Bond University

DOCTORAL THESIS

Quantifying Advertising Media Effectiveness: Insights from Data-Driven Modelling

Johnman, Mark

Award date:
2021

Quantifying Advertising Media Effectiveness: Insights from Data-Driven Modelling

Mark Johnman

Submitted in total fulfilment of the requirements of the degree of

Doctor of Philosophy (PhD)

May 2021

Bond Business School

Supervisors: Professor Bruce Vanstone and Associate Professor Adrian Gepp

This research was supported by an Australian Government Research Training Program Scholarship. This research was also supported by Rapid Media Pty Ltd as part of an Innovation Connections Grant ICG000569 from the Australian Government Department of Industry, Innovation and Science.

Abstract

"Half the money I spend on advertising is wasted; the trouble is I don't know which half" – John Wanamaker

Organisations invest considerable resources in advertising, with global spending on marketing in 2019 alone estimated at $618.7 billion (Handley 2019). In an increasingly competitive and noisy environment, accurately measuring the effectiveness of a company's media channels and optimizing the resources invested in them has become a critical business problem. Businesses that can understand the effects of their advertising are well-positioned to improve the return on their marketing investments and gain a competitive advantage. However, developing and implementing data-driven solutions to quantify advertising media effectiveness is a challenging task. There is an increasing variety of data that can be collected about consumers' interaction with media channels, such as display clicks, social media interactions and website page views. This leads to questions about how to best incorporate these consumer activity metrics, which are often used to represent owned and earned media channels, into models that quantify advertising media effectiveness. Additionally, many marketing managers struggle to interpret the results of statistical modelling and relate them to practical marketing calculations, such as media channel return on investment (ROI) and budget allocations.

This research advances the field of marketing analytics by using consumer and firm activity data to develop statistical models that marketing managers can use to measure advertising media effectiveness. Using data provided by an Australian media agency, this research connects the outputs of these models with practical marketing calculations, such as media channel ROI and budget allocation recommendations. This provides marketing managers with a methodology they can use to better understand the effectiveness of their advertising and to optimize their advertising budget allocations. This research finds that disaggregating marketing performance metrics, as well as including consumer activity metrics and indirect effects in advertising media effectiveness models, reveals a more complete picture of the marketing environment. Including these relationships can enhance the accuracy of media channel ROI and budget allocation calculations.

This research also investigates how data used in advertising media effectiveness models can be applied in a different context to help address non-marketing business

problems. In particular, this research examines how advertising data can be used to improve demand forecasts, which are important for strategic and operational planning. Producing more accurate demand forecasts can help companies become more profitable, efficient and effective. The results show that including advertising spend and calendar-based variables in time-series models can improve the accuracy of demand forecasts. Moreover, using relatively simple models augmented with exogenous variables can produce more accurate forecasts than more complex pure time-series models.

Since demand forecasts are often organised in a hierarchical structure, this research also examines how the performance of bottom-up and aggregate approaches to demand forecasting change across forecast horizons and as more data are collected. The findings suggest that when there is less data available, an aggregate forecasting approach produces more accurate forecasts, while a bottom-up approach becomes more accurate with more data. The results led to the development of an extension to time-series cross validation that reduces the sensitivity of results to the number of observations used in the initial training subset. This new technique, called repeated time-series cross validation (RTSCV), offers a more comprehensive way to assess time-series model performance.

Overall, this research leverages consumer and firm activity data to answer important business questions and improve strategic decision-making. The findings help researchers and industry professionals better interpret and translate advertising effectiveness models into practice. Tools are provided to help practitioners to better understand the effects of advertising and optimize marketing resources. Finally, this research demonstrates how marketing departments and media agencies can harness their data to provide additional value to clients by using it to address other business problems.

Keywords

Advertising media effectiveness, persistence modelling, consumer activity metrics, demand forecasting, hierarchical modelling, time-series cross validation, data-driven modelling

Declaration by Author

This thesis is submitted to Bond University in fulfilment of the requirements of the degree of Doctor of Philosophy (PhD).

This thesis represents my own original work towards this research degree and contains no material that has previously been submitted for a degree or diploma at this University or any other institution, except where due acknowledgement is made.

Name: Mark Johnman

Date: May 2021

Research Outputs

Peer-Reviewed Publications

These publications are not directly related to the dissertation, but were published during the Ph.D.

Johnman, Mark, Vanstone, Bruce James and Gepp, Adrian. 2018. Predicting FTSE 100 Returns and Volatility Using Sentiment Analysis. *Accounting and Finance*. 58: 253-274.

Johnman, Mark, Vanstone, Bruce James and Gepp, Adrian. 2019. Harnessing Investor Sentiment Using Big Data Analytics. *The Australasian Journal of Applied Finance*. 3. https://www.finsia.com/news-hub/ajaf/harnessing-investor-sentiment-using-big-data-analytics.

Conference Presentations

Johnman, Mark, Gepp, Adrian, & Vanstone, Bruce James. 2019. Using Customer Information and Bayesian Techniques to Enhance Persistence Modelling. *41st Annual ISMS Marketing Science Conference*. Rome, Italy. 87. https://www.abstractsonline.com/pp8/#!/6819/presentation/1381.

Ethics Declaration

The research associated with this thesis received ethics approval from the Bond University Human Research Ethics Committee. Ethics application number MJ00975.

Acknowledgements

I would like to thank Professor Bruce Vanstone and Dr Adrian Gepp for their support and mentoring throughout my Ph.D.

I would like to thank Rapid Media Pty Ltd for their support during my Ph.D. and for enabling me to work with industry data.

I would like to acknowledge Mitchell Horrocks and Sarah Tait for their help in designing my Research Week poster. I would also like to thank Robert Wrathall for the free supply of chocolate.

To my friends and family, thank you for your encouragement and support throughout this journey. Without you, I would not be the person that I am today.

I would also like to thank the Bond Business School for supporting my academic education and providing me with teaching and research opportunities throughout my Ph.D.

Finally, thank you God for taking me on this journey. Following you is the best decision I ever made.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

This research was conducted using client data provided by Rapid Media, an Australian media and communications agency. This represented a unique opportunity to conduct academic research with an industry partner, which will help ensure that the findings of this research are relevant to industry. It also answers the call of Wedel and Kannan (2016) for close collaborations between companies and academics to develop marketing models. This chapter lays the foundations of the dissertation. It introduces the research context and problem, along with the research questions and contributions. It closes by providing an overview of the structure of the remainder of the dissertation.

1.1 Background to the Research Problem

Firms around the world make substantial financial investments and exert considerable effort in attempts to influence consumer behaviour. The World Advertising and Research Center (WARC) estimated that $618.7 billion was spent globally on marketing in 2019 alone (Handley 2019). In essence, marketing seeks to stimulate favourable customer attitudes in order to increase revenue, and ultimately, the profit of the firm (Hanssens and Pauwels 2016). This is achieved through a wide variety of tools and tactics, which can be grouped under one of the four categories of the Four Ps framework developed by McCarthy (1960):

- **Product** – the good or service being offered to the customer.
- **Price** – decisions influencing how much a customer pays for the product.
- **Place** – how the product is distributed, such as via an online shopping website or an offline retail store.
- **Promotion** – marketing communications activities used to persuade the consumer to purchase the product, such as paid advertising, public relations and sales promotions.

The Four Ps framework outlines the marketing mix that marketers use in their attempts to build relationships with customers and increase company profits. The implementation of the promotion arm of the marketing mix was traditionally conducted through offline media channels, such as newspapers and television. However, the last twenty years has seen the emergence of, and increasing investment in, online media channels, such as paid search and social media. The rise of the internet has also caused the importance of earned and owned media channels to grow. Batra and Keller (2016) distinguish between paid, earned and owned media channels as follows:

- **Paid Media** – channels that a company pays for, with examples including television, radio and social media ads. These activities are organised internally within the company or outsourced to a third party, such as a media agency.
- **Owned Media** – marketing communications activities carried out in channels that are controlled by the company, such as their website and social media pages.
- **Earned Media** – marketing communications activities that are not carried out by the company or its agents, including online reviews, press coverage and word-of-mouth.

The media channels used by a company, along with the amount of funds allocated to them, depend on a wide variety of factors, including the size of the firm, target customers and the product or service being sold. When implementing marketing campaigns, companies often encounter substantial challenges, such as difficulties coordinating strategies across company departments and external media agencies. This is partly due to each stakeholder having their own responsibilities and objectives. Hanssens and Pauwels (2016) explain that this often leads to the use of different performance metrics among the stakeholders. For example, a digital advertising team might be aiming to minimise the cost per click (CPC), while the above-the-line marketing team may be seeking to maximise Gross Rating Points (GRPs). It can be difficult to reconcile these metrics and media channels into a cohesive system that measures their effectiveness at contributing to overall marketing objectives.

As a result, measuring the effectiveness of an organisation's media channels and optimizing the resources invested in them is an important problem to address. An insight of the importance of this to industry is given by the CEO of Inofec BV, an office equipment supplier analysed by Wiesel et al. (2011), who describes having a model to help allocate resources across media channels as a critical competitive advantage. The Marketing Science Institute (2018) also identifies measuring media efficacy as a key research priority for 2018 - 2020. Quantifying advertising media effectiveness involves producing outputs that describe how effective a media channel is at improving a marketing performance metric, such as sales or enquiries. These outputs are usually derived from statistical models that relate media channels to marketing performance metrics. This research refers to such models as advertising media effectiveness models.

In practice, there are several challenges that add to the complexity of developing advertising media effectiveness models. First, not all marketing campaigns have the same objectives. For example, some may be designed to drive short-term sales, while others may be trying to create brand awareness. Compounding the issue is the increasing complexity of the consumer’s path to purchase, which can involve a plethora of media channels spanning multiple devices and paid, earned and owned media categories (Batra and Keller 2016). This implies that marketers must measure the direct effects of marketing communications activities on marketing objectives, as well as their indirect effects via other media channels (Batra and Keller 2016). Finally, in practice, advertising media effectiveness models are only as good as the actionable and timely insights that they provide to marketers. Thus, it is important for these models to account for the different planning cycles of marketing communications activities (Kannan and Li 2017). For example, television campaigns need to be booked in advance, while online advertising campaigns can typically be implemented faster.

After model development, companies often encounter difficulties when implementing the results of advertising media effectiveness models. Data collection from disparate sources, as well as storage and organisation issues, can hinder the application and increase the cost of data analytics solutions. Moreover, many marketing managers do not have the marketing analytics expertise required to understand statistical modelling (Hanssens and Pauwels 2016). As a result, those who develop statistical models that measure advertising media effectiveness are often different from those who design and implement advertising campaigns (Hanssens and Pauwels 2016). This can lead to difficulties for marketing managers in interpreting the results of statistical modelling and relating them to practical marketing metrics, such as media channel return on investment (ROI) and cost per acquisition (CPA), as well as using them to refine the marketing budget. Marketing managers are also often reluctant to use data analytics solutions that give specific outputs, instead preferring to use decision support systems that allow them to see a range of options that can inform strategic decision-making (Albers 2012).

In summary, there are considerable challenges in quantifying advertising media effectiveness. However, the growing amount of data about consumers' interactions with media channels, such as webpage visits and Facebook post likes, provide opportunities to create statistical models that capture the complexity of the marketing environment (Wedel and Kannan 2016). This research aims to use this data to develop statistical models that accurately measure the effectiveness of the online and offline media channels a company uses. Furthermore, it seeks to connect outputs from these models to practical marketing calculations, such as media channel ROI and budget allocation recommendations, which can assist marketing managers in strategic decision-making.

This research also investigates how data used in advertising media effectiveness models can be utilised to improve solutions to non-marketing business problems. The emergence and use of data analytics is not limited to the marketing function of an organisation. Data analytics is playing an increasingly central role across all business functions as companies seek to harness both internal and external data to gain a competitive advantage. This presents opportunities for business functions to collaborate and leverage data from one another to improve their own decision-making processes. However, organisational silos can make it difficult to collect, interpret and integrate data from the different areas of a business. By investigating how advertising data can be used to improve solutions to other business problems, this research seeks to provide an example of the potential synergies from data-sharing between organisational functions. For marketing executives under increased pressure to demonstrate the economic value of the marketing department to the firm (Hanssens and Pauwels 2016), such an application provides a novel way the marketing team can provide additional value to the wider business. For media agencies, the ability to use existing advertising data to assist clients in solving additional business problems provides them with a competitive advantage in the marketplace.

1.2 Research Aims & Questions

Overall, this research aims to leverage consumer and firm activity data to answer important business questions and improve strategic decision-making. More specifically, this research aims to advance research and practice in the field of marketing analytics by:

- Using consumer and firm activity data to develop statistical models that companies can use to measure advertising media effectiveness;
- Translating these models into practice by connecting their outputs with practical marketing calculations, such as media channel ROI and budget allocation recommendations;
- Examining how data used in advertising media effectiveness models can be applied in a different context to help address non-marketing business problems.

In achieving these aims, this research produces findings that are applicable to marketing practitioners and researchers from a non-marketing background. The aims of this research are achieved by addressing four research questions, which are described in the remainder of this section.

Quantifying advertising media effectiveness is a well-established area of research. Historically, this problem has been addressed through econometric techniques, such as persistence modelling. Persistence modelling involves using multivariate time-series models, such as vector autoregressions (VARs), to capture the relationships between marketing variables. This approach focuses on letting the data dictate the relationships between marketing variables as opposed to predefining them. Other econometric solutions take a more customized approach, pre-defining the relationships between media channels and marketing performance metrics. Over the last decade, the growing amount of data about individual consumers' interactions with media channels has resulted in increased interest in more granular modelling approaches. This has seen the application of a diverse range of modelling techniques, such as survival models and Bayesian models, to investigate advertising media effectiveness at an individual-level. The increasing availability of individual-level data has also resulted in the emergence of new research streams, such as attribution modelling, which involves allocating credit to each of the advertisements a consumer encounters according to their contribution towards a desired outcome, such as a product sale.

The growth in data availability and rise of data analytics has seen the emergence of new modelling techniques and approaches to quantifying advertising media effectiveness. It is unclear how these advances relate and connect with each other. There are also questions around broader modelling issues, such as data collection and model assessment, as yet unanswered in the literature. For example, should individual-level or aggregate-level data be used to quantify advertising media effectiveness? While individual-level data can provide more granular insights into advertising media effectiveness, it can be challenging and costly for companies to collect data across multiple media channels and devices. Aggregated data, such as expenditure on individual media channels, on the other hand, are more readily obtainable by businesses, particularly those that do not have access to large financial resources (i.e. SMEs). Overall, the beginning of a new decade offers a timely moment to reflect upon and synthesise a research field's progress. It also provides an opportunity to create a consolidated platform that researchers from both marketing and non-marketing backgrounds can build on. This leads to the first research question:

- **Research Question 1 (RQ1)** – What is the state of the field of research in quantifying advertising media effectiveness over the past decade?

Addressing RQ1 provides the context needed to develop statistical models that organisations can use to measure advertising media effectiveness. Existing research demonstrates that using both online and offline media channels can produce both within-media and cross-media synergies (Pauwels et al. 2016b; Wiesel et al. 2011; Naik and Peters 2009). However, the numerous online and offline media channels available to choose from have made a customer's journey more akin to a network than the funnel traditionally used in the literature (Srinivasan et al. 2016). Additionally, the rise of owned and earned media channels has strengthened their ability to influence a company's financial performance, with free consumer and business content frequently being used by consumers in the purchase decision-making process. Despite this, advertising media effectiveness models do not always consider a company's full suite of media channels. In particular, when owned and earned media channels are considered, they are often included at the expense of paid media channels on the same platform (e.g. including a company's Facebook page, but not their Facebook advertising). Considering all paid, owned and earned media channels in models could help provide a clearer picture of advertising effectiveness. This would likely have ramifications for resource optimization, with the interactions between paid, owned and earned media channels potentially influencing the optimal budget that should be allocated to paid media channels.

Owned and earned media channels are usually represented in advertising effectiveness models with variables that capture data about consumers' interactions with them, such as

Facebook likes or website page views. Srinivasan et al. (2016) establish that these consumer activity metrics can serve as early performance indicators for brands. However, there is a lack of research showing how including them when measuring the effectiveness of paid media channels influences return on investment (ROI), cost per acquisition (CPA) and budget allocation calculations. Since these metrics are commonly used in industry for reporting and decision-making, connecting advertising effectiveness models to them is important so that their insights can be practically understood and implemented by marketing managers. Addressing RQ1 reveals that there is also an absence of research showing the implications of using disaggregated marketing performance metrics when calculating these metrics. For example, Dinner et al. (2014) consider online and offline performance measures, while Srinivasan et al. (2016) consider aggregated ones. Data collection, storage and organisation issues, as well as privacy concerns, generally make it easier to collect aggregate data. However, disaggregating data can reveal additional insights that are lost when the data are aggregated (Hyndman and Athanasopoulos 2018). This research addresses these issues through the following research question:

- **Research Question 2 (RQ2) -** What effects do consumer activity metrics and the level of aggregation in marketing performance metrics have on ROI, CPA and budget allocation calculations?

The third aim of this research is to examine how advertising data can be harnessed to address business problems typically outside the scope of the marketing department. More specifically, this research investigates how advertising data can be used to improve demand forecasts. Demand forecasting plays a crucial role in many businesses as it assists with operational and strategic decision-making. Producing accurate demand forecasts can help reduce the waste of perishable goods (Huber and Stuckenschmidt 2020), improve sales revenue and customer satisfaction (Fildes et al. 2019), as well as assist with long-term infrastructure planning (Ghalehkhondabi et al. 2019). The broad usage and wide-ranging importance of demand forecasting make it an important business problem. Addressing this problem also provides insight into how data used to measure advertising media effectiveness can be applied in a predictive context, as well as an explanatory one.

The demand forecasting literature contains mixed evidence of the effectiveness of including variables other than previous demand values in predictive models. In some circumstances, exogenous variables have been shown to improve model accuracy (Taieb and Hyndman 2014). However, other research finds that pure time-series approaches produce more accurate forecasts than ones with exogenous variables (Athanasopoulos et al. 2011). This has led Makridakis et al. (2020) to call for more empirical research into understanding the value of collecting exogenous data and determining the domains in which they practically improve forecasting performance. This leads to the third research question:

- **Research Question 3 (RQ3) -** To what extent can advertising spend and calendar-based indicator variables improve demand forecasts across different forecast horizons?

When investigating individual demand forecasts, it is also worthwhile considering how they perform in a hierarchical structure, such as individual store-level sales, city-wide sales and country-wide sales. Demand forecasts are often organised this way as different business departments require varying levels of granularity in forecasting. When producing demand forecasts for different levels within a hierarchy, it is usually important to ensure that they are coherent, or add up in a way that's consistent with the underlying data (Hyndman and Athanasopoulos 2018). This helps ensure that there is consistency in planning and execution across business departments. Different approaches to generating coherent forecasts include bottom-up forecasting and top-down forecasting. The bottom-up approach involves forecasting each bottom-level time series and adding up the results to produce forecasts for the higher levels. The top-down approach involves predicting the top-level time series directly and breaking these predictions down into forecasts for the lower-level time series using proportions. There is an absence of research into how these approaches perform across different forecast horizons, as well as how this might change as more data are collected and used in the model estimation process. These gaps in the literature are addressed by the fourth research question:

- **Research Question 4 (RQ4) -** How does the performance of bottom-up and aggregate approaches to demand forecasting change across forecast horizons and data size?

Overall, this research addresses important research issues that also have practical implications for industry. Addressing the four research questions will provide insights into how consumer and firm activity data can be leveraged in multiple contexts to improve strategic decision-making. The findings will also help researchers and industry professionals better interpret and translate advertising media effectiveness models into practice. Furthermore, this research provides tools to better understand the effects of advertising and optimize marketing resources.

1.3 Main Contributions of this Research

The overall aim of this Ph.D. is to leverage consumer and firm activity data to answer important business questions and improve strategic decision-making. In particular, this research advances the development of statistical models that companies can use to measure the effectiveness of their media channels. From a practical perspective, this research provides companies with a methodology that they can use to predict marketing performance metrics and better understand the effectiveness of their media channels. The insights from this process would help businesses to optimize the allocation of their marketing resources. This research also shows companies how they can connect the outputs of advertising media effectiveness models with practical marketing calculations, such as media channel ROI. This would assist marketing managers in interpreting and implementing the results of these models within their advertising campaigns. More broadly, this would help encourage the use of marketing analytics in organisational decision-making and improve the translation of advertising media effectiveness models into practice. Hanssens and Pauwels (2016) describe the benefits of increased use of marketing analytics as substantial, observing that organisations of any size and industry gain sustainable competitive advantages from using marketing analytics.

The research findings also have the potential to be generalised to other problems involving understanding the influence of multiple factors with carryover and interaction effects on performance metrics. For example, health organisations may be able to use insights from this research to assess the effectiveness of the variety of hospital touchpoints a patient is exposed to. Another field where this research may be useful is team sports analytics, where individual players could be considered as channels contributing to broader goals, such as winning or scoring a goal. Coaching staff would likely be interested in assessing how effective individual players are, along with their interaction effects on other teammates and competitors.

Additionally, this research advances the demand forecasting literature by showing how advertising data can be used to improve demand forecasts. This helps connect data analytics researchers from marketing and non-marketing backgrounds, which promotes cross-disciplinary research advances. It also demonstrates how data used in advertising media effectiveness models can be applied in a different context to help address non-marketing business problems. Practically, this promotes collaboration between marketing and other business functions, providing an example of how data in one function can be leveraged to assist another. Furthermore, it provides an example of how marketing departments and media agencies can use existing resources to add additional value to their clients. The following section summarises more specific contributions of the research according to each research question.

- **Research Question 1 (RQ1)** – What is the state of the field of research in quantifying advertising media effectiveness over the past decade?
    - Identifies the main methodological approaches used to quantify advertising media effectiveness and the requirements surrounding their use, as well as shows how they relate to each other.
    - Provides recommendations and future research suggestions regarding the data sources and model assessment techniques used to model advertising media effectiveness.
    - Outlines limitations, managerial implications and future research directions emerging from the literature.
    - Summarizes and synthesises the research field's progress over the past decade, creating a consolidated platform for future research. This makes the field more accessible to marketing practitioners and researchers from a non-marketing background.
    - Brings together individual-level and aggregate modelling approaches to quantifying advertising media effectiveness, pooling insights from persistence modelling and the newer attribution modelling literature. Understanding the nuances and types of advertising media effectiveness problems that these modelling approaches can address can help marketing managers derive clear insights and better use the information available to them.

- **Research Question 2 (RQ2) -** What effects do consumer activity metrics and the level of aggregation in marketing performance metrics have on ROI, CPA and budget allocation calculations?
    - Provides insight into the data marketing managers should be collecting when developing models that quantify advertising media effectiveness.
    - Shows how to connect advertising media effectiveness models with commonly used marketing calculations, such as media channel ROI and budget allocation calculations.
    - Presents a generalisable methodology for calculating the direct and indirect effects of each media channel and shows how to present this information in visualisations that are meaningful to marketing managers.
    - Addresses the call of Srinivasan et al. (2016) to investigate the extent to which their findings generalize by using a different product, data and media channel context. More specifically, an expanded number of online and offline media channels are included, higher frequency data are collected and enquiries is used instead of sales as the marketing performance metric.
    - Investigates the effects of including paid Facebook ads and owned Facebook page posts in the same model.
    - Examines the effects of including multiple metrics for paid search, namely advertising expenditure and engagement rate, in advertising media effectiveness models.
- **Research Question 3 (RQ3) -** To what extent can advertising spend and calendar-based indicator variables improve demand forecasts across different forecast horizons?
    - Addresses the call of Makridakis et al. (2020) for more empirical research into how exogenous variables can practically improve demand forecasts.
    - Provides insight into whether simpler models with advertising spend and calendar-based exogenous variables can outperform more complex pure time-series models. Since more complex time-series models are often harder to understand and implement in practice, it is

useful to investigate if comparative results can be achieved with simpler models augmented by exogenous variables.

