survey. IEEE Access. PP. 1-1. 10.1109/ACCESS.2022.3142859.

Mahdevari, S. & Khodabakhshi, M. B. (2021). A hybrid PSO-ANFIS model for predicting unstable zones in underground roadways. Tunnelling and Underground Space Technology. 117. 104167. 10.1016/j.tust.2021.104167.

Manchanda, P., Rossi, P. E., & Chintagunta, P. K. (2004). Response Modeling with Nonrandom Marketing-Mix Variables. Journal of Marketing Research, 41(4), 467–478. https://doi.org/10.1509/jmkr.41.4.467.47005

Medar, R., Rajpurohit, V. S. & Rashmi, B. (2017). "Impact of Training and Testing Data Splits on Accuracy of Time Series Forecasting in Machine Learning". International Conference on Computing, Communication, Control and Automation (ICCUBEA), Pune, India, 2017, pp. 1-6, doi: 10.1109/ICCUBEA.2017.8463779.

Méndez-Suárez, M. & Estevez, M. (2016). Cálculo del ROI de marketing en modelos de marketing mix, del ROMI, al valor creado del marketing para los accionistas EVAM. Universia Business Review. 52. 46.

Méndez-Suárez, M. (2021). Marketing Mix Modeling Using PLS-SEM, Bootstrapping the Model Coefficients. Mathematics. 9(15):1832. https://doi.org/10.3390/math9151832

Meta Platforms (2023). Robyn. A New Generation of Marketing Mix Modeling. Link: https://facebookexperimental.github.io/Robyn/

Michael J. W. Sr. & John C. (2011) Marketing Mix Modeling for the Tourism Industry: A Best Practices Approach, International Journal of Tourism Sciences, 11:1, 1-15, DOI: 10.1080/15980634.2011.11434633

Miron, J. A., & Zeldes, S. P. (1988). Seasonality, Cost Shocks, and the Production Smoothing Model of Inventories. Econometrica, 56(4), 877–908. https://doi.org/10.2307/1912703

Monnahan, C.C., Thorson, J.T. and Branch, T.A. (2017). Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo. Methods Ecol Evol, 8: 339-348. https://doi.org/10.1111/2041-210X.12681

Mottahedi, A. & Sereshki, F.& Ataei, M. (2018). Overbreak prediction in underground excavations using hybrid ANFIS-PSO model. Tunnelling and Underground Space Technology. 80. 1-9. 10.1016/j.tust.2018.05.023.

Myers, D. (1989). To be or not to be... stationary? That is the question. Mathematical Geology. 21. 347-362. 10.1007/BF00893695

Naik, P. & Raman, K. & Winer, R. (2005). Planning Marketing-Mix Strategies in the Presence of Interaction Effects. Marketing Science. 24. 25-34. 10.1287/mksc.1040.0083.

Nandy, A., Basu, A., & Ghosh, A. (2021). Robust inference for skewed data in health sciences. Journal of applied statistics, 49(8), 2093–2123. https://doi.org/10.1080/02664763.2021.1891527

Narayanan, S. & Desiraju, R. & Chintagunta, P. (2004). Return on Investment Implications for Pharmaceutical Promotional Expenditures:The Role of Marketing-Mix Interactions. Journal of Marketing - J MARKETING. 68. 90-105. 10.1509/jmkg.68.4.90.42734.

Nazir, Z., Kaldykhanov, D. K., Tolep, K. and Park, J. (2021). A Machine Learning Model Selection considering Tradeoffs between Accuracy and Interpretability. 13th International Conference on Information Technology and Electrical Engineering (ICITEE), Chiang Mai, Thailand, 2021, pp. 63-68, doi: 10.1109/ICITEE53064.2021.9611872.

Neijens, P. & Voorveld, H. (2015). Cross-Platform Advertising: Current Practices and Issues for the Future: Why the Gross Rating Point Metric Should Thrive in Today's Fragmented Media World. Journal of Advertising Research. 55. 362-367. 10.2501/JAR-2015-016.