- **Research Question 4 (RQ4) -** How does the performance of bottom-up and aggregate approaches to demand forecasting change across forecast horizons and data size?
    - Provides insights that can help organisations generate more accurate demand forecasts, which can improve profitability and productivity. For example, organisations may be able to produce more accurate demand forecasts by switching their forecasting approach as more data are collected and the forecast horizon increases, rather than adopting a one-size fits all approach.
    - Results in the development of an extension to time-series cross validation that reduces the effect of the number of data points used in the initial training subset on overall model performance. This new technique is termed repeated time-series cross validation (RTSCV) as it is analogous to repeated k-fold cross validation, an existing model evaluation technique.

1.4 Dissertation Structure

The remainder of this dissertation is structured as follows. The next chapter (Chapter 2) addresses RQ1 by presenting a review of the last decade of research into quantifying advertising media effectiveness. This review brings together individual-level and aggregate modelling approaches to quantifying advertising effectiveness. It outlines recommendations for data sources and model assessment, as well as discussing key findings and future research.

Based on the literature review, Chapter 3 presents the research that addresses RQ2. This chapter investigates how consumer activity metrics and aggregation levels of marketing performance metrics influence practical marketing calculations, such as the ROI or CPA of a media channel. The data, methodology and results of the analysis are outlined and synthesised into research and managerial implications.

The third and fourth research questions, RQ3 and RQ4, are addressed in Chapter 4. This chapter focuses on the use of advertising data in a predictive rather than an explanatory context. More specifically, this chapter examines whether including advertising spend and calendar-based variables can improve the accuracy of demand forecasts. It also explores how the performance of different approaches to predicting total demand respond to changes in data availability and forecast horizon. The data, methodology and results of this research are presented. Additionally, a new time-series model evaluation technique, repeated time-series cross validation (RTSCV), is developed and empirically evaluated.

Finally, conclusions of this research, along with future research directions, are discussed in Chapter 5.

Chapter 2 Quantifying Advertising Media Effectiveness

This chapter provides a systematic literature review of quantifying advertising media effectiveness over the past decade. A key finding is that individual-level and aggregate modelling approaches to quantifying advertising media effectiveness are rarely discussed together or compared. These approaches essentially address the same problem from different perspectives, with aggregate modelling approaches coming from a top-down perspective and individual-level modelling approaches coming from a bottom-up perspective. This chapter contributes to the literature by bringing together these two different approaches, pooling insights from persistence modelling and the newer attribution modelling literature. We provide recommendations regarding data sources and model assessment techniques, as well as identify limitations. Finally, we outline key findings emerging from the literature and call for more research into integrating aggregate and individual-levelling modelling approaches.

2.1 Introduction

The proliferation of advertising media channels and ways in which consumers can interact with them has made measuring their effectiveness increasingly complex. In today's marketing environment, the consumer path to purchase often resembles more of a network than a traditional hierarchical funnel (Srinivasan et al. 2016). While consumers may still go through different stages before purchasing, their media channel interactions are complex and non-linear, with it being common for consumers to return to media channels multiple times in their paths to purchase (Batra and Keller 2016).

Due to the sheer amount of content and competition in the marketplace, measuring the effectiveness of a firm's media channels and optimizing the resources invested in them has become increasingly important to gain a competitive advantage. This has been highlighted by the Marketing Science Institute (2018), which lists measuring media efficacy as a research priority for 2018 - 2020. At its core, this process involves using statistical models to relate media channels to marketing performance metrics, such as sales or conversions. The outputs of these models are used to derive measures that describe how effective each media channel is at improving the chosen marketing performance metric. We refer to these models as advertising media effectiveness models, and they can provide insights into how a company should allocate their resources among the available media channels.

The process of quantifying advertising media effectiveness has historically been conducted using econometrics techniques, such as persistence modelling. However, the increasing availability of individual-level data about a customer's exposure to media channels (touchpoints) has led to a greater focus on modelling approaches that extract value from this data. Over the last decade, this has resulted in the development of new modelling approaches, such as attribution modelling, defined by Jayawardane et al. (2015, 1) as "the matter of assigning credit [attribution value] to one or more advertising channels for influencing a desirable action (e.g. purchase, download or registration)".

To synthesise these developments, this chapter provides a systematic review of research into the process of quantifying advertising media effectiveness over the past ten years. The beginning of a new decade represents a timely opportunity to reflect upon and synthesise a research field's progress over the previous decade (Liu-Thompkins 2019). As discussed by Palmatier et al. (2018), review papers are useful as they provide researchers with a synthesised overview of a field that connects and reconciles the empirical pieces within it. This chapter complements the work of Liu-Thompkins (2019), who reviews online advertising research, as well as Mantrala and Kanuri (2016), who give a comprehensive overview of research into optimizing the allocation of marketing resources. As a point of difference, this review focuses on methodological approaches to calculating the effectiveness, or return-on-investment, of an organisation's advertising when conducted across multiple media channels. This chapter identifies the main methodological approaches used and the requirements surrounding their use, as well as showing how they relate to each other. In doing so, this chapter provides a more complete picture of the field and outlines managerial implications and directions for future research.

A key finding of this review is that individual-level modelling approaches are rarely discussed or compared with aggregate modelling approaches to quantifying advertising media effectiveness. This is despite these approaches essentially addressing the same issue from different perspectives, with aggregate modelling approaches coming from a top-down perspective and individual-level modelling approaches coming from a bottom-up perspective. Additionally, marketing managers are often confronted with a combination of both individual-level (e.g. clickstream data) and aggregate-level (e.g. TV expenditure) data in practice. Understanding the nuances and types of advertising media effectiveness problems that can be addressed by modelling approaches using these data can help marketing managers derive clear insights and better use the information available to them. From an academic perspective, bringing together well-established approaches to quantifying advertising media effectiveness (such as persistence modelling) with newer approaches based on individual-level data (such as attribution modelling) will enable faster progress and create a consolidated knowledge base for future research.

The rest of this chapter is organised as follows. First, we provide an overview of the systematic review process. Second, we synthesise the findings of the review process. In this discussion, we provide recommendations and future research suggestions regarding the data sources and model assessment techniques used to model advertising media effectiveness. Third, we outline limitations, managerial implications and future research directions emerging from the literature. In particular, we call for more research into integrating aggregate and individual-level modelling approaches to capture their respective strengths and offset their weaknesses.

2.2 Systematic Review Process

The sources for the literature review have been compiled and analysed using a systematic approach, following the process of Zupic and Čater (2015). We compiled the bibliometric data for the literature review in two stages, which are outlined below.

- **Stage 1 – Advertising Media Effectiveness Literature** – Two well-known databases, Web of Science (WoS) and Scopus, were searched most recently on May 02, 2020, for empirical articles published from 2010 to the present containing the following query in the title, abstract or keywords: *"marketing?mix model*" OR "media?mix model*" OR "media budget*" OR "optimal media budget*" OR "optimal media allocation" OR "marketing resource allocation" OR "optimal marketing polic*" OR (advertising "budget allocation") OR "profit impact of marketing*" OR "advertising elasticity" OR ("marketing effectiveness" elasticity) OR (measuring "advertising response") OR (marketing AND "attribution* *model*”)*[1]. The large number of search terms was used to capture syntactical variations of the same concept and were a consequence of the variety of keywords used in the literature. The search returned 231 unique records across Scopus and WoS. The abstracts in this list were then screened by the authors for relevance to quantifying advertising media effectiveness, reducing the list to 36 records. As this is a growing and diverse field, finding only 36 papers from the keyword set was not surprising. Most of the sources removed from the list are conceptual papers or focus on marketing resource optimization, rather than quantifying advertising media effectiveness. The data time period was chosen as the last decade has seen the development of new modelling approaches, such as attribution modelling, which have yet to be incorporated into existing advertising media effectiveness literature.

- **Stage 2 – Additional Sources** – 21 additional sources were manually added by following citations and from Google Scholar searches. This included two sources published by Google Research, three sources from attribution modelling review papers and seven from a prominent book chapter by Danaher (2017).

In total, the literature review includes 57 sources, 54 of which were indexed in the WoS or Scopus databases. The non-indexed sources are not published in academic journals and include a working paper (Abhishek et al. 2012) and two articles from Google Research (Chan and Perry 2017; Jin et al. 2017). The sources included in the literature review are tabulated in Appendix 3. After the sources were collected, the R package *bibliometrix*, developed by Aria and Cuccurullo (2017), was used to analyse the indexed sources. We used co-word analysis to identify the relationships and map out a conceptual structure of the field. Co-word analysis involves using the actual content of articles (e.g. keywords), rather than citations, as the unit of analysis (Zupic and Čater 2015). Figure 2-1 shows the relationships between the most commonly occurring keywords in the literature reviewed, with the line thickness indicating how often a pair of keywords have occurred together in an article and the colouring indicating groups of keywords that tend to occur together. Two distinct areas are apparent, with one focusing more on attribution modelling and online advertising, and the other consisting of alternative approaches to quantifying advertising media effectiveness. It is noted that these areas may have been

[1] A variety of other keyword combinations in addition to the ones presented in the thesis were trialled, with no material difference in results.

influenced by the manual selection process used in Stage 1 of the literature review. However, given the wide variety of publication sources that the literature review uncovered, it is unlikely that this influence is substantial.

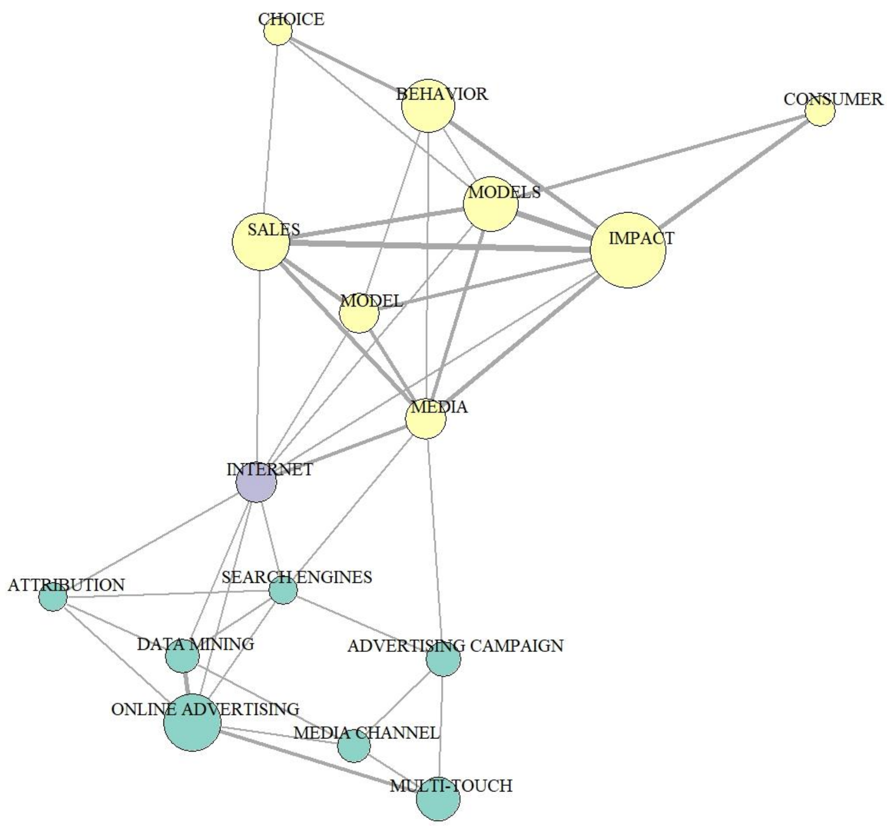


Figure 2-1 - Co-Occurrences of Top Keywords in the Literature Reviewed

There also seems to be a lack of consistency in the keywords and terminology used across the field, as shown in Figure 2-2, which limits comparability and could reduce the rate of the field's progress. The exception is attribution modelling literature as several of the keywords in the left panel of Figure 2-2 include the keyword "attribution". This contrasts with the right panel, which shows that the top 15 keywords across all of the non-attribution modelling sources are broad and varied. The lack of commonality across these keywords suggests that research into quantifying advertising media effectiveness lacks common terminology and that there is limited integration between research streams. It would help advance the field if research used common keywords and phrases. This would improve the organisation and accessibility of research, which would help researchers identify gaps, leverage findings from both streams of research and integrate approaches to pool their individual strengths. Researchers from other disciplines, such as data scientists, will also benefit, which will promote cross-disciplinary research and create additional contributions to the field.

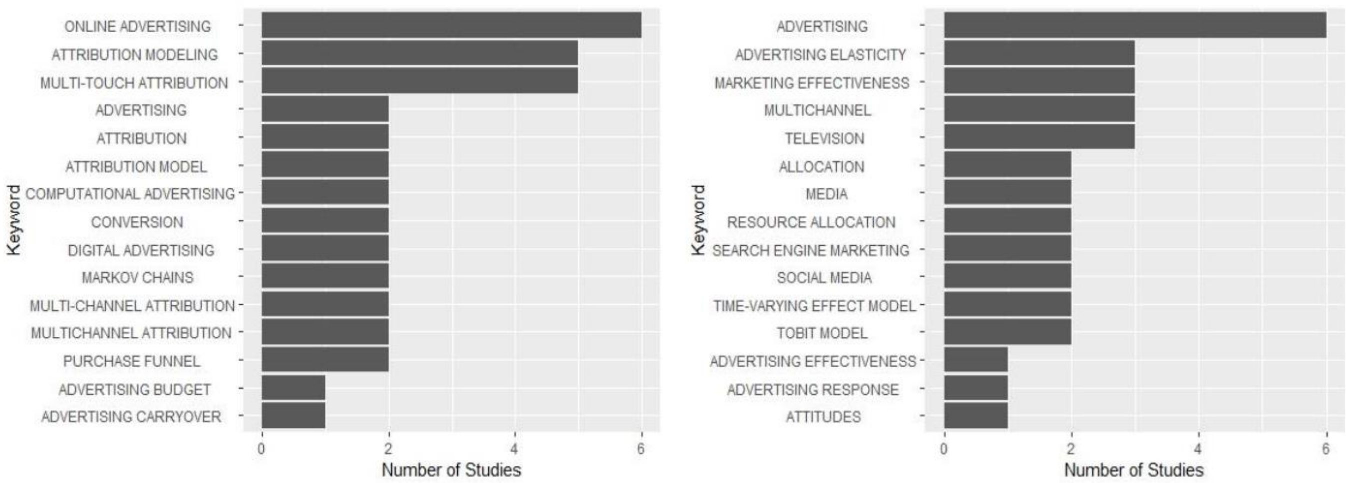


Figure 2-2 - Most Popular Keywords for Attribution Modelling Sources (left) vs Non-Attribution Modelling Sources (right)

2.3 Review of Advertising Media Effectiveness Research

While there are multiple modelling approaches to assessing the effectiveness of media channels, they all share the same underlying methodology: data collection, model development, model assessment and resource allocation.

2.3.1 Data Collection

The Data Collection stage involves gathering and preparing the data to be used in the advertising media effectiveness model. Due to its importance, we discuss data extensively in this review, highlighting issues and problems suited to both individual-level and aggregate-level data. The data used in the literature reviewed is consistent with that used in industry, with Nichols (2013), the co-founder of a marketing analytics firm that sold for $450m (USD), grouping the data into five main categories:

- **Company Information** – company sales and product data (e.g. pricing). This category usually contains the dependent variable of the advertising media effectiveness model, which is often a measure of financial performance (e.g. sales revenue) or conversions (e.g. number of sales).

- **Market Conditions** – exogenous variables that are uncontrolled by the company, but still influence customer purchase patterns. Examples include unemployment rates, consumer confidence measurements and weather patterns.
- **Marketing Activities** – data about a company's media channels, including expenditure, timing, location and advertisement content.
- **Customer Information** – customer details (e.g. demographics) and their interactions with media channels (e.g. clicks on an ad).
- **Competitor Information** – data about the behaviour of a company's competitors (e.g. advertising expenditure, product pricing).

Each category contains important information that can help advertising media effectiveness models be more reflective of reality and reduce omitted variable bias. However, this review finds that some literature ignores some of these categories. We recommend that future research consider data that spans all five categories. We also find that competitor information is often limited to aggregate measures of advertising expenditure or product pricing. Including more granular competitor marketing data (e.g. daily Facebook likes) could provide more detailed insights into how a company's advertising is affecting other companies. It would be valuable for future research to test this proposition.

The data used in the literature can be characterised as either individual-level or aggregate. Individual-level data comprise data collected about specific individuals over time. Such data are usually collected at the marketing touchpoint level (i.e. every interaction with a media channel is recorded) or accumulated for each individual over a certain timeframe, often daily. In contrast, aggregate data are typically collected on a weekly basis, with some instances of monthly collection and only two instances of daily collection. Another noticeable difference is that aggregate data tend to have been collected over a few years, while the data from individual-level studies are usually collected only over a few months. This means that it is difficult to assess how reliable models built on path to purchase data are over time.

Individual-level data are often collected over a reduced timeframe due to data collection and preparation issues. First, tracking individuals over multiple media channels usually involves integrating data from multiple sources, each with their own format. For example, Li and Kannan (2014) combine datasets from disparate sources such as

DoubleClick, Omniture Site Catalyst, referral engines and email campaigns. Second, a large portion of the datasets used in prior studies use cookies to track consumers' online movements. While collecting cookie information over time allows marketers and researchers to see a consumer's path to purchase, there are limitations to cookie-tracking:

- **Time Limitations** – cookies have set expiry dates (e.g. 30 days), meaning that consumer interactions with media channels after the expiry date are viewed as being part of a new path to purchase.
- **Cookie-Blocking** – consumers can disable browser cookies, meaning that each time they interact with a media channel, it is viewed as their first interaction, rather than a touchpoint in a single path to purchase (Jayawardane et al. 2015).
- **Multiple Devices** – cookies are unique to each device, meaning that consumer interactions on a mobile phone and a laptop are not considered part of the same path to purchase (Jayawardane et al. 2015).
- **Multiple Users of the Same Browser** – cookies only identify unique browsers, meaning that if multiple individuals use one browser, they are treated as the same consumer.

A more reliable alternative to tracking consumers is requiring them to login using their login details for another platform (e.g. Facebook). The advantage of this approach is that users can be personally identified and tracked across multiple devices and over an indefinite timeframe. However, data regarding customers' touchpoints ultimately lies in the hands of the platform provider. The same applies to third-party cookie data. This means that even if accurate individual-level data can be collected, it can still be costly for businesses to purchase the data.

Despite this, the advantages of individual-level data can outweigh this cost. By recording each consumer's interaction with a media channel, individual-level data offer a more complete picture of the effects of advertising. This can enable marketing managers to answer more granular questions, such as which individual ads are most effective at driving sales and whether there are differences in consumer behaviour depending upon the device being used (Kannan et al. 2016). Collecting data at a higher granularity also means that models are estimated using larger datasets, which can help reduce the variability caused by a large number of parameters relative to observations. Another advantage provided by individual-level data is the potential for segmentation and targeting of consumers (Kannan et al. 2016). Understanding the effects of advertising on individual consumers may help companies make more efficient real-time bidding decisions, as well as personalize advertisements and promotions (Anderl et al. 2016). For example, Li and Kannan (2014) developed an attribution model that can provide insight into whether and when to make a marketing intervention, such as an email to individual consumers who visit a website. These insights can help marketers optimize their advertisements on an individual basis, which could lead to increased conversions and sales.

While the proliferation of online media channels has provided marketers with opportunities to acquire individual-level customer data, many companies still only have access to data aggregated at a higher level (Kireyev et al. 2016). Some of the challenges with individual-level data include its susceptibility to inaccuracies (e.g. cookie deletion, incorrect self-reporting) and its tendency to be difficult and costly to collect. This is particularly true for small and medium size enterprises (SMEs), which often don't have the resources to collect and exploit individual-level marketing data. Additionally, changes to social media platforms' algorithms and user data policies can change the granularity and type of data companies have access to.

Moreover, while data for online media channels are potentially available at an individual level, the data for offline media channels are often only available at a more aggregated level. This makes it difficult to quantify the effects of online and offline media channels because the data granularity varies across the media channels. Despite these challenges, there are some promising approaches for linking aggregate offline and more granular online datasets, such as the Bayesian data-fusion approached proposed by Feit et al. (2013). Working with both online and offline media channels is also often challenging due to data collection from disparate sources, as well as storage and organisation issues associated with a large amount of data (Chan and Perry 2017).

Finally, it can be easier to collect aggregated data when different parties have control over different media channels. For example, a company runs their own website while a marketing agency coordinates online display and paid search channels. The growing concerns and regulations on consumer privacy, such as the European Union's General Data Protection Regulation (GDPR) and its provision to impose substantial fines, may also limit the ability of firms to collect and retain quality individual-level data.

Furthermore, models based on aggregate data have been shown to provide useful insights into advertising media effectiveness, with budget allocation informed by these insights resulting in improved financial performance (de Haan et al. 2016; Wiesel et al. 2011). Therefore, research that uses aggregate data to develop advertising media effectiveness models is also important and relevant to industry. While individual-level data can provide more granular insights and enable personalized targeting, it is important to evaluate its cost and quality before using it to quantify advertising media effectiveness.

2.3.2 Model Development

The Model Development stage uses the collected data to estimate one or more advertising media effectiveness models. Figure 2-3 presents a topology showing the modelling approaches identified from this review. In light of the conceptual structure of the literature shown in Figure 2-1, we have categorised these approaches by whether individual-level data or aggregate data are used.

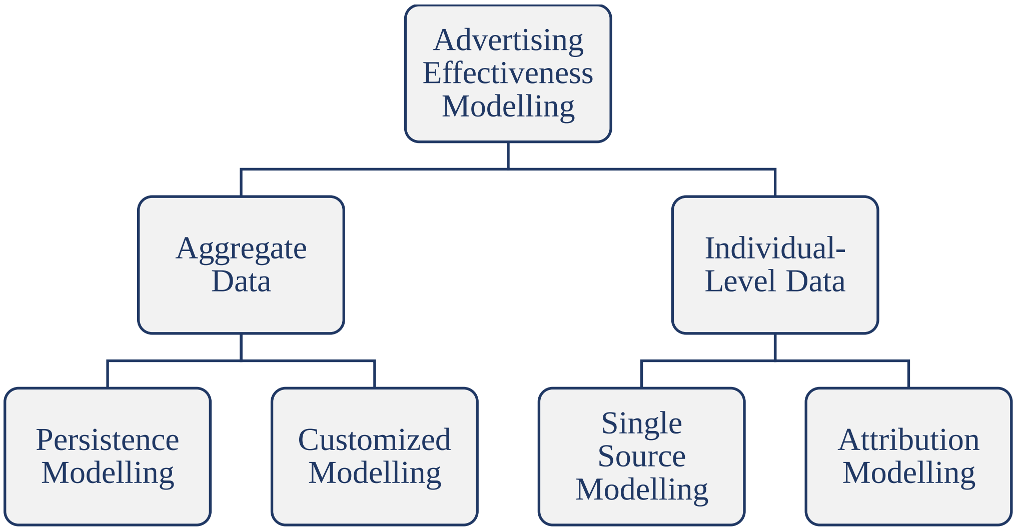


Figure 2-3 - Main Advertising Media Effectiveness Modelling Approaches

2.3.2.1 Aggregate Data Modelling Approaches

Within the aggregate data category, the two major modelling approaches are persistence modelling and customized modelling. Pioneered by Dekimpe and Hanssens (1995), persistence modelling uses multivariate time-series techniques to model the relationships between media channels and marketing performance metrics. Multivariate time-series models can simultaneously model the relationships between multiple variables, while handling endogeneity issues and incorporating exogenous variables, such as product pricing. Persistence modelling is typically used to address research questions requiring the disentangling of the short- and long-term effects of media channels on marketing performance metrics (Leeflang et al. 2009). This is usually done using impulse response functions, which can also provide insight into the direct and indirect effects of media channels over time. Persistence modelling allows the data to define the relationships between marketing variables, rather than pre-specifying relationships. This can make persistence models more flexible, generalisable and extensible than other models that require relationships to be pre-specified. However, this flexibility comes at the cost of the results often being harder to interpret.

We have classified papers in the literature reviewed that do not fall into the persistence modelling category, but still use aggregate data, as adopting a customized modelling approach. This is because while persistence modelling papers tend to focus on disentangling the short and long-term effects of marketing, customized modelling papers address a broader range of research questions that require the use of different modelling techniques. For example, Kumar et al. (2017) use a time-varying effects model to capture the time-varying synergistic effects of social media and traditional media on brand sales. In contrast with persistence modelling, researchers adopting the customized modelling approach tend to specify relationships between marketing communications activities and marketing performance metrics. For example, Dinner et al. (2014) assume that offline media channels influence online searches for a company, which, when combined with other online media channels, influence total sales. This enables researchers to construct models that are relatively easy to interpret and address specific research questions. However, this approach may reduce the generalisability of results, as the frameworks of pre-defined relationships may only be relevant to the datasets on which they are built. It can also be difficult to extend these frameworks as there is little guidance around how new marketing variables (e.g. a new media channel) should be incorporated. Additionally, omitted variable bias may be introduced if important relationships are not included in the modelling framework.

2.3.2.2 Individual-Level Data Modelling Approaches

The availability of individual customer data detailing their interactions with online media channels has heightened academic and industry interest in developing more granular models (Kannan et al. 2016). Such research can be broken into two streams:

single source modelling and attribution modelling. Single source modelling involves applying traditional marketing mix concepts to individual-level data, which are often accumulated at a daily level. All of the studies rely on data from a sample of customers, with three of the studies examining members of company loyalty programs (Danaher and Dagger 2013; Kumar et al. 2016; Saboo et al. 2016). Since single source studies typically use data from a sample of customers, it is important that adopters of this approach ensure that the sample is representative of all customers. Another challenge with single source modelling is the collection of individual-level media exposure data, with customer self-reporting potentially limiting the accuracy of research results (Danaher and Dagger 2013).

Unlike the other approaches, which usually involve some form of aggregation, attribution models tend to operate at a more granular level by considering individual marketing touchpoints. Since these data are usually not available for offline media channels, attribution modelling is typically limited to online media channels. Another difference between single source modelling and attribution modelling is that the single source approach will typically involve assessing the effectiveness of media channels using traditional approaches, such as elasticities. Attribution modelling, on the other hand, estimates media channel effectiveness by aggregating the attribution value of each channel for each conversion. Attribution models are generally based on either heuristics, which assign credit to media channels based on certain rules (e.g. last touch), or more data-driven modelling techniques (Jayawardane et al. 2015). These models typically use a measure of individual purchase outcome as the dependent variable and media exposure and control variables as the independent variables. Survival (Ji and Wang 2017; Hou et al. 2016) and Markov models (Anderl et al. 2016) have been shown to outperform common heuristic models, such as last touch, and more simplistic techniques, such as logistic regression, in predicting conversions. These models are closer representations of reality as they consider the sequencing and time effects of individual customer journeys. The main data-driven models used in the literature are summarized in Table 2-1, with the relatively limited use of neural networks representing an opportunity for further research.

Model Type	Description	Papers
Simple Probabilistic	Compiled using empirical probabilities rather than being estimated using statistical learning techniques	Shao and Li (2011), Dalessandro et al. (2012), Wooff and Anderson (2013), Yadagiri et al. (2015), Yin et al. (2016), Nisar and Yeung (2018), Yuvaraj et al. (2018)

Model Type	Description	Papers
Binomial	Binary response models, such as logistic regression, that are used to predict the probability of a binary event (e.g. conversion or not)	Shao and Li (2011), Sinha et al. (2014), Danaher and Van Heerde (2018), Yuvaraj et al. (2018), Li et al. (2019), Zhao et al. (2019)
Bayesian	Capture consumer heterogeneity by using Bayes theorem and prior distributions to take advantage of pre-existing information (e.g. managerial knowledge)	Li and Kannan (2014), Nottorf (2014), Xu et al. (2014), Brodersen et al. (2015)
Markov	Account for the sequence of a consumer's marketing touchpoints by describing their path to purchase as a series of transitions between states (media channels)	Abhishek et al. (2012), Anderl et al. (2016), Kaatz et al. (2019), Kakalejcik et al. (2019), Tinyakova et al. (2018)
Survival	Incorporate the time dimension by modelling the time until an event (e.g. conversion) occurs	Zhang et al. (2014), Hou et al. (2016), Ji et al. (2016), Ji and Wang (2017)
Neural Network	Flexible machine learning models designed to capture complex relationships between variables via interactions and non-linear functions	Ren et al. (2018)

Table 2-1 - Common Attribution Models Used in the Literature Reviewed

2.3.3 Model Assessment

Once an advertising media effectiveness model has been built, it is usually evaluated with goodness-of-fit tests and out-of-sample predictions. There is a wide range of statistical tests and model assessment methods used in the literature. The major modelling aspects assessed in the literature, along with relevant tests, are presented in Table 2-2. This table is an updated extension of Franses (2005), who discusses model assessment and diagnostics for marketing models. While caution should be taken when trying to compare literature across these modelling issues as they answer different research questions, the most commonly addressed modelling issues in the literature reviewed are in-sample fit and endogeneity.

Modelling Issue	Description	Example Diagnostics
In-Sample Fit	Examines how well the model fits the data used to estimate it	R^2, Adjusted R^2, AIC, BIC, log-likelihood, parameter significance
Out-Of-Sample Fit	Investigates how well the model fits data not used to estimate it, usually through running the model on a validation dataset or using cross-validation	Mean squared error (MSE), mean absolute error (MAE)
Autocorrelation	When errors are correlated with one another	Durbin-Watson test
Multicollinearity	When there is correlation between independent variables	Variance inflation factors
Heteroscedasticity	When the variance of errors is non-constant	Goldfeld-Quandt test
Endogeneity	When an independent variable is correlated with an error term	Hausman test
Error Distributional Assumptions	Checks that distributional assumptions made about the errors are valid	Jarque-Bera test (testing for normal distribution of errors)
Functional Form	Checks that the model has the correct functional form (e.g. does not have non-linearities that are unaccounted for)	Ramsey RESET test

Modelling Issue	Description	Example Diagnostics
Time-Series Properties	Checks for specific time-series issues like stationarity and cointegration	Unit root tests (e.g. ADF and KPSS tests), Johansen cointegration test
MCMC Conversion (Bayesian Statistics)	Checks that the MCMC iterations have converged to a steady state distribution	Gelman-Rubin test, trace plots

Table 2-2 - Modelling Issues & Diagnostics for Advertising Media Effectiveness Models

The Model Assessment stage also involves using the model's outputs to derive measures of advertising media effectiveness. Elasticities or impulse response functions are the dominant measures of advertising media effectiveness for marketing mix models, persistence models and single source models. Attribution models, on the other hand, assess media channel effectiveness by calculating attribution values. These are usually defined for each media channel as the incremental contribution to a performance metric (e.g. probability to purchase) from a customer's exposure to that media channel. If calculated at the individual exposure level, this value is usually aggregated over all conversions involving the media channel to provide the attribution value of the whole media channel.

Aside from testing for specific modelling issues, some researchers compare the results of their model with variations of it or models proposed by other researchers (Anderl et al. 2016; Ji and Wang 2017). Comparisons are extremely useful as they provide insight into the relative strengths and weaknesses of different models. This seems to occur more often in the attribution modelling literature, where researchers usually compare their model's performance against heuristic models, such as last touch, and Shao and Li's (2011) logistic regression model. Another assessment tool used by researchers is to adjust marketing budget allocations based on a model's results and evaluate how this influences one or more marketing performance metrics. While other tests check for common modelling issues, evaluating the implementation of the model's results on new, future data is the ultimate form of verification as it directly examines whether the model can actually improve marketing performance. This is an area that can be developed further, with relatively little research available on using attribution models to inform marketing resource allocations. Two exceptions to this are Geyik et al. (2014) and Danaher and Van Heerde (2018).

While model assessment is a vital part of developing an advertising media effectiveness model, it is not always given prominent attention in the literature. To facilitate the comparison and verification of advertising media effectiveness models, researchers should consider the modelling issues in Table 2-2 and clearly report the results from each test. Additionally, it would be valuable for researchers to practically evaluate how the implementation of their model's results improves marketing decision-making. Demonstrating how academic models can be applied in practice will encourage and facilitate the adoption of academic research by industry. This would, in turn, encourage industry support and provision of data, as well as promote faster development of academic research.

Finally, confidence intervals, which provide insight into the stability of parameters and predictions, are not commonly presented in the literature, with Dekimpe and Hanssens (1999), Li & Kannan (2014) and Kireyev et al. (2016) being notable exceptions. Hanssens and Pauwels (2016) report that risk is not comprehensively addressed in marketing academia or practice, with most studies that consider risk only performing best- and worst-case scenario testing using standard errors from parameter estimates. Hanssens and Dekimpe (2017) also identify that the investigation of the stability of marketing performance metrics has received less attention in the literature, while Albers (2012) explains that practitioners often prefer to see a range of solutions rather than specific numerical recommendations. Providing measures of risk and visual tools, such as dashboards, would allow marketers to see the trade-offs between marketing variables and may help to increase their trust and use of statistical models. This would also help researchers more thoroughly compare models and make appropriate recommendations.

2.3.4 Resource Allocation

After quantifying the effectiveness of a company's media channels, the next step is to leverage this information to improve financial performance. This can be done by making general recommendations, such as to increase spending on media channels with larger elasticities. Wiesel et al. (2011) take this approach, recommending that their partner firm decrease spending on flyers and increase spending on paid search. These changes resulted in a net profit increase that was 14 times larger than the status quo allocation (Wiesel et al. 2011). The other way to implement the findings from advertising media effectiveness models is to make specific media resource allocation recommendations. This involves deriving optimal marketing resource allocation formulas and substituting parameter estimates or elasticities from advertising media effectiveness models into them. An example of this is provided by de Haan et al. (2016), who show that the retailer providing data for their advertising media effectiveness model could increase revenues by 21% by switching from the current budget allocation to their recommendation. Overall, the literature is clear that applying the results of advertising media effectiveness models has the potential to increase a company's financial performance (de Haan et al. 2016; Kireyev et al. 2016; Danaher and Dagger 2013; Wiesel et al. 2011; Danaher and Van Heerde 2018; Geyik et al. 2014).

The two main approaches used to generate specific marketing resource allocation recommendations are dynamic and static optimization. This topic is the subject of a review by Mantrala and Kanuri (2016) and as a result, is not covered extensively in this chapter. However, since the outputs of advertising media effectiveness models often provide inputs for marketing resource optimization, a brief discussion is provided. Dynamic optimization typically involves defining an objective function that varies over time and then maximising (or minimising) it whilst adhering to any constraints. While dynamic resource optimization models are more complicated than static optimization rules, Kireyev et al. (2016) point out that dynamic allocation models often result in recommendations that are very similar to static ones. Additionally, static optimization models are by definition more stable as they do not change over time. They also tend to be more interpretable and easier to understand. Dorfman and Steiner (1954) produced one of the most influential works on static optimization, finding that the budget allocation of media channels should be based on the ratio of their elasticities. This principle is widely used for budget allocation within the literature reviewed (Pauwels et al. 2016b; Kireyev et al. 2016). Danaher and Dagger (2013) extend the principles of Dorfman and Steiner (1954) to derive a formula that factors in the different costs of each media channel:

$$Optimal\ Spend_k = \left(\frac{\eta_k}{\sum_{k=1}^{K} \eta_k}\right)\left(\frac{B}{C_k}\right), \quad (2\text{-}1)$$

where B is the total advertising budget, C_k and η_k are respectively the cost and advertising elasticity of media channel k, out of a total of K channels. de Haan et al. (2016) and Danaher and Van Heerde (2018) also adopt this approach and find improvements in marketing performance.

While most of the literature makes specific recommendations when performing budget optimization, industry does not often implement these types of optimization solutions (Albers 2012). Heuristic methods are widely used in practice to determine marketing resource allocation. Examples of these methods include setting the marketing budget based on a fixed percentage of historical sales or estimates of the advertising expenditure of competitor firms (West et al. 2014). In a survey completed by 125 US marketers, West et al. (2014) find that heuristics are popular as they are easy to use and provide general rules that can be adapted depending on the situation. Another key reason for the use of heuristics in marketing resource allocation is that marketers tend to be distrustful of solutions that give specific outputs, instead preferring to use tools that allow them to see a range of options, which they can use to help them make a final decision (Albers 2012). Examples of such tools include Microsoft Excel and dashboards, which allow managers to interact with the data-driven models and formulas behind advertising media effectiveness estimates and budget recommendations. This helps managers assess trade-offs and understand recommendations, which encourages their implementation (Albers 2012). This practice also reinforces our earlier recommendations to consider measures of risk and visual tools when assessing and comparing models.

The survey completed by West et al. (2014) reveals that the marketing budget setting process often involves both heuristics and data-driven models, with heuristics helping to provide insight into the reasonableness of the results of data-driven models (West et al. 2014). Thus, marketers often use multiple methods to inform marketing resource allocation decisions. It would be beneficial for researchers to keep this in mind when illustrating how management can use their models for decision-making.

2.4 Key Findings and Managerial Implications

The key managerial themes and insights that can be observed throughout the literature are summarized in Table 2-3.

Finding	**Example Papers**
There are synergies between media channels. Companies benefit from coordinating their online and offline media channels, as well as their paid, owned and earned media channels	Naik and Raman (2003), Naik and Peters (2009), Wiesel et al. (2011), Dinner et al. (2014), Kumar et al. (2016), Pauwels et al. (2016b), Frison et al. (2014), Srinivasan et al. (2016), Pauwels et al. (2016a), Kireyev et al. (2016)
Advertising media effectiveness and synergistic effects of media channels on marketing performance metrics can vary over time	Osinga et al. (2010), Kumar et al. (2017)

Finding	Example Papers
Media channels are often grouped into one of a few prominent categorisations for analysis: Paid, Owned and Earned; Firm-Initiated Channels (FICs) and Customer-Initiated Channels (CICs); Online and Offline	Paid, Owned and Earned Media Channels: Pauwels et al. (2016b), Srinivasan et al. (2016); FICs and CICs: Wiesel et al. (2011), de Haan et al. (2016) Online and Offline Media Channels: Naik and Peters (2009), Dinner et al. (2014)
Customer-initiated media channels tend to have a greater influence on marketing performance metrics than firm-initiated media channels	Wiesel et al. (2011), Dinner et al. (2014), de Haan et al. (2016), Pauwels et al. (2016b)
Traditional media channels, such as TV, can still be effective. They tend to influence consumers earlier on their path to purchase and drive them to online media channels, such as paid search	Bollinger et al. (2013), Danaher and Dagger (2013), Liaukonyte et al. (2015), Joo et al. (2016), Srinivasan et al. (2016), Pauwels et al. (2016a)
Last touch attribution models tend to underestimate the effect of firm-initiated media channels, such as display advertising, and overestimate the effect of customer-initiated media channels, such as paid search advertising	Abhishek et al. (2012), Li and Kannan (2014), Nottorf (2014), Xu et al. (2014), Anderl et al. (2016)
Including consumer attitudinal metrics, as well as measures from owned and earned media channels, can help forecast sales	Srinivasan et al. (2010), Hanssens et al. (2014), Liaukonyte et al. (2015), Kumar et al. (2016), Pauwels et al. (2016b), Srinivasan et al. (2016)

Table 2-3 – Key Observations from the Literature Reviewed

A consistent finding across the literature is that there are synergies between media channels and that companies benefit from coordinating their online and offline media campaigns. This is often referred to as omnichannel management (Verhoef et al. 2015). However, Kolsarici and Vakratsas (2018) find that these synergistic effects are not always positive, sometimes having an antagonistic (negative) effect on marketing performance metrics. Additionally, Kumar et al. (2017) show that the effectiveness of social media and TV, along with their synergistic effects, can vary over time.

Therefore, in order to create accurate advertising media effectiveness models, it is important to incorporate both the direct and indirect effects of online and offline media channels on company sales (Dinner et al. 2014). Direct effects refer to the influence of a media channel on sales, while indirect effects refer to the influence of a media channel on another media channel, which in turn drives sales. While media channels with large direct effects (e.g. paid search) are easier to observe, media channels with large indirect effects (e.g. display) are often still important, albeit harder to track. Li and Kannan (2014) demonstrate that firm-initiated media channels, such as display, tend to have larger indirect effects. This is because they typically influence the early stages of the consumer path to purchase, making consumers aware of a product, before customer-initiated media channels help them make a final decision. Thus, firm-initiated media channels remain valuable and should not be dismissed by marketers. The ability to capture this network of relationships is an important advantage that more data-driven advertising media effectiveness models have over heuristic models, such as last touch, which tend to overestimate the effect of customer-initiated media channels. While heuristic models are widely used in industry due to their simplicity and ease of use, more data-driven advertising media effectiveness models can provide marketers with a more accurate understanding of the effectiveness of their media channels.

Future research should build on these findings by further investigating the direct and indirect effects of media channels on marketing performance metrics. Given the ongoing shift of media budgets from offline to online media channels, an area of interest is the time-varying synergistic effects of online and offline media channels (Kumar and Gupta 2016). As noted by Kolsarici and Vakratsas (2018), there has been less research into the antagonistic effects of media channels relative to their synergistic effects, particularly its underlying causes.

It is also important to consider the limitations of advertising media effectiveness models so that their results are not misinterpreted. If a media channel appears ineffective according to the results of an advertising media effectiveness model, it does not necessarily imply that the media channel itself is ineffective. It could also mean that the advertising creative is performing poorly, rather than the media channel itself. Another limitation of advertising media effectiveness models is their inability to identify emerging media channels. For example, advertising using Facebook when it first started would likely have seemed ineffective when considered in a model. However, persisting with Facebook would likely have generated much better results over time than reducing advertising on the platform. Thus, there is clearly still a crucial role to play for marketers, whose experience and expertise can help identify emerging media channels and situations where a model's results may not reflect the underlying reality. We believe that the combination of advertising media effectiveness models and marketers' judgements, such as the identification of future trends, will be crucial in helping companies optimize their marketing performance moving forward.

Finally, advertising media effectiveness models are built using datasets for specific companies and industries. The conclusions that some models draw about media channels may not be generalisable to some other companies because of company and industry differences, such as the media preferences of their customers. Other contextual differences that may hinder the generalisability of results include different advertisement (e.g. video quality) and product (e.g. price) characteristics. More studies that consider multiple brands and companies from multiple industries, such as Pauwels et al. (2016b), are important in advancing understanding in this field.

2.5 Recommendations for Future Research

Throughout this chapter, we have made several best-practice recommendations for future research that are summarized in Table 2-4.

Stage	**Recommendations for Future Research**
Data Collection	Include multiple social media channels; Include all categories (e.g. paid, owned and earned) of media channels; Investigate the optimal metrics to use for owned and earned media channels; Incorporate advertising metadata (e.g. design information); Investigate using data from the five categories identified in section 2.3.1, including more granular competitor data
Model Development	Connect individual-level modelling approaches with more aggregate-level modelling approaches; Investigate the use of Bayesian methods to incorporate prior customer and advertising information to make models more accurate; Consider the time-varying synergistic and antagonistic effects of online and offline media channels
Model Assessment	Provide measures of risk (e.g. confidence intervals); Consider modelling issues (e.g. endogeneity) and report how they are handled; Use model comparisons to show how an advertising media effectiveness model improves on the extant literature; Practically apply advertising media effectiveness models across multiple companies and industries to verify results and produce empirical generalizations
Resource Allocation	Investigate resource allocation based on attribution modelling; Report a range of results, rather than just specific numbers (e.g. consider different scenarios)

Table 2-4 – Recommendations for Future Research in Quantifying Advertising Media Effectiveness

We subsequently discuss key future research directions for quantifying advertising media effectiveness. These have been derived from future research areas suggested by the literature reviewed and intuitions about the process of quantifying advertising media effectiveness. We group these future research directions by the two critical components of measuring advertising media effectiveness: data and modelling.

2.5.1 Data

A key area of future research is taking advantage of the amount of data available about marketing communications activities. For example, Kolsarici and Vakratsas (2018) and Kumar et al. (2017) suggest including competitor information when measuring synergistic effects as a direction for future research. The growth of online media channels has led to the development of many additional measures of consumer behaviour, such as website clicks and video views. Wedel and Kannan (2016) explain the need for marketing models to take advantage of this information, stating that current research has only scratched the surface of information available about consumer media channel interactions. Given the amount of information available, research is needed to investigate how the choice of marketing metrics (e.g. clicks, impressions, expenditure, ratings, etc) in advertising media effectiveness models influences their output (Liu-Thompkins 2019).

In addition to behavioural data, media channels themselves contain a wealth of metadata (e.g. design, content information) that is often only represented as a single data point (e.g. expenditure) (Wedel and Kannan 2016). Investigating how the characteristics of advertisements, such as their content and design features (e.g. color), can moderate their effectiveness, is a potential area for future research. Unstructured data, such as social media posts, could also provide valuable information for advertising media effectiveness models. Furthermore, Hanssens and Pauwels (2016) highlight the need for better integration of soft, attitudinal metrics (e.g. survey based) with hard, behavioural metrics (e.g. click-through rate) in marketing models. The literature suggests that incorporating these additional metrics into advertising media effectiveness models can add valuable information to them. For example, in an investigation covering $3.4 billion in spending by 20 brands, Liaukonyte et al. (2015) found that television advertising influences online shopping, with advertising content playing a key role. Srinivasan et al. (2016) also note that consumer activity data can serve as early indicators of advertising media effectiveness, with their results suggesting that if a TV campaign failed to generate increased paid search clicks, sales volumes were only likely to marginally increase.

Finally, analysing media channels in categories allows marketers to assess advertising media effectiveness at a more aggregated level, such as how effective owned media channels are compared to paid media channels. This would be useful for strategic decision-making. The categorisation of the media channels identified in the literature is shown in Table 2-5.

Media Channel	Paid, Owned or Earned	FIC or CIC	Online or Offline
Offline Media (i.e. television, radio, print)	Paid	FIC	Offline
Salesforce	Paid	FIC, CIC	Offline
Paid Search	Paid	CIC	Online
Display	Paid	FIC	Online
Affiliates	Paid	FIC	Online
Retargeting	Paid	FIC	Online
Email	Owned	FIC	Online
Organic Search	Owned	CIC	Online
Website (Direct)	Owned	CIC	Online
Referrals	Earned	CIC	Online
Reviews	Earned	CIC	Online
Mobile	Paid, Owned	FIC, CIC	Online
Social Media	Paid, Owned, Earned	FIC, CIC	Online
Audio	Paid, Owned, Earned	FIC, CIC	Online
Video	Paid, Owned, Earned	FIC, CIC	Online

Table 2-5 - Media Channel Classification in the Literature Reviewed

Interestingly, some media channels can belong to multiple categories (e.g. Facebook can be a paid media channel through advertising and an owned media channel through a company page). However, advertising media effectiveness models in the literature tend to categorize each media channel into only one category. To help more accurately capture the direct and indirect effects that occur, future research should include all of the categories of a media channel. Additionally, when developing models, researchers would benefit from considering all the media channels a company uses, not just the ones that fall under the paid media category. Earned and owned media channels may exert their own influence on sales or have synergistic relationships with paid media channels. Thus, considering paid, earned and owned media channels are critical for gaining a clear perspective of the effectiveness of media channels and necessary to provide accurate recommendations. Another notable observation is the absence of multiple social media channels, such as Facebook and Instagram, in a single study. Given the widespread usage and advertising across social media, it would be valuable for future research to consider all of the individual social media channels that a company uses.