Ng, E., Wang, Z., Dai, A. (2021). Bayesian Time Varying Coefficient Model with Applications to Marketing Mix Modeling. arXiv. Lin: https://arxiv.org/abs/2106.03322

Nishio, M. & Arakawa, A. (2019). Performance of Hamiltonian Monte Carlo and No-U-Turn Sampler for estimating genetic parameters and breeding values. Genet Sel Evol 51, 73 https://doi.org/10.1186/s12711-019-0515-1

Noushabadi, A. S. & Dashti, A. & Raji, M. & Zarei, A. & Mohammadi, A. H. (2020). Estimation of cetane numbers of biodiesel and diesel oils using regression and PSO-ANFIS models. Renewable Energy, Elsevier, vol. 158(C), pages 465-473.. Link: https://ideas.repec.org/a/eee/renene/v158y2020icp465-473.html

Okokpujie, I.P., Sinebe, J.E. (2023). An overview of the study of ANN-GA, ANN-PSO, ANFIS-GA, ANFIS-PSO and ANFIS-FCM predictions analysis on tool wear during machining process. Journal Européen des Systèmes Automatisés, Vol. 56, No. 2, pp. 269-280. https://doi.org/10.18280/jesa.560212

Pappa, E. & Kontogeorgis, G. & Garefalakis, A. & Kyriakogkonas, P. (2022). APPLYING A MARKETING MIX MODEL FOR COMPETITIVE STRATEGIC MANAGEMENT. Actual Problems of Economics. 10-11. 16 -29. 10.32752/1993-6788-2022-1-256-257-16-29.

Philippe, A. & Broz-Giroux, A. (2013). Modeling advertising impact at campaign level: Empirical generalizations relative to long-term advertising profit contribution and its antecedents. Marketing Letters. 25. 10.1007/s11002-013-9252-3.

Ponzini, G. T., & Schofield, C. A. (2019). Randomized controlled trials evaluating the effectiveness of a direct-to-consumer marketing campaign for social anxiety disorder. Stigma and Health, 4(4), 450–461. https://doi.org/10.1037/sah0000163

Qian, Y., & Xie, H. (2011). No Customer Left Behind: A Distribution-Free Bayesian Approach to Accounting for Missing Xs in Marketing Models. Marketing Science, 30(4), 717–736. http://www.jstor.org/stable/23012021

Ramachandran, K.M. & Tsokos, C.P. (2009). Mathematical Statistics with Applications. Academic Press.

Rettie, R., Grandcolas, U. & Deakins, B. (2005). Text message advertising: Response rates and branding effects. Journal of Targeting, Measurement and Analysis for Marketing. Volume 13, pages 304–312. https://doi.org/10.1057/palgrave.jt.5740158

Robert, C. & Elvira, V. & Tawn, N. & Wu, C. (2018). Accelerating MCMC algorithms. Wiley Interdisciplinary Reviews: Computational Statistics. 10. 10.1002/wics.1435.

Sam, K. M. and Chatwin, C. (2019). Understanding Chinese gamblers' adoption of online casinos based on e-marketing mix model. The Journal of Gambling Business and Economics, 12 (2). pp. 67-87. ISSN 1751-8008

Schröer, Christoph & Kruse, Felix & Marx Gómez, Jorge. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. Procedia Computer Science. 181. 526-534. 10.1016/j.procs.2021.01.199.

Shamshirband, S., Hadipoor, M., Baghban, A., Mosavi, A., Bukor, J., & Várkonyi-Kóczy, A. (2019). Developing an ANFIS-PSO Model to Predict Mercury Emissions in Combustion Flue Gases. Mathematics, 7(10), 965. https://doi.org/10.3390/math7100965

Shihab, K. V. & Pillai, G. (2018). Recent advances in Neuro-fuzzy system: A survey. Knowledge-Based Systems. 152. 10.1016/j.knosys.2018.04.014.

Stevenson, M. & Porter, J. (2009). Fuzzy Time Series Forecasting Using Percentage Change as the Universe of Discourse. Proceedings of World Academy of Science, Engineering and Technology. 55.

Suardika, I. K., & Sari D., M. (2021). The Impact of Brand, Product Quality and Price on Sales Volume of Samana Mart Stores. International Journal of Social Science and Business, 5(2), 256–261. https://doi.org/10.23887/ijssb.v5i2.34766

Subbey, S. & Haugland, D. (2008). The Hamiltonian Monte Carlo Algorithm in Parameter Estimation and Uncertainty. ECMOR 2008 - 11th European Conference on the Mathematics of Oil Recovery. 10.3997/2214-4609.20146437.

Sun, Y. & Wang, Y. & Jin, Y. & Chan, D. & Koehler, J. (2017). Geo-level Bayesian Hierarchical Media Mix Modeling. Google Research. Link: https://research.google/pubs/pub46000/

Tariq, M.H., Abbas, T., Abrar, M.F., & Iqbal, A. (2017). EWOM and brand awareness impact on consumer purchase intention: mediating role of brand image.