2.5.2 Modelling

To provide an overview of the strengths and weaknesses of the various modelling approaches, we adopt a similar method to Leeflang et al. (2009) by rating each approach in Figure 2-3 according to the four criteria in Table 2-6. We use a “+” to indicate that the approach tends to cope well, a “-” to indicate that the approach tends to cope poorly and a “+/-” to mean that it tends to cope neither well nor poorly.

Modelling Approach	**Data Accessibility**	**Generalisability/ Extensibility**	**Interpretability**	**Flexibility**
Persistence	+: aggregate data generally easier to collect	+: new media channels can usually be added in as an additional parameter to the model	+/-: lack of defined relationships can make results harder to interpret	+: data-driven as it does not impose defined relationships between media channels
Customized		+/-: often impose pre-defined relationships between media channels, which can make it harder to add new media channels to the framework	+: defined relationships typically make results easier to interpret	+/-: models with pre-defined relationships are less likely to highlight other relationships in the data
Single Source	-: individual-level data generally harder to collect	+: can usually easily add new media channels	+/-: depends on the model used, can range from easy (e.g. logistic regression) to difficult (e.g. Bayesian)	+: individual-level data can capture more granular insights
Attribution		+/-: new online media channels are generally easy to add, but offline media exposures are difficult to capture		

Table 2-6 - Ratings of Modelling Approaches in the Literature Reviewed

As Table 2-6 and the preceding discussion suggests, the modelling approach suitable for a research study depends upon the purpose of the analysis and the data available. Persistence modelling is often used in studies that involve aggregated and endogenous variables. However, when allowing for parameter variation over time, a customized modelling approach, such as state-space modelling incorporating time-varying effects, is typically more appropriate (Leeflang et al. 2009). More broadly, Kumar and Gupta (2016) note the increasing demand for dynamic models over traditional static modelling, with Osinga et al. (2010) and Kumar et al. (2017) showing that the effects of marketing can vary over time. The customized modelling approach is also often used when modelling requires the incorporation of aggregate, cross-sectional units, such as different markets (Dinner et al. 2014) and product categories (Wang et al. 2017).

A key trend over the last decade has been the increase in the number of studies examining the effectiveness of media channels on an individual-level via attribution modelling. This is shown in Figure 2-4 and Figure 2-5, which break down the empirical studies in the literature reviewed by modelling approach, publishing year and model estimation method.

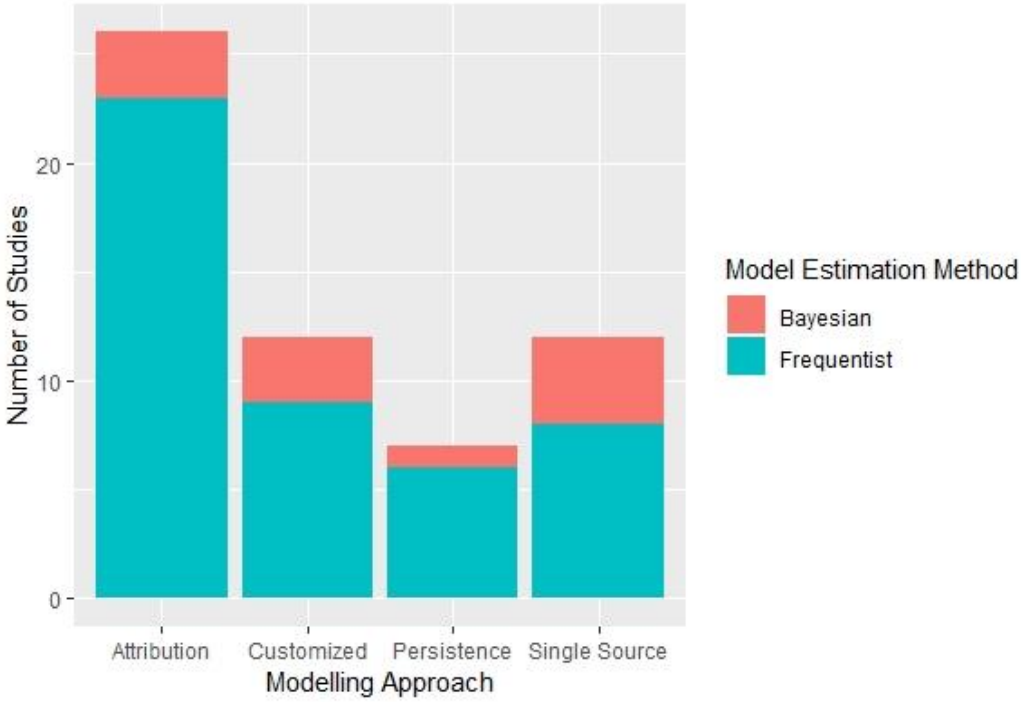


Figure 2-4 - Literature Reviewed by Modelling Approach & Model Estimation Method

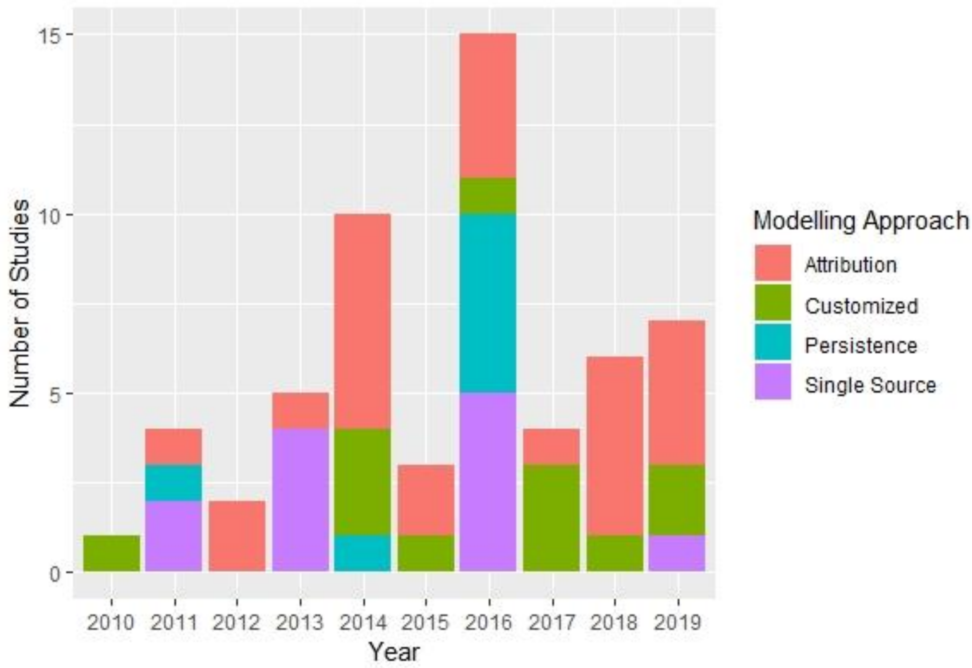


Figure 2-5 - Literature Reviewed by Publishing Year and Modelling Approach

The growth in attribution modelling studies has been driven by the increased availability of individual-level data and with it, the prospect of quantifying advertising media effectiveness on an individual level. An important question is how attribution modelling relates to conventional advertising media effectiveness concepts, such as advertising elasticities. Danaher and Van Heerde (2018) are one of the first to address this issue, deriving a formula that links attribution to advertising elasticities. The authors show that the attribution of a media channel is proportional to the elasticity of the response variable to the media channel multiplied by the number of exposures to that channel. This suggests that while elasticities capture the potential influence of a media channel regardless of exposure to it, attribution measures capture its realized influence given the number of exposures. Thus, budget allocation using attribution measures can result in a bias towards media channels that consumers are highly exposed to, rather than ones that are truly effective. For three brands of a North American speciality retailer, Danaher and Van Heerde (2018) find that optimal budget allocation based on elasticities outperforms those based on attribution measures. The connection of elasticities and attribution measures provides many opportunities for future research, such as investigating how advertising budgets can be optimized at an individual level.

A related area for future research is reconciling newer, individual-level modelling approaches, such as attribution modelling, with more traditional aggregate modelling approaches, such as persistence modelling. While one argument is to solely focus on the individual-level modelling approaches as they can provide more granular insights, data collection and organisation issues can sometimes hinder the implementation of these approaches in practice. Existing research has demonstrated that aggregate modelling approaches can provide insights into advertising media effectiveness that improve company financial performance, suggesting that there is still a place for these approaches. The big-picture perspective provided by aggregate modelling approaches may make them more suitable for strategic decision-making, while the granular insights provided by the individual-level modelling approaches may make them more useful for tactical decision-making and targeting. We call for future research to investigate how marketers can take advantage of both modelling approaches, or how to aggregate the results from granular attribution models to answer higher-level strategic questions.

Figure 2-4 also reveals that only 11 studies involve models estimated using Bayesian techniques, whereas 43 studies are based on models estimated using frequentist methods. Bayesian modelling can help deal with datasets of different granularities, incorporate additional known information into models and reduce the uncertainty of parameter estimates. For example, Gallego et al. (2019) produce more accurate sales forecasts by using a Bayesian structural time-series (BSTS) model that can incorporate a priori managerial information. Additionally, Wang et al. (2017) use hierarchical Bayesian models to pool data from multiple shampoo and soda brands to create more informative priors. They find that the hierarchical Bayesian models reduce the uncertainty of the estimates of media channel effectiveness and sales. However, selecting a prior for Bayesian models can be challenging and can heavily influence the model's accuracy. Models also usually need to be estimated using relatively complex techniques, such as MCMC methods. Nonetheless, with increasing computing power, there are further opportunities to explore the use of Bayesian estimation methods in quantifying advertising media effectiveness. These methods can be used to incorporate additional information into advertising media effectiveness models, such as customer demographic data or advertising metadata. Dynamic models that use Bayesian techniques may also be a useful starting point for reconciling individual-level modelling approaches with aggregate ones. For example, Bayesian structural time-series models (BSTS) have now been used for both aggregate data (Gallego et al. 2019) and individual-level data (Brodersen et al. 2015), while Feit et al. (2013) develop a Bayesian data-fusion approach to combine individual-level usage data from digital media channels with aggregated data from an offline media channel.

2.6 Conclusion

Quantifying the effectiveness of media channels, along with incorporating big-data techniques to handle data integration from multiple, disparate sources, has been identified as an important area for future research in marketing analytics (Kannan et al. 2016; Kannan and Li 2017; Wedel and Kannan 2016). This chapter comprehensively reviews research into quantifying advertising media effectiveness over the last decade. We synthesise the main data sources, methodological approaches and model assessment techniques of past research in the field, as well as discuss key managerial insights contained within it. We also identify some of the limitations of existing advertising media effectiveness models and provide best-practice recommendations for future research, which are summarized in Table 2-4.

This chapter finds that companies require an omnichannel management approach to marketing because it best reflects the complex and highly integrated nature of the consumer purchasing environment. Given their role in measuring the effectiveness of marketing efforts, it is important for advertising media effectiveness models to capture the complexities of the consumer path to purchase, as well as the direct and indirect effects of media channels upon marketing performance metrics. This can be achieved through aggregate data modelling approaches, such as VAR models, or individual-level modelling approaches, such as survival models. Additionally, there is a need for researchers in this field to take advantage of and incorporate advertising metadata (e.g. an advertisement's position on a website) and consumer activity data (e.g. Facebook likes). This data can help models create a more accurate picture of the consumer path to purchase and the relationships between media channels. Increasing collaboration with industry by practically implementing advertising media effectiveness models can also provide more insight into how these models can improve the decision-making of marketing managers. Finally, we call for more integration between aggregate data and individual-level data modelling approaches, building on the groundwork laid by this review.

Chapter 3 Measuring Advertising Media Effectiveness

This chapter investigates how consumer activity metrics influence practical marketing calculations, such as the ROI or CPA of a media channel. We also examine how the aggregation level of a marketing performance metric affects these measures of media channel effectiveness. We find that disaggregating marketing performance metrics, as well as including consumer activity metrics and indirect effects, reveals a more complete picture of the marketing landscape, particularly for media channels encountered in the early stages of the marketing funnel. Excluding these relationships can reduce the accuracy of ROI, CPA and budget allocation calculations. The results also show that owned media channels can influence other owned media channels, as well as marketing performance metrics. Additionally, the effects of paid and owned media channels on marketing performance metrics differ, with the immediate effects of owned media channels tending to be smaller, but more long-lasting, than those of paid media channels. Finally, we find that representing customer-initiated paid media channels, such as paid search, with both a consumer activity metric and firm activity metric in persistence models can help separate the effects of increases in advertising expenditure and consumer engagement on marketing performance.

3.1 Introduction

In an increasingly competitive global economy, measuring the effectiveness of media channels and optimizing the resources invested in them is critical to a firm obtaining a competitive advantage. However, determining the effectiveness of marketing investments is challenging due to the availability of numerous online and offline media channels, which results in a more complicated and non-linear consumer path to purchase (Batra and Keller 2016). Over the last decade, the allocation of advertising expenditure has shifted towards online advertising, with an estimated 23.1% of expenditure going to Alphabet (Google) in 2019 compared to 6.2% in 2010 (Handley 2019). Nevertheless, offline media channels still play a role in many companies' marketing strategies, with 29.9% of advertising expenditure spent on TV in 2019 compared to 34.1% in 2010 (Handley 2019).

Along with paid media channels, companies are exposed to consumers through owned and earned media channels. Owned media channels are controlled by firms, such as company Facebook pages, while earned media channels are driven by consumers, such as product reviews (Pauwels et al. 2016b). Additionally, there is an increasing variety of data businesses can collect about consumers' interactions with their media channels, such as website page views, Facebook reactions and paid search clicks. These consumer activity metrics are particularly important for representing owned and earned media channels as marketing expenditure is usually not associated with these channels. Marketers also often use them to evaluate and design marketing strategies. Questions arise as to which consumer activity metrics companies should be tracking, along with how this data can be used to help marketers identify and quantify the relationships between a company's media channels and its marketing performance metrics, such as sales and enquiries. The extant marketing literature has made some progress in addressing these issues. Existing research streams investigate how to measure media effectiveness and synergy across online and offline media channels, as well as how consumer attitude and behavioural data can be used in this process. Appendix 2 provides a summary table showing how this research compares to relevant marketing literature.

Danaher and Dagger (2013) examine the effectiveness of ten media channels with either sales or profit as the outcome measure. Their results suggest that traditional offline media remain effective, with television and direct mail having the strongest influence on sales and profit, while email and paid search are the best-performing online media channels. Frison et al. (2014) conduct a deeper analysis of the offline media category by examining the effectiveness of six offline media channels across 261 brands. They find that billboard and cinema advertising have no statistically significant sales elasticity, while television and magazine advertising do. Using only online media channels, Kireyev et al. (2016) develop a VECM to predict bank account applications based on paid search and display media channels. They find that display has a delayed positive effect on paid search, leading to more applications via this media channel. Dinner et al. (2014) investigate the influence of online advertising on online and offline sales, finding that positive cross-media advertising effects exist and are almost as strong as own-media effects. Interestingly, they find that online advertising is more effective than traditional advertising in terms of overall sales, primarily due to the strength of its cross effects on the offline channel (Dinner et al. 2014).

Wiesel et al. (2011) investigate the role of online and offline media channels at different stages of the customer purchase decision. Their results show that marketing communications activities have online-offline synergies and can directly influence both early and late stages in the consumer purchase decision. They also report that online, customer-initiated media channels are more effective than offline, firm-initiated ones. This finding is confirmed by de Haan et al. (2016), who quantify advertising media effectiveness over a wide range of media channels, reporting that customer-initiated media channels have higher elasticities than firm-initiated media channels. They also break customer-initiated media channels into content-integrated and content-separated categories. Content-integrated media channels are part of a website's content (e.g. price-comparison website), while content-separated media channels are not. de Haan et al. (2016) find that content-integrated media channels are more effective than content-separated media channels.

Naik and Peters (2009) develop a model that separates synergy into within-media synergy (i.e. within offline media channels) and cross-media synergy (i.e. between online and offline media channels). Overall, Naik and Peters (2009) find evidence of the presence of cross-media synergy, where offline media channels interact with online media channels to drive online visits to a website, as well as within media synergies. Pauwels et al. (2016b) investigate how media channel synergy differs across familiar and unfamiliar brands. They find that while online media channels are more effective than offline media channels, within-online synergy is greater than online-offline synergy for familiar brands, but not for unfamiliar brands. Thus, offline media channels can still be useful for unfamiliar brands as they build awareness about their products and services.

Srinivasan et al. (2010) analyse the path to purchase by examining the value of including consumer mindset metrics in a sales response model. They find that consumer mindset metrics of awareness, consideration, and liking help explain sales variance and act as early warning signals for sales performance. Additionally, Hanssens et al. (2014) investigate the effect of adding consumer attitude metrics, such as the percentage of survey respondents aware of the brand, to marketing mix models, finding that their inclusion can improve sales forecasts and budget allocation. Pauwels et al. (2016a) find that over a third of the offline store traffic for an apparel retailer comes indirectly through electronic word-of-mouth (eWOM) and organic search. Thus, marketing managers should track eWOM content metrics to fully capture the effects of paid media channels on marketing performance metrics. Finally, Srinivasan et al. (2016) investigate the effects of traditional marketing mix variables and online consumer activity on fast moving consumer goods (FMCG) brand sales. They find that TV advertising has a direct effect on sales, as well as an indirect effect via consumer activity metrics, such as paid search clicks. This means that tracking consumer activity metrics can act as early performance indicators of advertising campaigns.

These studies present strong evidence of within-media and cross-media synergies, thereby highlighting the value of using both online and offline media channels. The extant marketing literature also demonstrates that consumer mindset and activity metrics can improve sales response models and provide a more complete picture of the consumer path to purchase. However, there is a lack of research showing how including consumer activity metrics when modelling the effects of paid media channels affects return on investment (ROI), cost per acquisition (CPA) and budget allocation calculations. These metrics are commonly used in practice by marketing managers for reporting and decision-making purposes, making them important to consider in a research setting. There is also an absence of research showing the implications of using disaggregated marketing performance metrics when calculating these metrics. For example, Dinner et al. (2014) consider online and offline performance measures, while Srinivasan et al. (2016) consider aggregated ones. It is not clear whether one level of aggregation is preferable to the other. While it is often easier to collect aggregated data, disaggregated data can contain additional information that is lost by aggregation (Hyndman and Athanasopoulos 2018). This chapter addresses these issues by answering the following research questions, which are derived from RQ2:

1. What effects do consumer activity metrics have on ROI, CPA and budget allocation calculations?
2. What are the implications of using disaggregated marketing performance metrics when performing elasticity, ROI, CPA and budget allocation calculations?
3. To what extent do the findings of Srinivasan et al. (2016) generalise to different product, data and expanded media channel contexts?

This chapter makes three main contributions to the extant marketing literature. First, we complement the work of Srinivasan et al. (2016) by including an expanded number of online and offline media channels, collecting higher frequency data and using enquiries instead of sales as the marketing performance metric. These differences help us address the call of Srinivasan et al. (2016) to investigate the extent to which their findings generalize. Our research is also focused around a high involvement product category (luxury travel), whereas Srinivasan et al. (2016) focused on FMCG, a low involvement product category. While the effects of consumer activity metrics in high involvement product categories have been previously examined (Wiesel et al. 2011), the effects of traditional media are usually not accounted for (Srinivasan et al. 2016).

Our second contribution is to investigate the effects of including paid Facebook ads and owned Facebook page posts in the same model. When owned and earned media channels are considered, they are often included at the expense of paid media channels on the same platform (e.g. including a company's Facebook page, but not their Facebook advertising). To understand the full picture of a company's marketing performance, models need to consider all their paid, owned and earned media channels. We also examine the effects of including multiple metrics for paid search, namely advertising expenditure and engagement rate. Including multiple metrics for customer-initiated paid media channels is needed to separate the effects of the firm's advertising expenditure and the consumer's response (e.g. engagement rate), which may be driven by other factors, such as other media channels or advertisement quality. This is less important for firm-initiated media channels (e.g. TV) as consumers do not choose when they are exposed to them. Since consumers can seek out customer-initiated paid media channels themselves, the effects of that channel are unlikely to be fully captured by a metric representing the firm's investment in the channel. Thus, customer-initiated paid media channels should be represented with both a firm activity metric (e.g. expenditure) and a consumer activity metric (e.g. engagement rate).

Third, we make an empirical contribution to the literature by investigating how including consumer activity metrics, disaggregating marketing performance metrics and incorporating indirect effects influences elasticity, ROI, CPA and budget allocation calculations. Furthermore, we present a generalisable methodology for calculating the direct and indirect effects of each media channel and show how to present this information in visualisations that are meaningful to marketing managers.

Overall, we build on research examining the effects of incorporating consumer activity metrics into media channel effectiveness calculations (Srinivasan et al. 2016). We extend this research by investigating how these metrics influence ROI, CPA and budget allocation calculations, which are of high importance to marketing managers. We also generalise previous research by applying our methodology in a different product, data and media channel context.

3.2 Conceptual Framework

We adopt a conceptual framework based on the findings of Srinivasan et al. (2016), who show that advertising variables can influence marketing performance metrics directly, while also having an indirect effect via consumer activity metrics. Figure 3-1 outlines the relationships between the key components of the framework, namely advertising spend variables, consumer activity metrics and marketing performance metrics. The underlying idea is that consumers generally move through awareness (cognition) to liking (affection) to purchase (conation) stages on the path to purchase (Srinivasan et al. 2016). Individual consumers can progress through these stages at different speeds and exit at any stage. From a firm's perspective, advertising can stimulate an immediate desired response from a consumer, such as a sale. Alternatively, advertising can stimulate a consumer's progression to a different stage of the path to purchase, such as a searching for the firm online. This is captured by consumer activity metrics, which are consumer-initiated interactions with media channels, such as browsing a company website or Facebook page, that represent intermediate steps in the path to purchase. Consumer activity metrics may in turn have their own effects on response variables. For example, consumers may see an advertisement on TV, search for the product in Google, read about the product on the firm's website and then make a purchase or enquiry. Essentially, advertising can stimulate an immediate response (direct effect) or a consumer activity (e.g. Google search, Facebook search), which can then flow on to a response (indirect effect). This research seeks to empirically identify and quantify these direct and indirect paths, as well as investigate their effects on commonly used marketing calculations, such as ROI, CPA and budget allocation.

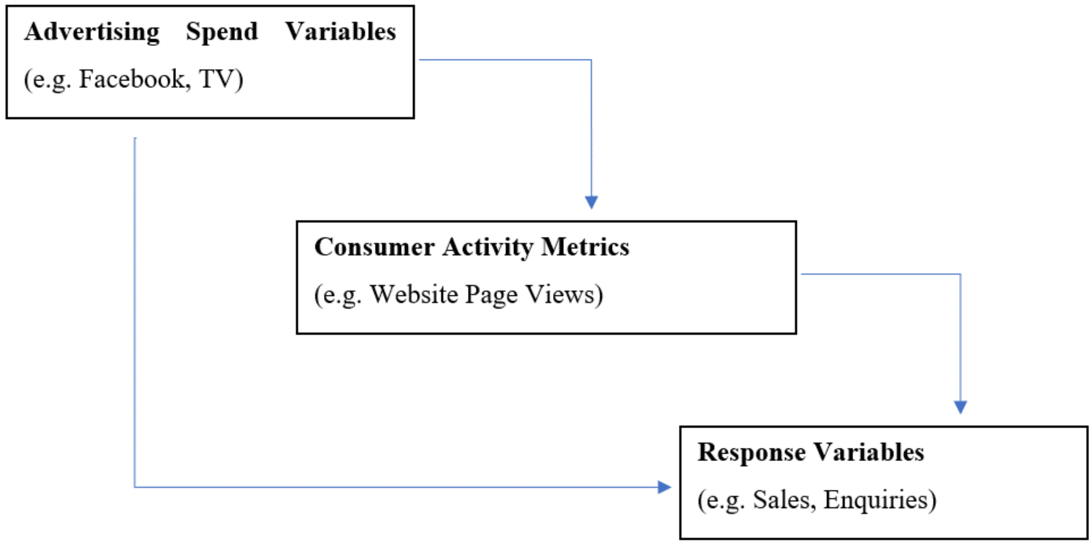


Figure 3-1 - Conceptual Framework of Customer Path to Purchase

3.3 Data

The data for this research were provided by an Australian media agency, which develops and coordinates advertising across paid media channels. The client company is a luxury travel brand that books holidays for consumers. The marketing performance metrics used by the travel brand are online and offline enquiries. Marketing is conducted across online and offline paid media channels that are operated by the media agency, as well as owned media channels that are operated by the travel brand. Data were collected daily between 01 April 2018 and 29 April 2019 (394 observations). Data were collected at an aggregate level as individual-level data was not available to the media agency. Using aggregate data helps ensure that the methodology is generalisable to all the media agency's clients, not just those with the resources to collect individual-level data. Aggregated data are also less problematic to collect from a legal perspective, with increased regulatory and public focus on privacy making companies more resistant to sharing individual customer data.

Before descriptive statistics were calculated, zero values in several variables were imputed due to data entry errors using the median of the non-zero values. The median was used so that the imputed values would not be influenced by outliers. The imputation process resulted in changes to four values for paid search spend and paid search impressions (used to calculate paid search engagement rate), three values for website page views and eight values for online enquiries. Table 3-1 describes the variables used in the analysis and provides descriptive statistics for them, along with the results of an Augmented Dickey-Fuller (ADF) test for stationarity.

Variable	Description	Mean	Median	SD	Skew	ADF Test Stat
Offline Enquiries	Number of enquires made via phone calls to the travel brand's call centre	153.67	136.00	105.42	0.24	-3.144 *
Online Enquiries	Number of enquiries made by submitting a form via the travel brand's website	58.62	51.00	33.02	1.88	-4.7174 *
Display Spend	Expenditure (AUD)	330.03	289.88	253.33	2.82	-4.5251 *
Facebook Ads Spend	Expenditure (AUD)	3863.16	3624.31	2309.13	1.06	-5.2315 *
Magazine Spend	Expenditure (AUD)	1548.24	0.00	4844.59	4.50	-7.4335 *
Outdoor Spend	Expenditure (AUD)	490.90	0.00	1120.03	2.28	3.4229 *
Paid Search Spend	Expenditure (AUD)	1423.42	1267.23	844.93	1.34	-5.1311 *
TV Spend	Expenditure (AUD) on traditional over the air TV	2243.86	843	3112.78	1.93	-2.6055
Video Spend	Expenditure (AUD) on online streaming content	615.22	479.86	569.81	1.47	-3.6388 *
Website Page Views	Total page views on travel brand's website	8211.85	7611.50	3026.50	0.97	-4.7488 *
Facebook Page Posts Engagement	Number of post interactions (likes, comments and shares)	563.19	461.00	789.21	11.84	-5.8967 *
Paid Search Engagement Rate	Number of clicks / number of impressions	0.16	0.16	0.06	-0.66	-3.571 *

Table 3-1 - Descriptive Statistics of Daily Enquiry and Media Channel Variables (* indicates statistical significance at 10%)

As Table 3-1 shows, most enquiries are made via phone calls to the travel brand's call centre, with there being on average three times as many offline enquiries per day as opposed to online enquiries. This may be because the travel brand books luxury holidays, which are often booked by older, more financially well-established consumers who may prefer to directly discuss their holiday plans with a consultant rather than go through the process online. Additionally, luxury holidays tend to be more customized and complicated to organise, meaning that prospective customers may find it easier to discuss their plans over the phone rather than explain them via an online form.

Table 3-1 also reveals strong skewness in most of the offline media channels, with them having low medians compared to the mean. This is because offline media campaigns, apart from TV, tend to be more infrequently used by the media agency for the travel brand. Outdoor (out-of-home) expenditure is evenly spaced over the number of days the campaign is running (e.g. if a billboard costs $10,000 and runs for 10 days, expenditure is recorded as $1000 per day). Additionally, several market conditions variables, namely the AUD/CHF and AUD/USD exchange rates, were collected and considered for use due to their possible influence on the travel client's business. However, these were found to be insignificant in the models and were removed from the analysis.

Further analysis of online and offline enquiries reveals reduced enquiries on Saturdays and Sundays. This can be seen in Figure 3-2, which particularly highlights a reduction in average offline enquiries on weekends due to reduced call centre operations.

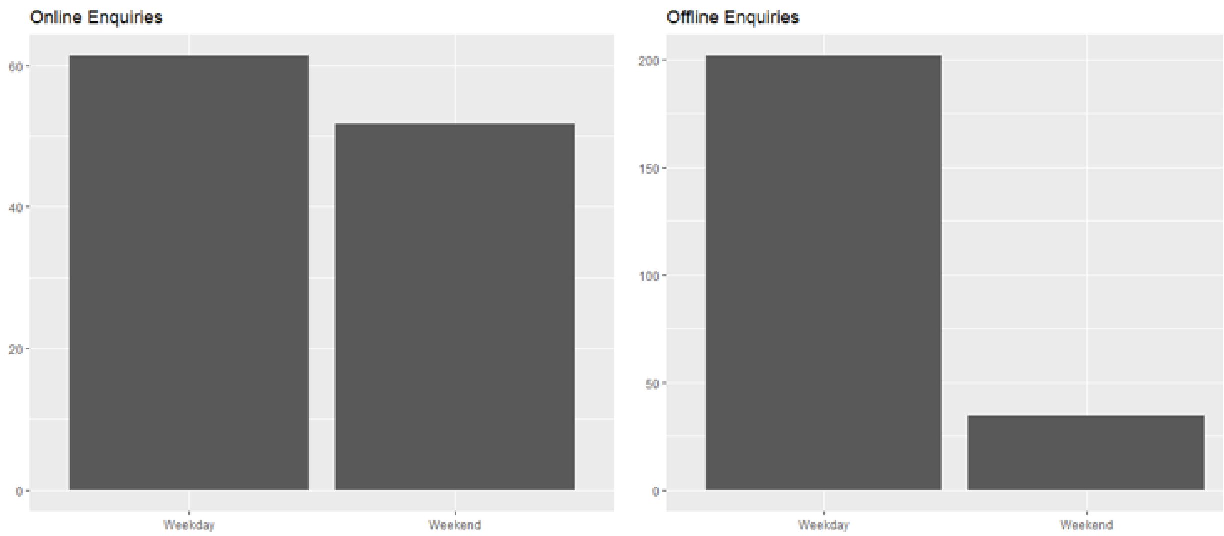


Figure 3-2 - Enquiries on Weekdays and Weekends

Additionally, Figure 3-3 shows an increase in average online and offline enquiries following a company rebrand that launched on 01 October 2018.

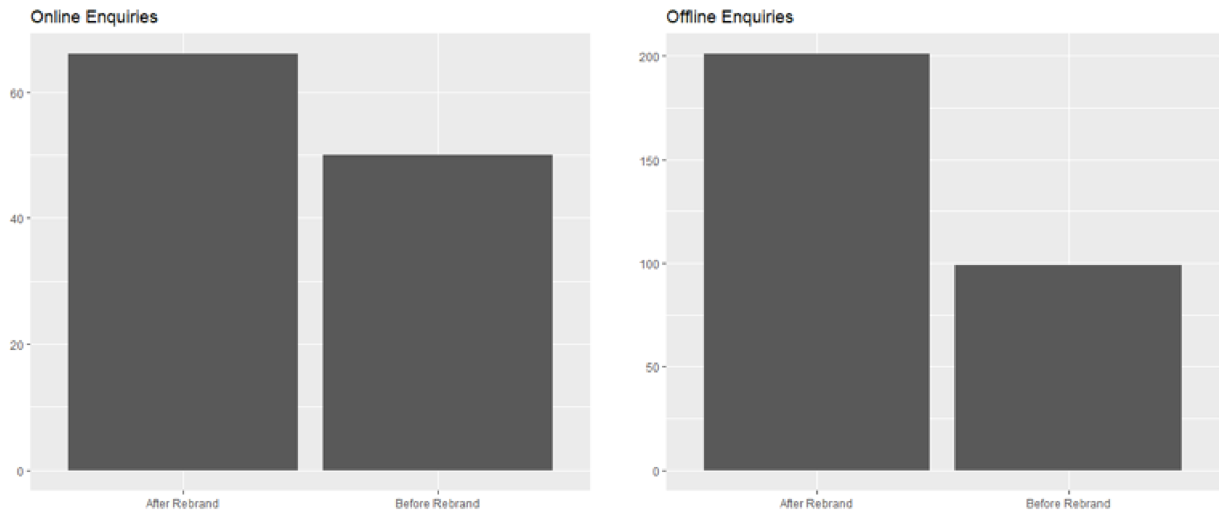


Figure 3-3 - Enquiries Before and After Company Rebrand

To facilitate the analysis of advertising effectiveness at a more aggregated level (e.g. how effective is online versus offline media channels), the variables relating to media channels in the dataset have been classified according to several prominent categorisations used in the literature. These are whether a media channel is online or offline, whether it is firm-initiated (FIC) or customer-initiated (CIC) and whether it is paid, owned or earned. An additional category describing whether a variable is an indicator of advertising quality or volume has also been included. For example, an increase in paid search spend will increase the amount of people who see the ad, while an increase in paid search engagement, with everything else kept constant, suggests that the paid search ads' ability to engage consumers have improved. The Quality/Volume category can help measure and separate the effect an increase in advertising quality, as well as expenditure, has on enquiries. This can provide insights into the trade-off between simply increasing advertising expenditure versus investing in improving the advertisements. The variable classifications are shown in Table 3-2.

Media Channel	**Online/Offline**	**FIC/CIC**	**Paid/Owned/ Earned**	**Quality/ Volume**
Display Spend	Online	FIC	Paid	Volume
Facebook Ads Spend	Online	FIC	Paid	Volume
Magazine Spend	Offline	FIC	Paid	Volume
Outdoor Spend	Offline	FIC	Paid	Volume
Paid Search Spend	Online	CIC	Paid	Volume
TV Spend	Offline	FIC	Paid	Volume
Video Spend	Online	FIC	Paid	Volume
Website Page Views	Online	CIC	Owned	Volume

Media Channel	Online/Offline	FIC/CIC	Paid/Owned/ Earned	Quality/ Volume
Paid Search Engagement Rate	Online	CIC	Paid	Quality
Facebook Page Posts Engagement	Online	CIC	Owned	Quality

Table 3-2 - Media Channel Classification

3.4 Methodology

This research requires a flexible methodology that can model direct and indirect relationships between media channels, consumer activity metrics and marketing performance metrics over time. Persistence modelling is a well-established approach to this task, using multivariate time-series techniques to simultaneously model the short and long-term relationships between variables, while handling endogeneity issues and incorporating exogenous variables. Persistence models also allow the data to define the relationships between variables, rather than pre-specifying them, which ensures that the methodology is generalisable to companies with different sets of media channels. The steps in the modelling approach are outlined in Table 3-3.

Methodological Step	Relevant Literature	Research Question
1. Stationarity Testing Augmented Dickey-Fuller (ADF) test Structural break test	Fuller (1996), Perron (1989)	Are variables stationary or evolving over time? Are there structural breaks in variables?
2. Model Building VAR model	Dekimpe and Hanssens (1999)	How do endogenous variables interact in the short run?
3. Advertising Effectiveness Impulse Response Functions (IRFs)	Lütkepohl (2007)	What is resulting change in an endogenous variable when another endogenous variable is shocked by 1 standard deviation?
4. Marketing Metrics Return on Investment (ROI), Cost Per Acquisition (CPA), Budget Allocation	ROI/CPA - Dinner et al. (2014) Budget Allocation - Danaher and Dagger (2013), de Haan et al. (2016)	How many enquiries does $1 of advertising spend gain? How much advertising spend is required to elicit one enquiry? How can budget allocation be improved based on elasticities of paid media channels?

Table 3-3 - Overview of Methodological Steps for Quantifying Advertising Media Effectiveness

First, stationarity testing was performed on all variables to determine whether differencing or cointegration testing was required. The results of this can be found in Table 3-1, which shows that all variables aside from TV advertising expenditure are stationary. While TV Spend did not report a stationary result for the ADF test, it was found to be stationary after a structural break test was conducted. The non-stationary result in the ADF test is likely due to the on-off nature of TV advertising campaigns for the luxury travel brand. Given the stationary testing results, a vector autoregression (VAR) model was chosen and specified as:

$$Y_t = A + \sum_{p=1}^{n} \beta_p Y_{t-p} + \theta X_t + \varepsilon_t, t = 1, 2, \ldots, T, \quad (3\text{-}1)$$

where Y_t is a $m \times 1$ vector of endogenous variables, A is a $m \times 1$ vector of intercepts, β_p is a $m \times m$ matrix of parameters for lag p, X_t is a vector of exogenous variables, θ is the parameter matrix for the exogenous variables and ε_t is a $m \times 1$ vector of zero mean, independent and identically distributed residuals. A $Y_t = \ln(W_t + 1)$ transformation was applied to the advertising variables (W_t) before modelling. A log-log specification allows parameter estimates to be interpreted as elasticities, captures the diminishing returns effect of advertising variables, as well as the multiplicative relationships between them. Adding one before this transformation prevented undefined results in variables that contain values of zero for some days (e.g. no TV advertising expenditure). Two exogenous variables were also included in the model. A weekend dummy variable was added to account for reduced enquiries on Saturdays and Sundays, while another dummy variable was added to control for the increase in enquiries following the company's rebrand on 01 October 2018. In the model building stage, an optimal lag length of 1 was selected using the Bayesian information criterion (BIC). The BIC was used to select the optimal lag length as it tends to produce more parsimonious models than other information criterion.

Since all variables in a VAR model depend on each other, individual coefficient estimates only provide limited information on the reaction of the system to a shock or the effect of one variable on another. Instead, impulse response functions (IRFs) were used to quantify how a shock (i.e. 1 standard deviation increase) to a variable affects another variable after a certain time period (e.g. t = 0, t = 1), holding all else constant. Since a log-log specification was used when estimating the model, the resulting IRFs can be treated as elasticities. The IRFs of the variables in the VAR model, such as TV spend, provide measures of their effects on other variables, such as website page views or offline enquiries, and show how those effects diffuse over time. IRF estimates for each variable are obtained for 30 periods (days) after the shock so that the full effects of the variable can be captured and because advertising expenditure for the travel brand is finalized thirty days in advance. Specifically, the IRFs are estimated using the Cholesky (orthogonal) approach, which involves applying the Cholesky decomposition to the VAR model's variance-covariance matrix and using this information to incorporate contemporaneous information about the shocks of one variable to another (Lütkepohl 2007).

Since orthogonal IRFs are sensitive to variable ordering, researchers must specify a causal ordering among the endogenous variables in the VAR model. The variable ordering for this research is shown in Figure 3-4, with variables being ordered by the categories outlined in the conceptual framework (see 3.2). As a robustness check, several different variable orderings were tried within each category. Additionally, results were replicated using generalized IRFs, which do not require the specification of a causal ordering. These variations produced no material differences in results. In terms of consumer activity metrics, the paid search engagement rate is ordered first as it is directly related to paid search advertising spend (Dinner et al. 2014). The paid search engagement rate could also be influenced by other spend variables, such as someone searching for the brand after seeing a billboard advertisement. Engagement rates for display, video and Facebook advertising are not included as these are purchased on a cost-per-impressions basis, rather than a cost-per-click basis. Consumers also do not get a choice when they see display, video and Facebook advertising, whereas they encounter paid search advertising after searching a term related to the travel band. This means that the engagement rates for display, video and Facebook advertising would be driven primarily by expenditure in their own media channels, whereas paid search advertising can be driven by expenditure in other media channels as well as its own. Facebook page posts engagement are included next as seeing paid advertising could lead consumers to the travel brand's Facebook page. Finally, website page views are included as online enquiries are made through the travel brand's website, meaning that its website is often the last step in the path to purchase.

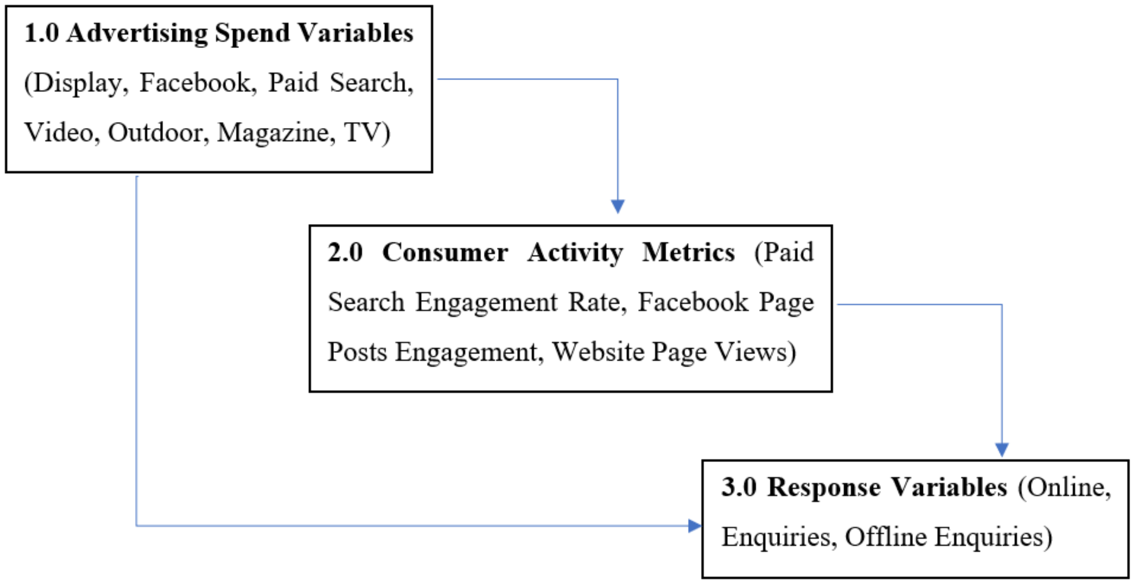


Figure 3-4 – VAR Model Variable Ordering

The orthogonal IRFs were bootstrapped to enable standard errors to be estimated. First, a VAR model was estimated with the original data, producing OLS/MLE estimates of coefficients (β and θ) and errors (ε). The errors were then sampled with replacement 10,000 times (number of draws – N) and used to build 10,000 new time series of Y. For each new time series, a VAR model and IRFs were estimated. An example bootstrapping process with an intercept A is outlined below (O'Hara 2018):

1. Let $n \in [1, N], t \in [1, T]$ and fix $Y_0, Y_{-1}, \ldots, Y_{-p+1}$. Using β, calculate:
$$Y_1^n = A + \beta_1 Y_0 + \beta_2 Y_{-1} + \cdots + \beta_p Y_{-p+1} + \varepsilon_1^n$$
2. Then using Y_1^n:
$$Y_2^n = A + \beta_1 Y_1^n + \beta_2 Y_0 + \cdots + \beta_p Y_{-p+2} + \varepsilon_2^n$$
3. Continue calculations like this:
$$Y_3^n = A + \beta_1 Y_2^n + \beta_2 Y_1^n + \cdots + \beta_p Y_{-p+3} + \varepsilon_3^n$$
$$\ldots$$
$$Y_T^n = A + \beta_1 Y_{T-1}^n + \beta_2 Y_{T-2}^n + \cdots + \beta_p Y_{T-p} + \varepsilon_T^n$$
4. With new time series Y^n, estimate VAR model and store β^n
5. Go back to Step 1 and repeat for $n = n + 1$
6. After repeating process N times, estimate IRFs for each β^n

After bootstrapping, the mean and standard errors of the bootstrapped IRFs were used to assess whether each IRF is statistically different from zero at a 10% level of significance (de Haan et al. 2016; Pauwels et al. 2016b). Summing up all the statistically significant IRF estimates yields the total direct effect of one endogenous variable (impulse) on another (response). The results of these marketing effectiveness calculations are discussed in sections 3.5.1 and 3.5.2, while the ROI, CPA and budget allocation calculations are shown in section 3.5.3. ROI, CPA and budget allocation values are detailed in section 3.5.3 as their calculations are built on IRF values, which are discussed in sections 3.5.1 and 3.5.2. Explaining the calculations in this order helps prevent the need to go back to earlier sections when reading about ROI, CPA and budget allocation results.

3.5 Results

The results are discussed over three sections. First, the direct marketing effects uncovered in the analysis are discussed, followed by the indirect marketing effects. These results are then used to perform ROI, CPA and budget allocation calculations.

3.5.1 Individual Effects

The discussion of the direct marketing effects is broken down into separate sections, with each section focusing on a separate consumer activity metric or enquiry variable. For each variable, the statistically significant IRF estimates are summarized in two visualisations, with one showing the breakdown of impulses over time and the other highlighting the total direct effects[2]. Table 3-4 summarizes all total direct effects, along with their start period, end period and duration. Table 3-5 summarizes the orthogonal IRF results (total direct effects) in matrix form for additional understanding.

Impulse	**Response**	**Total IRF**	**Start Period**	**End Period**	**Total Periods**
Facebook Ads Spend	Facebook Page Posts Engagement	0.4677	1	5	5
Paid Search Engagement Rate	Offline Enquiries	0.2048	2	9	8
Paid Search Spend	Offline Enquiries	0.0871	2	3	2
TV Spend	Offline Enquiries	0.0829	1	1	1
Magazine Spend	Offline Enquiries	0.0738	1	1	1
Website Page Views	Offline Enquiries	0.0656	2	2	1
Display Spend	Offline Enquiries	0.0415	1	1	1
Facebook Ads Spend	Online Enquiries	0.5864	1	12	12

[2] Confidence intervals are excluded from the visualisations as they crowd the plots, making it difficult to compare multiple variables on the same graph.

Impulse	Response	Total IRF	Start Period	End Period	Total Periods
Facebook Page Posts Engagement	Online Enquiries	0.4648	2	16	15
Website Page Views	Online Enquiries	0.2093	1	2	2
Outdoor Spend	Online Enquiries	0.0907	2	5	4
Magazine Spend	Online Enquiries	0.0404	2	2	1
Display Spend	Paid Search Engagement Rate	0.0179	3	9	7
TV Spend	Paid Search Engagement Rate	0.0054	3	4	2
Paid Search Spend	Paid Search Engagement Rate	-0.0069	1	1	1
Facebook Ads Spend	Website Page Views	0.3636	1	9	9
Facebook Page Posts Engagement	Website Page Views	0.2985	2	15	14
Video Spend	Website Page Views	0.2227	2	14	13
Paid Search Engagement Rate	Website Page Views	0.1628	1	7	7
Paid Search Spend	Website Page Views	0.0786	1	2	2
Magazine Spend	Website Page Views	0.0372	2	2	1

Table 3-4 - Total Direct Effects of Media Channels on Consumer Activity Metrics and Enquiry Variables

Impulse	Paid Search Engagement Rate	Facebook Page Posts Engagement	Website Page Views	Online Enquiries	Offline Enquiries
Display Spend	0.0179	NA	NA	NA	0.0415
Facebook Ads Spend	NA	0.4677	0.3636	0.5864	NA
Facebook Page Posts Engagement	NA	NA	0.2985	0.4648	NA
Magazine Spend	NA	NA	0.0372	0.0404	0.0738
Outdoor Spend	NA	NA	NA	0.0907	NA
Paid Search Engagement Rate	NA	NA	0.1628	NA	0.2048
Paid Search Spend	-0.0069	NA	0.0786	NA	0.0871
TV Spend	0.0054	NA	NA	NA	0.0829
Video Spend	NA	NA	0.2227	NA	NA
Website Page Views	NA	NA	NA	0.2093	0.0656

Table 3-5 - Matrix of Orthogonal IRF Results (NA means not statistically significant or applicable)

3.5.1.1 Online Enquiries

Five media channels have a statistically significant effect on online enquiries. As shown in Figure 3-5 and Figure 3-6, Facebook ads spend has the largest total direct effect, followed by Facebook page posts engagement. The statistically significant result for Facebook page posts engagement matches Kumar et al. (2016), who found that firm-generated social media content has a positive effect on customers' behaviour. It also shows that increasing owned media channel engagement results in more enquiries, meaning that it is worthwhile for companies to invest in increasing consumer engagement (Srinivasan et al. 2016). The large effects for Facebook ads and Facebook page posts highlight the importance of Facebook media channels in driving online enquiries for this company. Furthermore, it shows that to understand the full benefits of a social media platform, firms should consider both the paid and owned advertising aspects of it. Interestingly, there are differences in the size and timing of the paid and owned Facebook effects. Facebook ads expenditure has a large and immediate effect on online enquiries, before having a reduced effect over subsequent periods. Facebook page posts engagement has a smaller initial effect than Facebook ads expenditure, but its effect becomes larger after the fourth period and lasts an additional four periods. As found in Srinivasan et al. (2016), the effects of Facebook page posts engagement were also delayed, not starting until the second period. This suggests that paid media channels have larger and more immediate effects, while the effects of owned media channels tend to be delayed and smaller initially, but more sustained over time.