Taylor, S.J., Letham, B. (2017). Forecasting at scale. Facebook Inc. PeerJ Preprints 5:e3190v2 https://doi.org/10.7287/peerj.preprints.3190v2

Think with Google (2022). The MMM Handbook. A Guide for Developing Impactful Marketing Mix Models. Link: https://www.thinkwithgoogle.com/_qs/documents/17950/The_MMM_Handbook.pdf

Thorsten, T. & Herbert, B. (2007) Jarque–Bera Test and its Competitors for Testing Normality – A Power Comparison, Journal of Applied Statistics, 34:1, 87-105, DOI: 10.1080/02664760600994539

Varian, H. & Choi, H. (2009). Predicting the Present with Google Trends. Economic Record. 88. 10.2139/ssrn.1659302.

Varma, S., & Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics, 7(1), 91.

Wagenmakers, E.J., Farrell, S. (2004). AIC model selection using Akaike weights. Psychonomic Bulletin & Review 11, 192–196. https://doi.org/10.3758/BF03206482

Witte, A., Baardman, M. & Weide, W. (2022). Unified Marketing Measurement - The Power of Blending Methodologies. Think With Google. Link: https://www.thinkwithgoogle.com/_qs/documents/13385/TwGxOP_Unified_Marketing_Measurement.pdf

Xu, Y., & Goodacre, R. (2018). On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. Journal of analysis and testing, 2(3), 249–262. https://doi.org/10.1007/s41664-018-0068-2

Yamashita, T. & Yamashita, K. & Kamimura, R. (2007). A Stepwise AIC Method for Variable Selection in Linear Regression. Communications in Statistics—Theory and Methods. 36. 2395-2403. 10.1080/03610920701215639.

Yoo, B. & Donthu, N. & Lee, S.. (2000). An Examination of Selected Marketing Mix Elements and Brand Equity. Journal of The Academy of Marketing Science - J ACAD MARK SCI. 28. 195-211. 10.1177/0092070300282002.

Zanias, G.P. (1994). The long run, causality, and forecasting in the advertising-sales relationship. J. Forecast., 13: 601-610. https://doi.org/10.1002/for.3980130704

Zantedeschi, D., Feit, E.M., & Bradlow, E.T. (2016). Measuring Multi-Channel Advertising Effectiveness Using Consumer-Level Advertising Response Data. Management Science, 63, 2706.

Zheng, A. & Casari, A. (2018). Feature engineering for machine learning : principles and techniques for data scientists. First edition. O'Reilly Media, Inc.

Zhong-guo, Z. (1997). Forecasting sales and price for existing single-family homes: A VAR model with error correction. The Journal of Real Estate Research, 14(1), 155-167. Retrieved from https://ezproxy.cc.lut.fi/scholarly-journals/forecasting-sales-price-existing-single-family/docview/200268858/se-2

Zhou, J., Grazian, C. & Ormerod, J. (2023). Skew-Normal Posterior Approximations. arXiv.
Link: https://arxiv.org/abs/2302.08614

Appendix 1. Publishing this thesis

This thesis will be confidential for a limited period of time of 2 years as per LUT policy.

Appendix 2. Github Link for Stored Model Results & Code

Case-company data is confidential hence only exported model results and code is shared.

https://github.com/mujtabaanwer/ThesisLUT.git

Appendix 3. Dependencies for Hamiltonian MCMC Simulations using STAN

During this development of this thesis, considerable amount of time was spent in figuring out the correct dependencies for the Hamiltonian MCMC built using STAN without any errors. The correct dependencies were discovered using online research and trial and error. The efforts failed on Windows operating system despite consulting DevOps practitioners. Due to this, a Macbook (with 32GB RAM) was used for implementation. The following dependencies were installed:

- Xcode Compiler for Mac
- Specific Anaconda Python Version and Release. Note that other versions did not enable PyStan to execute code for Hamiltonian MCMC at all.
    - Anaconda3-2020.07-MacOSX-x86_64.pkg 462.3M 2020-07-23 12:16:42 e095c487d2837e4c984d0fcd2217be42c615504a2d7d6651095d34c2fa9807a0
- conda install -c conda-forge pystan
- conda install setuptools
- install brew from command line
- brew install gcc