The direct effects from the outdoor and magazine media channels show the presence of cross-channel effects, where offline media channels drive online conversion (Dinner et al. 2014; Srinivasan et al. 2016). More specifically, some consumers are driven to make an online enquiry after seeing a billboard or magazine advertisement. The influence of the offline media channels on online enquiries is smaller and lasts for less time than that of the online media channels. The large and short-lasting effect of website page views on online enquiries shows that the more time consumers spend on the travel brand's website, the more likely they are to make an online enquiry. Additionally, the more traffic driven to the website, the more likely an immediate online enquiry will be made. Thus, the travel brand's website is an important part of the path to purchase.

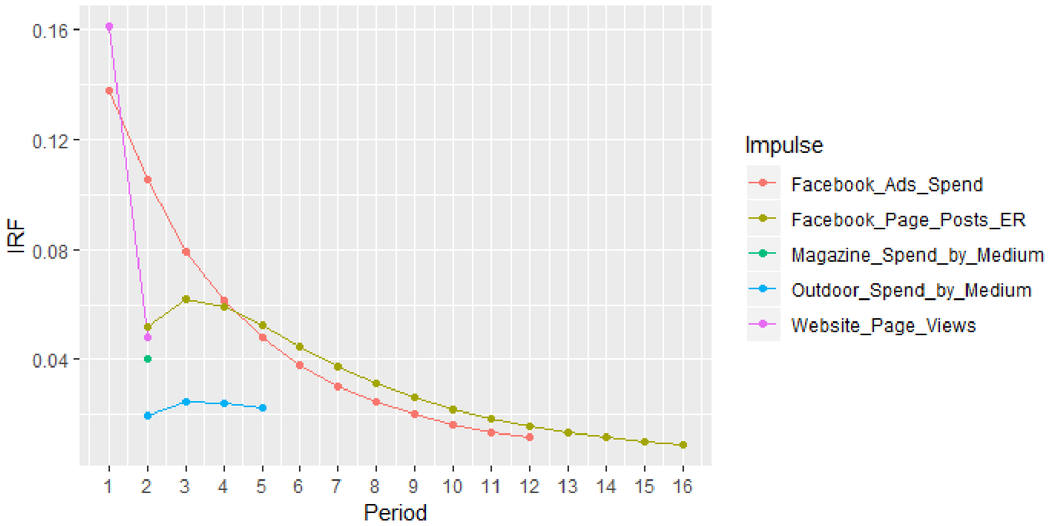


Figure 3-5 - IRF Effects on Online Enquiries Over Time

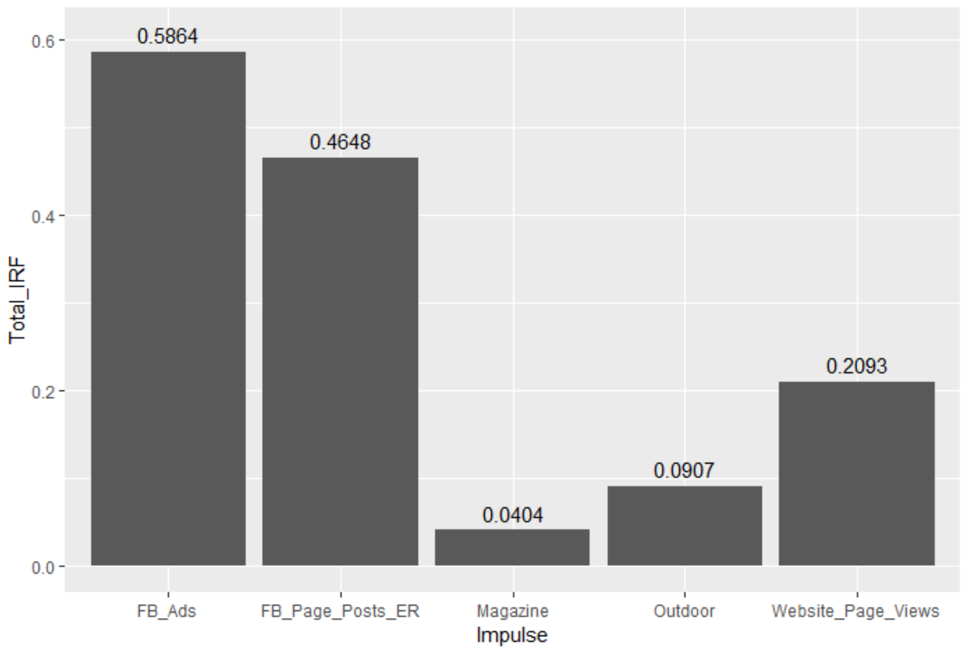


Figure 3-6 - IRF Totals for Online Enquiries

3.5.1.2 Offline Enquiries

Figure 3-7 and Figure 3-8 show that both online and offline media channels have direct effects on offline enquiries, with TV advertising having the largest effect in a single period. The effects of the offline media channels are larger than the online media channels on a per period basis, but the influence of paid search lasts for longer. The effects of the offline media channels are immediate, meaning that consumers generally see an advertisement and make an offline enquiry on the same day. This contrasts to the more delayed effects of the online media channels, indicating that consumers who search online for brand spend some additional time gathering additional information before making an offline enquiry. Interestingly, both variables related to paid search, namely expenditure and engagement rate, have statistically significant direct effects on offline enquiries. Increasing the paid search engagement rate by one standard deviation has a smaller immediate effect than increasing paid search expenditure, but the effect is more sustained over time. Thus, it can be more beneficial for firms to focus on increasing the engagement rate of paid search, rather than simply spending more money on the media channel. This could involve improving the quality of paid search advertisements or increasing investment in other media channels that might drive consumers to a paid search advertisement. The presence of direct effects for both paid search variables suggest that including a consumer activity metric, in addition to a firm activity metric, can provide additional insight into the effects of customer-initiated paid media channels.

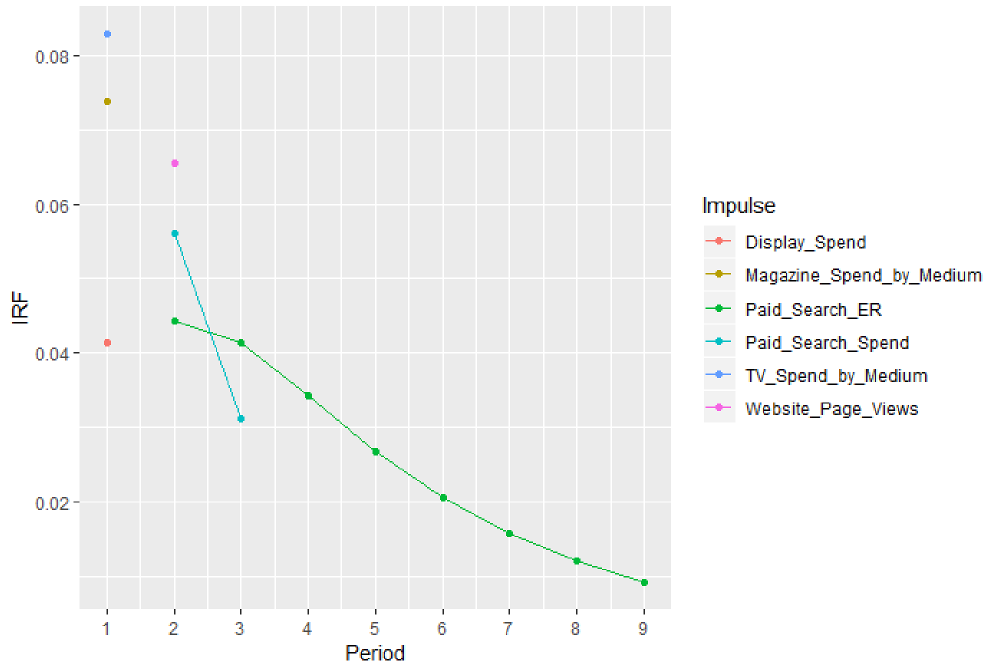


Figure 3-7 - IRF Effects on Offline Enquiries Over Time

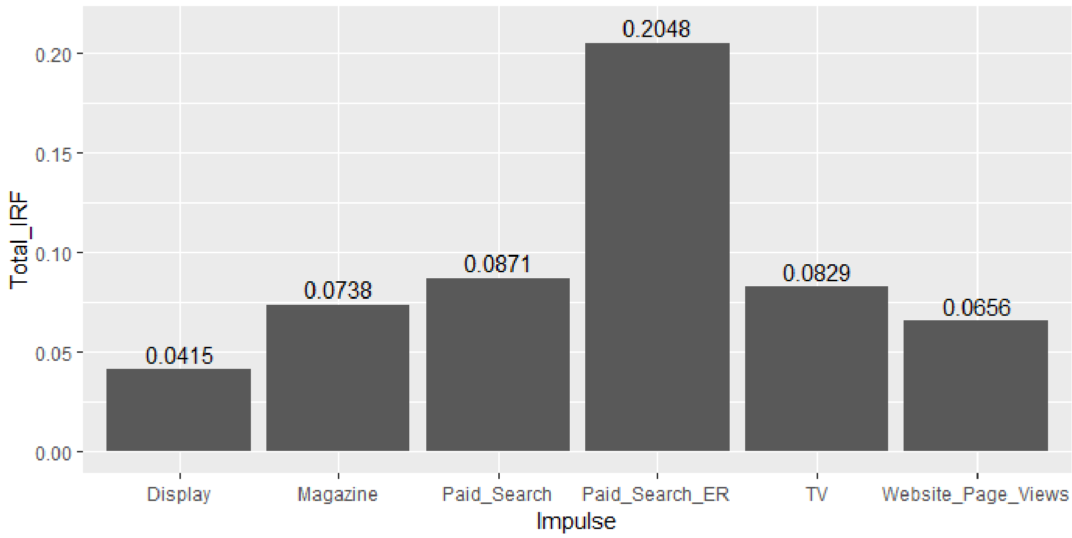


Figure 3-8 - IRF Totals for Offline Enquiries

3.5.1.3 Paid Search Engagement Rate

Figure 3-9 and Figure 3-10 reveal the existence of three direct effects on the engagement rate of paid search advertising. Increasing TV and display advertising expenditure have delayed positive effects on the paid search engagement rate. TV and display are firm-initiated media channels, which appear to generate awareness about the travel brand and prompt consumers to search for more information about it online. As seen in sections 3.5.1.1 and 3.5.1.2, this can lead to a consumer making an enquiry. The positive effect of display advertising matches the results of Kireyev et al. (2016), who find that display ads can increase search conversion and clicks. The relationship between TV and paid search advertising was also identified by Srinivasan et al. (2016), showing that offline media channels can positively influence online consumer activity metrics. The effects for TV and display are not felt until the third day, suggesting that consumers who decide to further investigate the travel brand online may not immediately do so after seeing an advertisement. These consumers contrast those described in section 3.5.1.2, who make an offline enquiry on the same day as being exposed to TV and display advertising.

Interestingly, increasing paid search expenditure has an immediate negative effect on the paid search engagement rate. This finding reinforces the results of Dinner et al. (2014), who find that paid search expenditure had a negative association with click-through rates. Dinner et al. (2014) suggest that this negative relationship is a result of firms bidding on more esoteric keywords with lower click-through rates as their paid search advertising budgets increase. Therefore, firms should be careful about selecting appropriate keywords for paid search advertising and be aware of the diminishing returns of increasing their overall budget allocation towards this media channel.

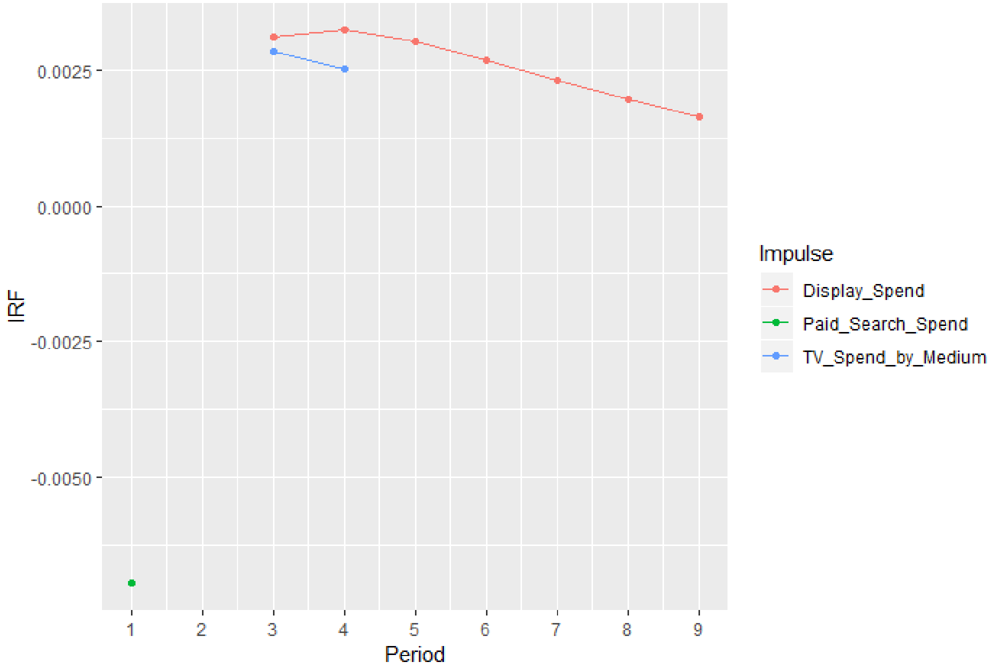


Figure 3-9 - IRF Effects on Paid Search Engagement Rate Over Time

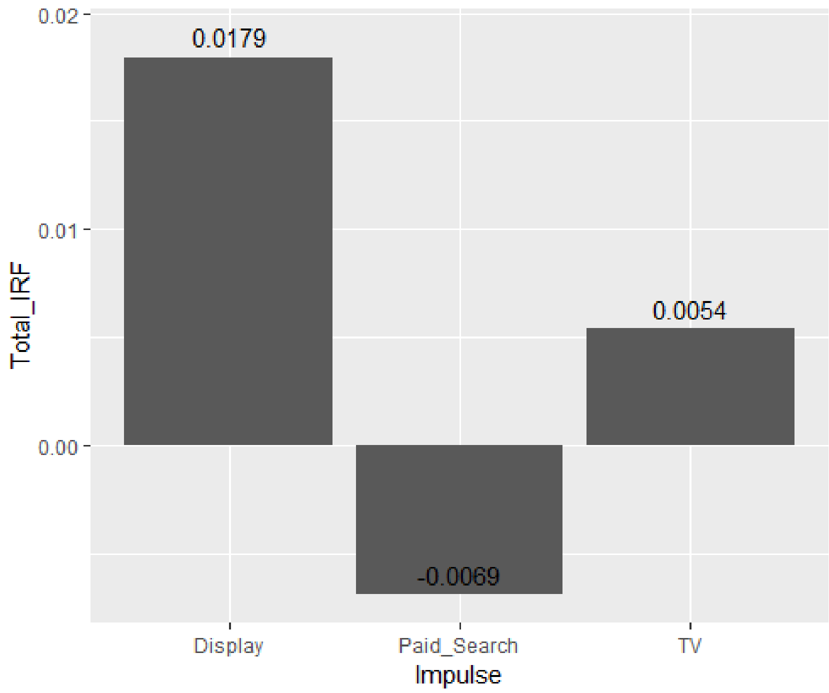


Figure 3-10 - IRF Totals for Paid Search Engagement Rate

3.5.1.4 Facebook Page Posts Engagement

Paid Facebook advertising is the only media channel to have statistically significant effects on Facebook page posts engagement, as can be seen in Figure 3-11 and Figure 3-12. The effects last for five days and show that increasing investment in paid Facebook ads can increase engagement with posts from a company's Facebook page. More specifically, after being exposed to a paid Facebook ad, consumers are driven to the travel brand's owned Facebook page and begin engaging with its content. This shows that paid media channels can improve owned media channels, as well as having a direct effect on customer conversion. Considering this, advertising content creators across paid and owned media channels should coordinate their efforts to take advantage of the relationships between them. The direct linkage between paid and owned Facebook advertising also reinforces the importance of considering all aspects of a company's social media presence when assessing advertising effectiveness.

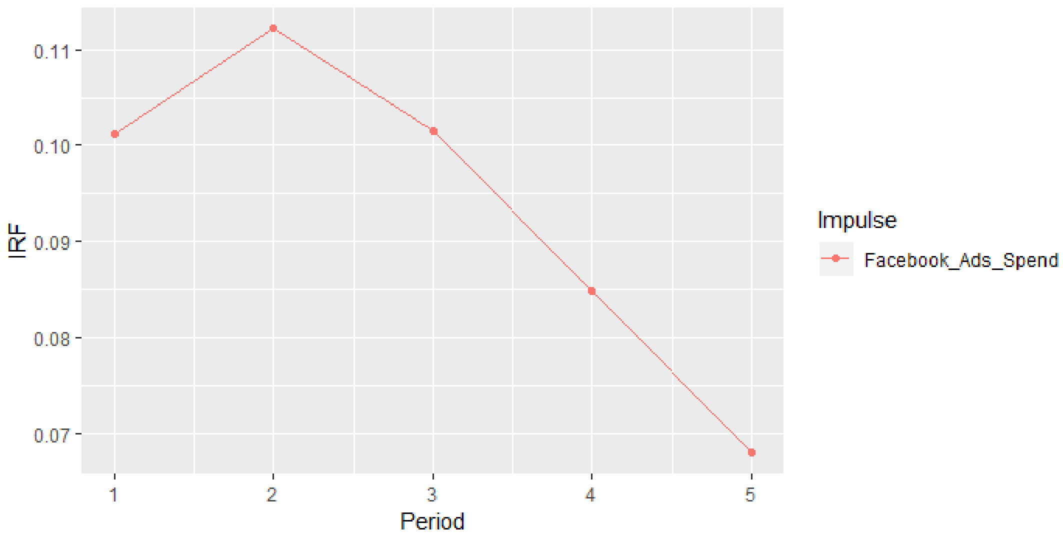


Figure 3-11 - IRF Effects on Facebook Page Posts Engagement Over Time

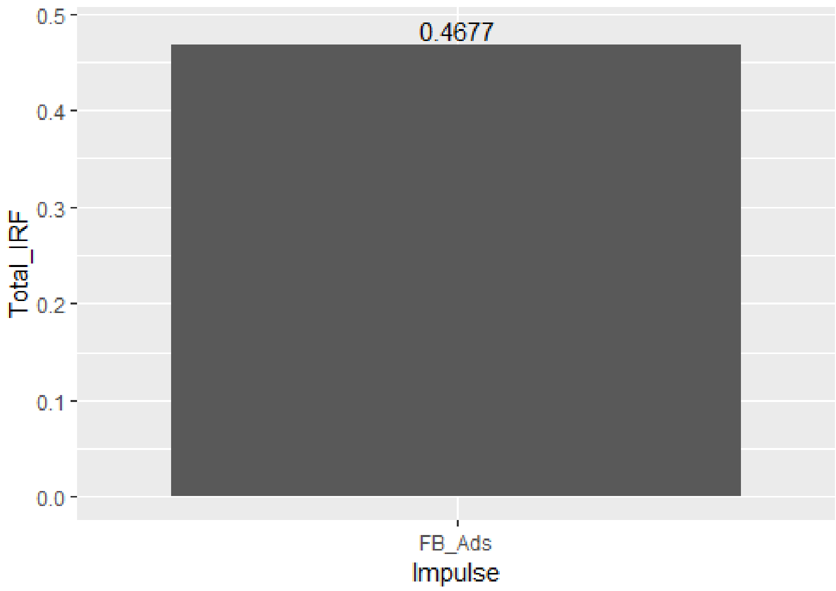


Figure 3-12 - IRF Totals for Facebook Page Posts Engagement

3.5.1.5 Website Page Views

As Figure 3-13 and Figure 3-14 show, website page views are predominantly directly influenced by online media channels, with paid Facebook advertising have the largest effect. The positive effects of Facebook ads, magazines, paid search and Facebook page posts demonstrate that paid and owned media channels can not only directly increase enquiries, but also drive consumers to find out more information about the travel brand via its website (Wiesel et al. 2011). Notably, the effect of increasing the paid search engagement rate lasts longer and is larger than increasing paid search expenditure. This suggests that it may be more effective for the travel brand to focus on improving the engagement rate of paid search advertising, rather than simply increasing its budget allocation. The negative relationship between paid search expenditure and the paid search engagement rate identified in section 3.5.1.3 reinforces this finding. Thus, simply increasing expenditure in an advertising channel is not always the best way to improve its effectiveness. Another interesting finding relates to paid video advertising. While video advertising does have any statistically significant effects on enquiries or other consumer activity metrics, it has a relatively long-lasting positive effect on website page views. This suggests that while video advertising may not immediately incite the average consumer to make an enquiry, it does drive them to the travel brand’s website, which may in turn result in an online or offline enquiry. The effects of paid video advertising echo the conclusions of Danaher and Dagger (2013), who report that while some online media do not have a direct effect on purchase outcomes, they can increase traffic to a retailer's website, which may increase the chance of a sale.

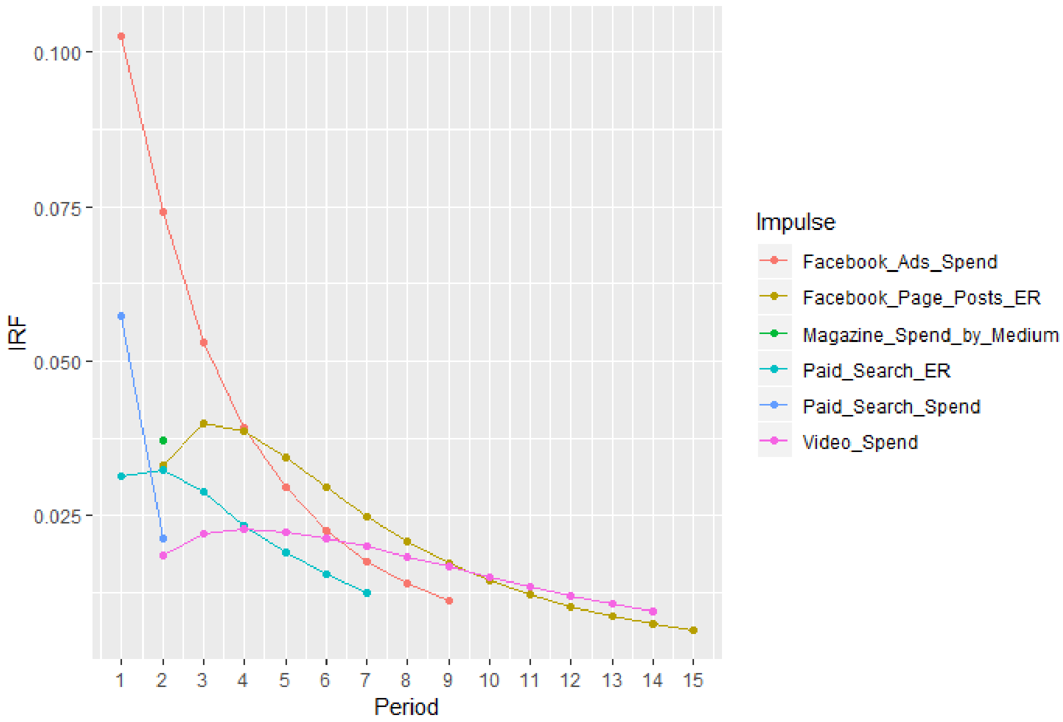


Figure 3-13 - IRF Effects on Website Page Views Over Time

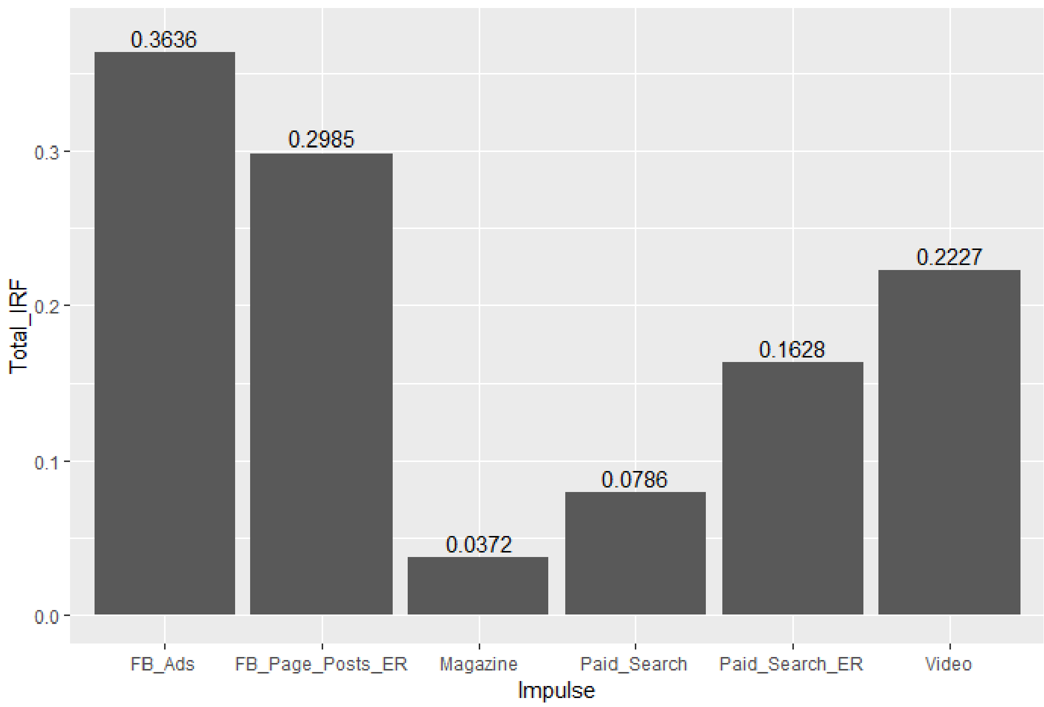


Figure 3-14 - IRF Totals for Website Page Views

3.5.2 Direct and Indirect Effects

The orthogonal IRF estimates previously discussed describe the direct effects of a shock to one variable on another, such as the effect of an increase in TV expenditure on offline enquiries. However, relationships between marketing variables are not always so straightforward. While media channels may have direct effects on marketing performance metrics, they may also exert indirect effects via other media channels. For example, one marketing variable may have an effect on another marketing variable, which in turn could have an effect on a marketing performance metric. Indeed, Dinner et al. (2014) and Srinivasan et al. (2016) show that TV advertising creates awareness that drives consumers to paid search advertising, which in turn results in additional sales. Thus, a shock to TV expenditure might not have a direct effect on sales, but could have an indirect effect via paid search. More generally, while paid media channels can influence marketing performance metrics directly, they can also influence them indirectly via consumer activity metrics (Srinivasan et al. 2016). To build a complete picture of the effectiveness of paid media channels, both direct and indirect effects need to be considered.

The results presented in section 3.5.1 show that the travel brand's paid media channels have statistically significant effects on the consumer activity and marketing performance metrics. In turn, the consumer activity metrics have statistically significant relationships with online and offline enquiries. Thus, the travel brand's paid media channels have indirect effects on online and offline enquiries via these consumer activity metrics. Each indirect effect can be quantified by multiplying the total direct effects in the path from the paid media channel to the marketing performance metric together. For example, if a shock of 1 to Facebook ads spend increases website page views by 0.1 and a shock of 1 to website page views increases online enquiries by 0.2, the indirect effect of Facebook ads spend on online enquiries is 0.02 (0.1×0.2). The total effect of a paid media channel on a marketing performance metric is calculated by adding up its direct and indirect effects. This approach is similar to Srinivasan et al. (2016), who trace the effects of shocks to marketing variables using generalized impulse response functions (GIRFs) and Granger Causality, finding both direct and indirect effects. For example, increasing TV advertising by 1 GRP for a leading brand of a low-involvement, paper-based product not only increases sales by 137 units directly, but also increases paid search clicks by 9 units, each of which in turn increases sales by 40 units (Srinivasan et al. 2016). Granger

Causality is not used for this research as the causal ordering among variables is given by research and indicated in the orthogonal IRFs.

An example of the calculation of the indirect and total effects of magazine advertising on online and offline enquiries is presented below. As seen in section 3.5.1, magazine spend has a direct effect on offline enquiries of 0.0738, a direct effect on online enquiries of 0.0404 and a direct effect on website page views of 0.0372. However, website page views also have a direct effect on online enquiries of 0.2093 and offline enquiries of 0.0656. Therefore, magazine spend has indirect effects on online enquiries and offline enquiries via website page views:

$$Magazine\ Spend\ \rightarrow\ Website\ Page\ Views\ \rightarrow\ Online\ Enquiries$$
$$=\ 0.0372\ \times\ 0.2093\ =\ 0.0078 \tag{3-2}$$

$$Magazine\ Spend\ \rightarrow\ Website\ Page\ Views\ \rightarrow\ Offline\ Enquiries$$
$$=\ 0.0372\ \times\ 0.0656\ =\ 0.0024 \tag{3-3}$$

$$Total\ Effect\ of\ Magazine\ Spend\ on\ Online\ Enquiries$$
$$=\ Direct\ Effects\ +\ Indirect\ Effects$$
$$=\ 0.0404\ +\ 0.0078\ =\ 0.0482 \tag{3-4}$$

$$Total\ Effect\ of\ Magazine\ Spend\ on\ Offline\ Enquiries$$
$$=\ Direct\ Effects\ +\ Indirect\ Effects$$
$$=\ 0.0738\ +\ 0.0024\ =\ 0.0762 \tag{3-5}$$

Appendix 1 provides a detailed breakdown of the direct and indirect effects for each impulse variable, while Figure 3-15 and Table 3-6 provide an overall summary of these relationships. An appendix was created because including a detailed breakdown of the results of each variable in the main text would reduce its readability. Figure 3-15 shows the percentage breakdown of the total direct and indirect effects of each media channel on online and offline enquiries. As can be seen, several media channels only have indirect effects on the marketing performance metrics. This shows the importance of considering indirect effects to capture all the relationships between media channels and marketing performance metrics. Table 3-6 details the total direct and indirect effects of each impulse variable on online and offline enquiries. While direct effects tend to be larger than indirect effects for this company, indirect effects are still sizeable, contributing 13.82% of the total effects for offline enquiries and 26.07% of the total effects for online enquiries. Additionally, Appendix 1 shows that the total effects of media channels on marketing performance metrics become greater and last for longer when indirect effects are included.

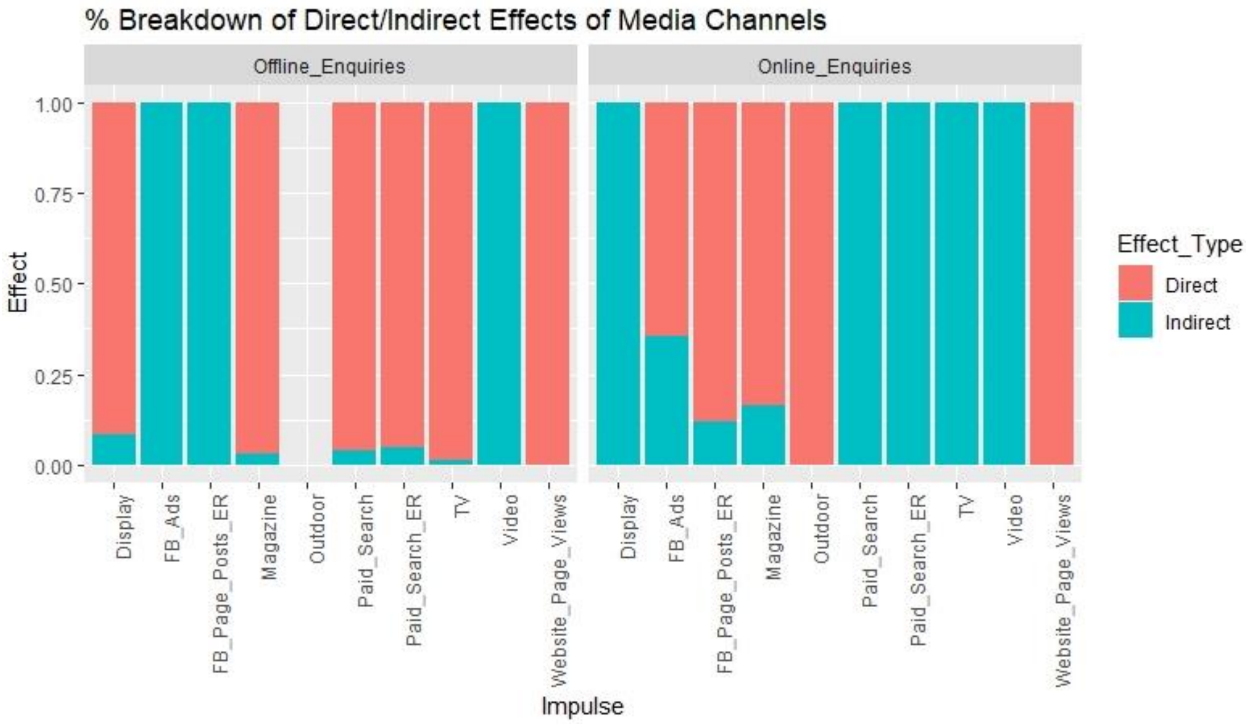


Figure 3-15 - Percentage Breakdown of Direct and Indirect Effects of Media Channels on Online and Offline Enquiries

Impulse	Response	Direct	Indirect	Total
Display Spend	Offline Enquiries	0.0415	0.0039	0.0454
Display Spend	Online Enquiries	0.0000	0.0006	0.0006
Facebook Ads Spend	Offline Enquiries	0.0000	0.0330	0.0330
Facebook Ads Spend	Online Enquiries	0.5864	0.3227	0.9091
Facebook Page Posts Engagement	Offline Enquiries	0.0000	0.0196	0.0196
Facebook Page Posts Engagement	Online Enquiries	0.4648	0.0625	0.5273
Magazine Spend	Offline Enquiries	0.0738	0.0024	0.0762
Magazine Spend	Online Enquiries	0.0404	0.0078	0.0482
Outdoor Spend	Offline Enquiries	0.0000	0.0000	0.0000
Outdoor Spend	Online Enquiries	0.0907	0.0000	0.0907
Paid Search Engagement Rate	Offline Enquiries	0.2048	0.0107	0.2155
Paid Search Engagement Rate	Online Enquiries	0.0000	0.0341	0.0341
Paid Search Spend	Offline Enquiries	0.0871	0.0037	0.0908
Paid Search Spend	Online Enquiries	0.0000	0.0162	0.0162
TV Spend	Offline Enquiries	0.0829	0.0012	0.0841
TV Spend	Online Enquiries	0.0000	0.0002	0.0002
Video Spend	Offline Enquiries	0.0000	0.0146	0.0146
Video Spend	Online Enquiries	0.0000	0.0466	0.0466
Website Page Views	Offline Enquiries	0.0656	0.0000	0.0656

Impulse	Response	Direct	Indirect	Total
Website Page Views	Online Enquiries	0.2093	0.0000	0.2093
All Impulses	Offline Enquiries	0.5557 (86.18%)	0.0891 (13.82%)	0.6448
All Impulses	Online Enquiries	1.3916 (73.93%)	0.4907 (26.07%)	1.8823

Table 3-6 - Direct, Indirect and Total Effects of Media Channels on Online and Offline Enquiries

3.5.3 ROI/CPA, Total Elasticity & Budget Allocation

Impulse response functions (IRFs) provide insight into the presence, size and length of the effects of changes in one marketing variable on another. While disaggregating enquiries into online and offline components yields a more granular and complete picture of marketing effects, the results can be misleading when considering the overall effectiveness of a paid media channel on total enquiries. For example, Table 3-6 shows that the media channels' IRFs are larger for online than offline enquiries. However, there are fewer online enquiries than offline enquiries on a daily basis, meaning that a one standard deviation shock to online enquiries will not generate as many enquiries as the same shock to offline enquiries. In order to quantify the effectiveness of a media channel on total enquiries, its effects on both online and offline enquiries need to be reconciled. Since IRFs can be treated as elasticities due to the VAR model specification, calculating the effect of a media channel on total enquiries is equivalent to calculating a total elasticity. This can be done using the process of Dinner et al. (2014), who observe that an elasticity is the percentage change of one variable in response to a change in another. Therefore, total elasticity can be calculated as the weighted average between online and offline elasticities, with the weights being online and offline enquiries as a percentage of total enquiries over the whole dataset (Dinner et al. 2014). Since there are more offline than online enquiries on a daily basis, the offline elasticity (IRF) of a media channel will contribute more to the total elasticity than the online elasticity. As an example, the total elasticity for magazine advertising, including both direct and indirect effects, is:

$$\begin{aligned}\eta_{Total} &= \%\,Offline\ Enquiries\ \times \eta_{offline} + \%\,Online\ Enquiries\ \times \eta_{online} \\ &= 0.7239 \times 0.0762 + 0.2761 \times 0.0482 = 0.0685\end{aligned} \tag{3-6}$$

Marketing elasticities can provide insights that can help marketing managers improve the allocation of their media channel investments. Basing marketing expenditure decisions on data, as well as experience, can also add more transparency and robustness to the budget allocation process. While it is difficult to derive a closed-form solution to optimal media spending based on elasticities derived from IRFs, recommendations for improvement can still be made using the optimization rules from other scenarios. This research follows the budget allocation procedure of Danaher and Dagger (2013), who show that the optimal budget allocation for a media channel can be calculated by dividing a media channel's total elasticity (η_i) by the sum of all of the total elasticities ($\sum_{\mathrm{i}=1}^{\mathrm{K}} \eta_{\mathrm{i}}$). While the formula used by Danaher and Dagger (2013) is based on levels of spending instead of shocks to spending (IRFs), it still represents a useful starting point for improving the media budget allocation of the luxury travel brand. The elasticities and recommended budget allocation can be re-estimated as more data is collected to capture consumer responses to the budget allocation adjustments. Since the goal here is to maximise total enquiries, total elasticities are used in the budget allocation calculations. For example, the recommended budget allocation for magazine advertising is:

$$Recommended\ Budget\ Allocation = \frac{\eta_{Magazine}}{\sum_{i=1}^{K} \eta_i} \qquad \textit{(3-7)}$$

$$= \frac{0.0685}{0.0330+0.2749+0.0702+0.0234+0.0250+0.0685+0.0609} = 12.32\%$$

In addition to improving budget allocation for a media channel, marketing managers are also usually interested in its ROI. This analysis reports the incremental revenue ROI of all paid advertising expenditure using the baseline-lift method between 01 April 2018 and 29 April 2019 (Farris et al. 2015). Revenue ROI is reported as there is no data available to relate enquiries with the luxury travel brand to its net profit (Farris et al. 2015). More specifically, the ROI of each paid media channel is calculated in the same way as Dinner et al. (2014), relating a media channel's online (η_{online}) and offline ($\eta_{offline}$) elasticities to its ROI to obtain the following:

$$ROI = \eta_{online}\frac{E_{online}}{A} + \eta_{offline}\frac{E_{offline}}{A}, \qquad \textit{(3-8)}$$

where E_{online} and $E_{offline}$ are average daily online and offline enquiries and A is the average daily expenditure in the media channel. There is no subtraction of one at the end of the ROI calculation as the ROI is being expressed in terms of revenue per advertising dollar instead of profit per advertising dollar. In addition to each paid media channel's ROI, we also calculate its CPA. CPA refers to the advertising expenditure required to obtain one unit of a marketing performance metric, while ROI refers to the amount of a marketing performance metric obtained for one dollar of advertising expenditure. ROI is typically used for continuous metrics, such as sales, while CPA tends to be used for discrete metrics, such as enquiries, as they cannot be broken into fractions. Since CPA is simply ROI from a different perspective, it can be calculated as:

$$CPA = \frac{1}{ROI} \tag{3-9}$$

As an example, the ROI and CPA calculations for magazine advertising are:

$$ROI_{Magazine} = 0.0482\frac{58.62}{1548.24} + 0.0762\frac{153.67}{1548.24} \approx 0.0094\ enquiries \tag{3-10}$$

$$CPA = \frac{1}{ROI_{Magazine}} = \$106.43 \tag{3-11}$$

These results suggest that every \$1 of investment in magazine advertising resulted in 0.0094 enquiries. Alternatively, from a more meaningful CPA perspective, every \$106.43 spent on magazine advertising resulted in one enquiry. It should be noted these are historical measures of performance and are affected by the level of historical expenditure on magazine advertising. For assessing media channel effectiveness and improving budget allocation in the future, the elasticities themselves should be used as these capture diminishing returns to advertising spend and measure media channel potential.

To provide insight into how the use of consumer activity metrics and disaggregated marketing performance metrics affects total elasticity, ROI, CPA and budget allocation calculations, five different models are compared:

- Model 1 – online/offline enquiries with consumer activity metrics
- Model 2 – total enquiries with consumer activity metrics
- Model 3 – online/offline enquiries with only paid media channels
- Model 4 – total enquiries with only paid media channels
- Model 5 – online/offline enquiries with consumer activity metrics and indirect effects included (Model 1 with indirect effects included)

Models 2-4 were created by re-estimating the VAR model with different endogenous variables. Model 1 and Model 5 calculations are based on the full VAR model, but Model 1 only uses direct effects for elasticity calculations, while Model 5 includes both direct and indirect effects (see 3.5.2). Removing the consumer activity metrics in Models 3 and 4 reveals their influence on elasticity, ROI, CPA and budget allocation calculations. Similarly, the use of total enquiries in Models 2 and 4 compared to online and offline enquiries in other Models 1, 3 and 5 highlights the differences between using aggregated and disaggregated marketing performance metrics. Finally, Model 1 and Model 5 reveal how the inclusion of indirect effects can influence measurements of advertising media effectiveness. Performance measures for each model are presented in Table 3-7 to provide a statistical point of comparison between them.

	Online Enquiries		**Offline Enquiries**		**Total Enquiries**	
	Model 1, 5	**Model 3**	**Model 1, 5**	**Model 3**	**Model 2**	**Model 4**
Rolling Out-Of-Sample RMSE	0.3369	0.3499	0.4728	0.4404	0.3459	0.3555
N-Step Out-Of-Sample RMSE	0.5009	0.4970	0.6542	0.6540	0.5541	0.5508
In-Sample RMSE	0.3468	0.3537	0.6450	0.6544	0.2870	0.2977
Adjusted R^2	0.5033	0.4867	0.6999	0.6952	0.7974	0.7889
AIC	328.7338	338.7648	804.8529	808.1371	219.0587	242.5115
BIC	392.3148	390.4244	868.4338	859.7966	278.6659	290.1972

Table 3-7 – Online, Offline and Total Enquiries Model Performance Measures

The RMSE values in Table 3-7 are based on predictions of the marketing performance metrics in each model (online/offline or total enquiries). The in-sample RMSE values are produced using models built with the first 364 observations of data, while the n-step out-of-sample RMSE values were created after using these models to predict the last 30 observations of data. The rolling out-of-sample RMSE values were based on predictions created via an expanding window approach, with a model being fitted on data up until observation *t-1* and then used to predict observation *t*. This was done for the last 30 observations of data, producing 30 one-step ahead predictions from models fitted on all the previous observations. Rolling forecasts can be more accurate than n-step ahead forecasts as they are made using models built with more recent data. However, n-step ahead forecasts are still useful as they can make predictions further into the future. Including RMSE values for both types of predictions provides insight into the models' ability to forecast in both the short-term and long-term.

Table 3-7 shows that models with consumer activity metrics perform better in terms of rolling out-of-sample and in-sample predictions for online and total enquiries, suggesting that their inclusion can improve a model's ability to make short-term predictions. However, in the case of n-step out-of-sample predictions, models without consumer activity metrics had lower RMSE values. This suggests more parsimonious models are more effective for longer-term forecasting. In terms of model fit, models with consumer activity metrics tended to have better Adjusted R^2 and AIC values, showing that these variables can add extra value to models. However, models with consumer activity metrics tended to have lower BIC values, meaning that if the goal is to create a parsimonious model, these variables could be excluded. Interestingly, the Adjusted R^2 for the offline enquiries equations is higher than that of online enquiries. This is because offline enquiries is more strongly driven by its own lags, as well as the weekend and after rebrand exogenous variables. Overall, there is not a large difference between the models in terms of forecasting performance or model fit. The consumer activity metrics appear to add a little bit of predictive value, particularly for online enquiries, but not a large amount.

For each model, the total elasticity, ROI and CPA of each paid media channel was calculated. The total elasticities were also used to determine the budget allocation recommended by each model. All calculations were rounded to four decimal places for display purposes. The ROI and CPA figures are shown in Table 3-8 and Figure 3-16 to Figure 3-17. Since ROI and CPA values are inverses of each other, they provide the same conclusions, albeit from different perspectives. More specifically, smaller CPA and larger ROI values indicate that a media channel is more effective as it generates enquiries at a lower cost.

Media Channel	**CPA (ROI)**				
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Display	51.73 (0.0193)	70.51 (0.0141)	45.14 (0.0222)	59.25 (0.0169)	47.09 (0.0212)
Facebook	112.37 (0.0089)	128.01 (0.0078)	87.70 (0.0114)	73.63 (0.0136)	66.18 (0.0151)
Paid Search	106.29 (0.0094)	165.93 (0.0060)	81.36 (0.0123)	132.76 (0.0075)	95.49 (0.0105)
Video	-	26.99 (0.0370)	55.41 (0.0180)	19.44 (0.0515)	123.61 (0.0081)
Outdoor	92.36 (0.0108)	-	-	-	92.36 (0.0108)
Magazine	112.88 (0.0089)	-	108.82 (0.0092)	-	106.43 (0.0094)
TV	176.23 (0.0057)	-	162.28 (0.0062)	-	173.66 (0.0058)

Table 3-8 –ROI and CPA Values for Paid Media Channels

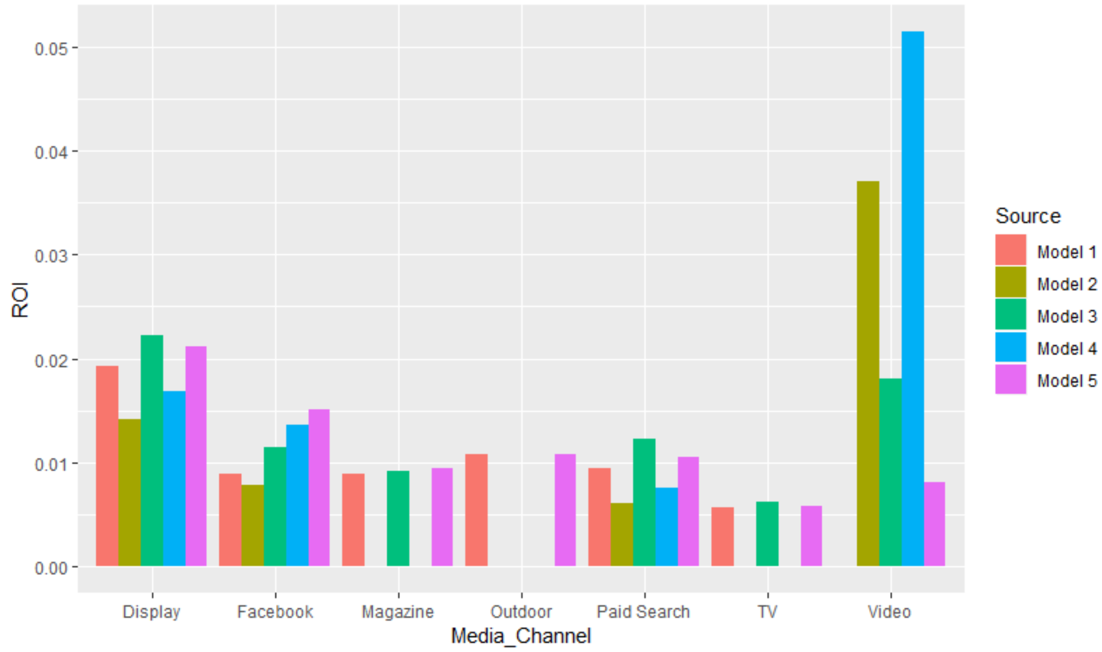


Figure 3-16 – Comparison of ROI Values for Paid Media Channels

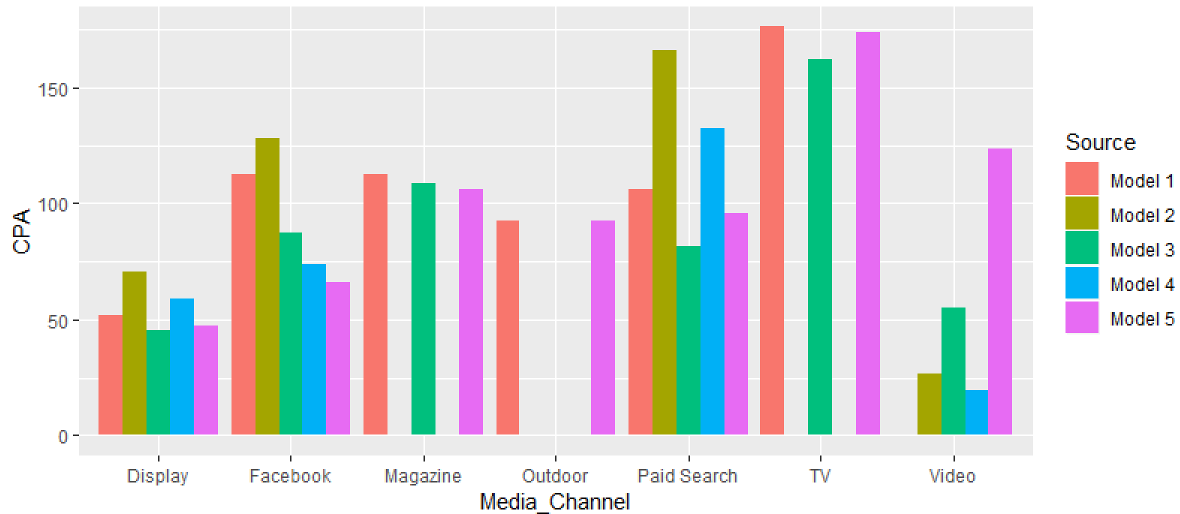


Figure 3-17 – Comparison of CPA Values for Paid Media Channels

Table 3-8 and Figure 3-16 to Figure 3-17 reveal considerable differences between the models with online and offline enquiries (Models 1, 3 and 5) and the ones with only total enquiries (Models 2 and 4). For example, display advertising has the lowest CPA for Models 1, 3 and 5, while video advertising has the lowest CPA for Models 2 and 4. Interestingly, the models with total enquiries do not find statistically significant effects for offline media channels. It is only when the enquiries are disaggregated that these effects are uncovered. Furthermore, Model 5 is the only model to detect statistically significant effects between all of the media channels and the marketing performance metrics. This suggests that aggregating enquiries or excluding indirect effects results in the loss of some information about media channel effectiveness.

There are also differences in ROI and CPA values between the models. For Models 2 and 4, these differences do not change the order of the paid media channels in terms of cost effectiveness. For Models 1, 3 and 5, the ROI and CPA values are more varied depending on whether on whether consumer activity metrics and indirect effects are included. There are some similarities, with TV being the least effective and display the most effective media channel from a CPA and ROI perspective. However, there are clear differences, such as the CPA of video advertising being much larger in Model 5 compared to Model 3. When the indirect effects of media channels are accounted for, the CPA of Facebook advertising improves considerably, falling from $112.37 in Model 1 to $66.18 in Model 5. This is because the inclusion of indirect effects captures the relationships between Facebook advertising, Facebook page posts engagement and website page views, which flow on to online and offline enquiries. Across the models, the CPAs of the paid media channels are higher when consumer activity metrics are included. This is most likely because the consumer activity metrics mediate some of their effects on online and offline enquiries, making their CPAs seem higher than they really are. This suggests that if consumer activity metrics are included in the analysis, the indirect relationships between media channels and marketing performance metrics should be accounted for in ROI and CPA calculations. Overall, ROI and CPA values vary depending on whether the underlying model incorporates disaggregated marketing performance metrics and consumer activity metrics, as well as whether the indirect effects of media channels are accounted for.

The total elasticities for the paid media channels in each model are presented in Table 3-9 and Figure 3-18. The current budget allocation of the travel brand is compared to the recommended budget allocation recommended by each of the models in Table 3-10 and Figure 3-19. As budget allocation calculations are directly based on total elasticities, larger budget allocations are given to media channels with larger total elasticities.

Media Channel	**Total Elasticity (used in budget allocation)**				
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Display	0.0300	0.0220	0.0344	0.0262	0.0330
Facebook	0.1619	0.1422	0.2075	0.2472	0.2749
Paid Search	0.0631	0.0404	0.0824	0.0505	0.0702
Video	-	0.1074	0.0523	0.1491	0.0234
Outdoor	0.0250	-	-	-	0.0250

Media Channel	Total Elasticity (used in budget allocation)				
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Magazine	0.0646	-	0.0670	-	0.0685
TV	0.0600	-	0.0651	-	0.0609

Table 3-9 - Total Elasticity Values for Paid Media Channels

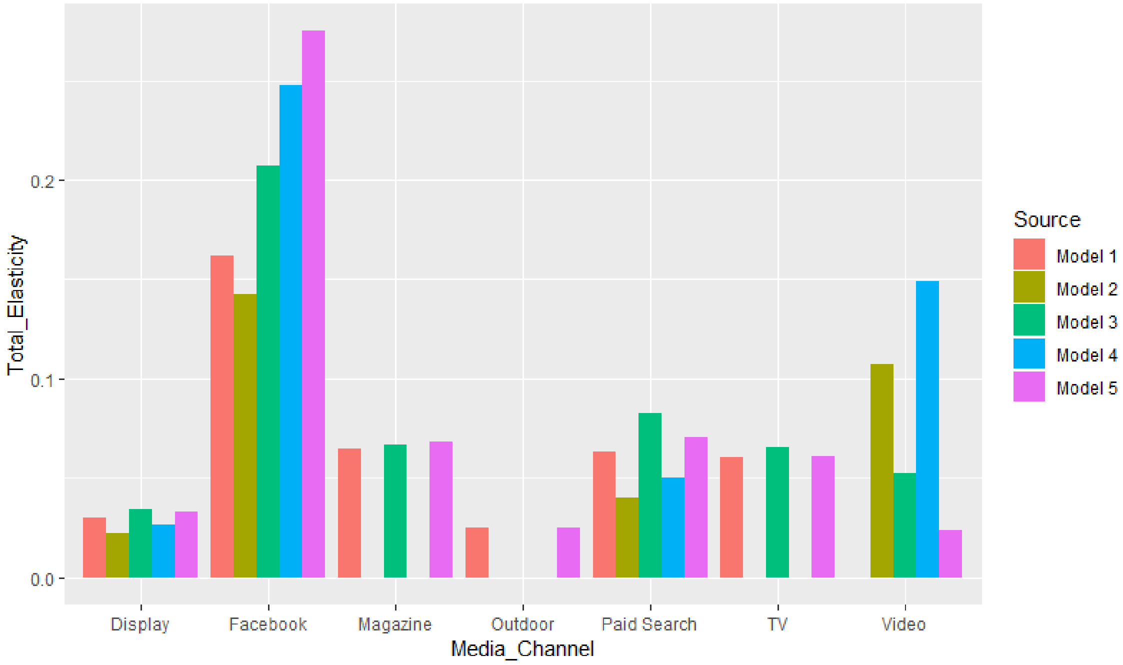


Figure 3-18 - Comparison of Total Elasticity Values for Paid Media Channels

Media Channel	Current Allocation	Recommended Allocation				
		Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Display	3.14%	7.43%	7.07%	6.77%	5.55%	5.94%
Facebook	36.74%	40.00%	45.57%	40.78%	52.25%	49.44%
Paid Search	13.54%	15.59%	12.95%	16.20%	10.68%	12.63%
Video	5.85%	0.00%	34.41%	10.28%	31.52%	4.22%
Outdoor	4.67%	6.19%	0.00%	0.00%	0.00%	4.50%
Magazine	14.72%	15.97%	0.00%	13.17%	0.00%	12.32%
TV	21.34%	14.82%	0.00%	12.80%	0.00%	10.95%

Table 3-10 - Current vs Recommended Budget Allocations

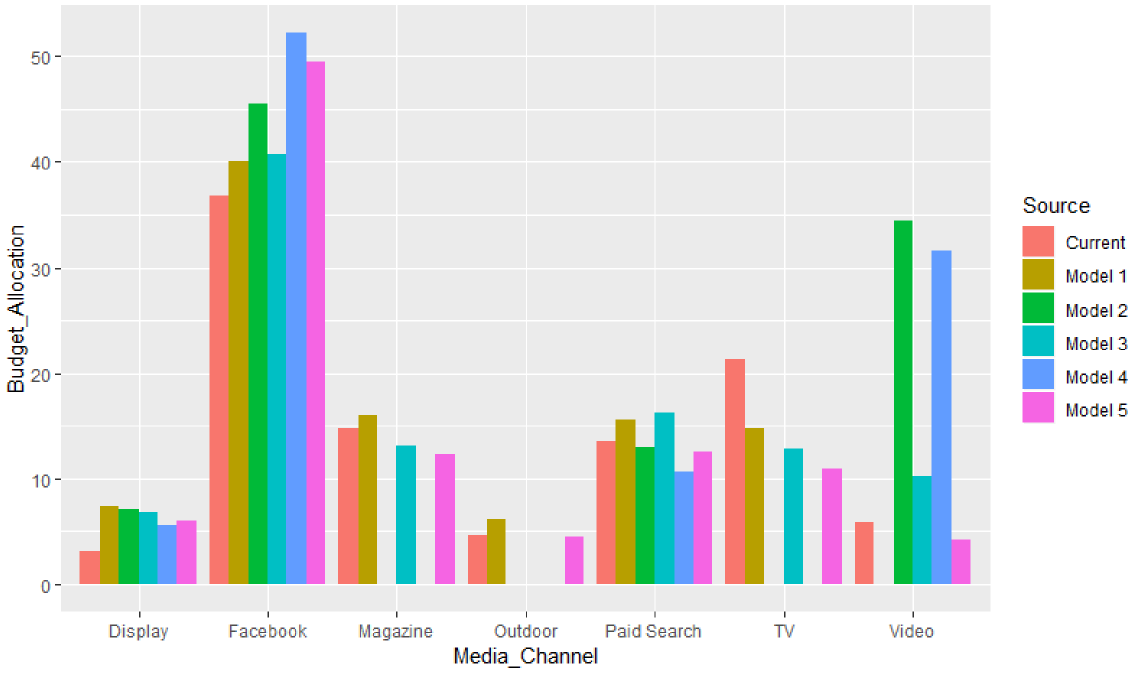


Figure 3-19 - Comparison of Recommended Budget Allocations

It is noteworthy that Table 3-9 and Figure 3-18 reveal seemingly conflicting results between the total elasticity and CPA calculations. For example, Facebook and TV advertising are more effective media channels according to the total elasticity calculations, yet have higher CPAs in Table 3-8. This is because ROI and CPA values are historical measures of performance and are influenced by the amount spent in the media channel in the period they are calculated for. Thus, having a high CPA does not necessarily mean that a media channel is ineffective, rather it means that the budget allocation in that period was probably too high relative to other media channels. An elasticity, on the other hand, represents a media channel's potential effectiveness and is not influenced by the amount spent in the channel. This is why elasticities are used in budget allocation calculations, instead of ROI or CPA values.

As can be seen in Table 3-10 and Figure 3-19, all models suggest that Facebook advertising is the most effective paid media channel. All models recommend concentrating media investment in this channel, typically suggesting an increase in current expenditure of 10-15%. While this results in a budget allocation of 40-50% towards Facebook advertising, this media channel does include both messenger and news feed ads. The other clear change recommended by the models which detect a statistically significant effect for TV advertising is reducing its budget allocation by 6-10%. Smaller budget changes are recommended to display, outdoor and magazine advertising when statistically significant effects are uncovered by the models. Paid search consistently has a recommended allocation of 10-15% across all models, meaning that its current allocation of 13-14% is unlikely to need much adjustment.

Despite these similarities, there are clear differences between the budget allocation recommendations. The recommended budget allocation for video expenditure is conflicting, with the models built using total enquiries suggesting a much larger allocation than the ones built using online and offline enquiries. Additionally, the total enquiries models allocate no budget towards offline media channels as they do not find any statistically significant effects for these media channels. Model 5 is the only model to recommend a budget allocation that involves all paid media channels. It suggests concentrating media investment in Facebook advertising, with smaller stakes in TV, display, video, outdoor and paid search advertising. From discussions with the media agency, the budget allocation of Model 5 is the one closest to their expectations. Despite this, it is difficult to verify its validity without conducting a field experiment. To provide insight into how the recommended budget allocation might perform in practice, Model 5 was fitted to the first 380 days in the analysis and subsequently used to predict total enquiries for the last 14 days based on the actual and recommended media budget allocations. The results of this analysis can be seen in Figure 3-20, which shows that following the recommended budget allocation would have improved total enquiries by 2.16% (3830 enquiries as opposed to 3749). While this may not seem like much, even a small percentage difference can result in large financial gains. For example, if 10% of enquiries result in a sale, following the recommended budget allocation would result in approximately 8 additional sales. If each luxury travel holiday sale is worth approximately $10,000, the company would gain $80,000 in additional revenue over the course of the two-week period. If this scenario continued over the course of one year, this would result in $2.08m in additional revenue. Thus, using advertising media effectiveness models to inform budget allocation decisions can result in substantial financial gains (Wiesel et al. 2011).

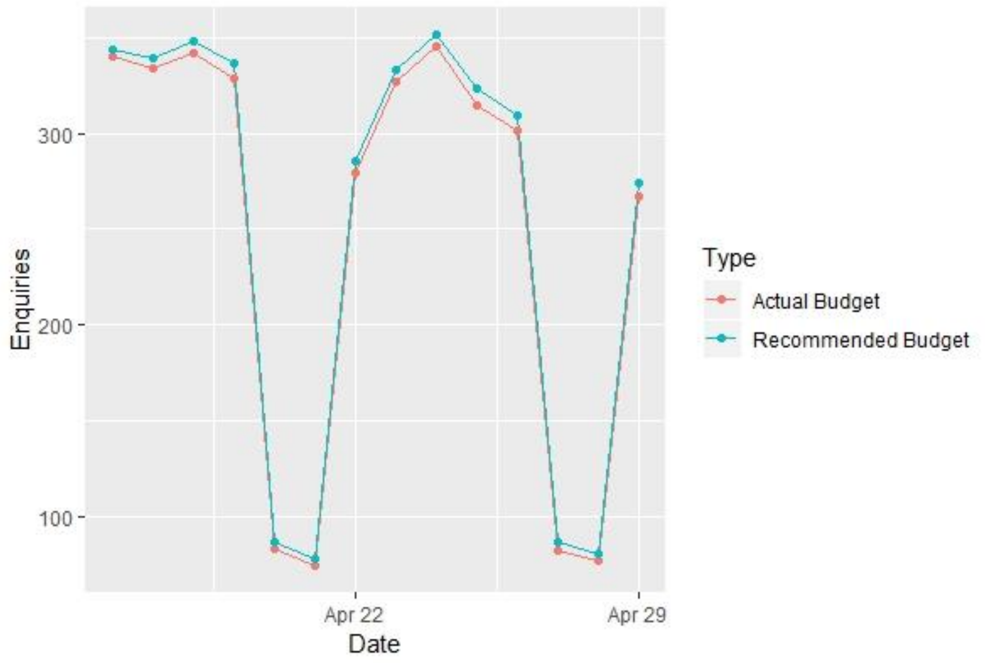


Figure 3-20 - Predicted Total Enquiries of Actual vs Recommended Media Budget Allocations

Overall, budget allocations vary considerably between models depending upon whether aggregate or disaggregate enquiries are used. While models built using total enquiries do not detect the contributions of offline media channels, it is likely that these channels should be part of the media channel mix. This is because the luxury travel brand's customers are often older, wealthier people who tend to engage with offline media channels, such as TV or magazines. Given the contributions of offline media channels are detected for models using disaggregated enquiries data, it is likely that models using aggregate marketing performance metrics lose some information through aggregation (Hyndman and Athanasopoulos 2018). For models with and without consumer activity metrics, the budget allocation differences tend to be not so substantial. However, when indirect effects are accounted for, there is a marked increase in the allocation towards paid Facebook advertising as its influence on owned Facebook advertising is included the total elasticity calculations. Therefore, since the indirect effects of media channels occur through consumer activity metrics, it is important to include these variables in the modelling process.

3.6 Research & Managerial Implications

While previous research has shown that consumer activity metrics can improve sales response models, it is not clear how including these metrics influences practical managerial calculations, such as the ROI or CPA of a media channel. The effects of using aggregated or disaggregated marketing performance metrics in the modelling process have also yet to be explored. This research addresses these gaps in the literature by comparing ROI, CPA and budget allocation calculations derived from persistence models built with and without these features. This chapter builds on the work of Srinivasan et al. (2016) by expanding their research to include a company's full suite of media channels, including owned and paid Facebook advertising, and using data collected at a higher frequency. These differences, along with the use of enquiries as the marketing performance metric and a high involvement product, provide insight into the extent to which the findings of Srinivasan et al. (2016) generalize.

This research finds that paid media channels can directly influence both marketing performance metrics and consumer activity metrics, which in turn can directly influence marketing performance metrics. This highlights that while paid media channels may not always directly influence marketing performance metrics, they may still influence variables that have their own direct relationship with them. For example, video advertising may drive consumers to a company's website, which may then incite them to make an enquiry or purchase. Thus, media channels without a direct effect on marketing performance metrics may still have a role to play in a company's media channel mix via driving consumers to other parts of the marketing funnel. In practice, marketing managers should monitor consumer activity metrics for online and offline media campaigns, as well as marketing performance metrics (Srinivasan et al. 2016). Including consumer activity metrics when modelling advertising media effectiveness provides a more complete picture of marketing effects, which is helpful for marketing managers in reporting and decision-making. This confirms the findings of Srinivasan et al. (2016) in an expanded context with more media channels, a high involvement product and daily data.

The results also shed light on the specific effects of media channels on marketing performance metrics. First, the effects of online media channels tend to last for longer than offline media channels. While the effects of online (offline) media channels tend to be larger on online (offline) enquiries, online (offline) media channels do still influence offline (online) enquiries. The presence of cross-channel effects is an affirmation of previous research (Dinner et al. 2014; Wiesel et al. 2011; Srinivasan et al. 2016). Additionally, this research reinforces the findings of Dinner et al. (2014) around paid search advertising, with increasing paid search expenditure having a negative effect on the paid search engagement rate. At a strategic level, marketing managers should be careful not to allocate too much of the marketing budget to paid search advertising. At a tactical level, media buyers should focus their attention on bidding for relevant and specific keywords, rather than a broad set that potentially generates less engagement.

This research is the first to include paid and owned Facebook advertising in a persistence model. We find that increasing the engagement on the travel brand's Facebook page posts has a positive effect on website page views and online enquiries. Additionally, we find that increasing the travel brand's paid Facebook advertising improves its Facebook page posts engagement. Thus, owned media channels can influence other owned media channels, as well as marketing performance metrics. The connection between paid and owned Facebook advertising suggests that there are benefits from coordinating the management of paid and owned media channels. Often this is not done in practice, with the management of paid media channels being given to media agencies, while owned media channels are managed internally. The size and timing of the effects of paid and owned media channels on marketing performance metrics also differ, with the initial effects of owned media channels tending to be smaller, but lasting for longer, than those of paid media channels.

Additionally, we find that both metrics for paid search advertising, expenditure and engagement rate, have positive effects on offline enquiries and website page views. Interestingly, the effects of increasing paid search expenditure were larger in the short-term, but the effects of increasing the paid search engagement rate lasted for longer and were larger overall. This highlights that simply increasing advertising expenditure is not necessarily the best way to improve a media channel's effectiveness. It also shows that leveraging different features of a media channel can have different effects on marketing performance metrics. More broadly, this finding reveals the importance of considering both a firm activity metric and a consumer activity metric for customer-initiated paid media channels when modelling advertising effectiveness. Including these metrics separates the effects of volume changes, such as an increase in expenditure, from more subtle changes, such as additional consumer engagement due to an increase in advertisement quality. This is particularly important for customer-initiated paid media channels as more consumers may choose to engage with them after seeing advertising in other media channels, such as TV, rather than just being exposed to the media channel due to extra expenditure in it.

The final major finding is that elasticity, ROI, CPA and budget allocation calculations for paid media channels vary depending upon whether marketing performance metrics are disaggregated, consumer activity metrics are included in the underlying model and indirect effects are accounted for. Disaggregating marketing performance metrics, including consumer activity metrics and accounting for indirect effects reveal a more complete picture of the marketing landscape, particularly for media channels encountered in the early stages of the marketing funnel. More specifically, breaking total enquiries into online and offline components provides insight into which media channels are influencing which enquiry channel. It also helps detect relationships between media channels and marketing performance metrics that do not appear when the enquiries data are aggregated. This is an example of hierarchical aggregation losing some of the information contained in the underlying data (Hyndman and Athanasopoulos 2018). Including consumer activity metrics when modelling the effectiveness of media channels enables the indirect relationships between paid media channels and marketing performance metrics to be captured. These relationships form an important part of a company's overall marketing landscape. Excluding them can reduce the accuracy of ROI, CPA and budget allocation calculations. In particular, companies should seek to include consumer activity metrics that capture their owned and earned media channels, as well as consumer engagement for customer-initiated paid media channels.

From a managerial perspective, modelling the direct and indirect effects of media channels is helpful as it reveals the different effects advertising expenditure is having, the relative size of those effects and how they develop over time. Visualising these relationships in network diagrams, as well as translating advertising elasticities into metrics more commonly used in industry, such as ROIs and CPAs, can assist marketing managers in interpreting and communicating these results to a non-technical audience. The data-driven approach presented in this research can increase transparency and visibility over the marketing budget, as well as provide insight into the optimal budget allocation. The methodology is generalisable to companies of varying sizes and media channel mixes, and can accommodate different types of marketing performance metrics, such as sales and enquiries.

3.7 Future Research

The limitations of this research present opportunities for future research. First, the findings have not been verified in a real-life application. Implementing the budget allocation recommendations in a field experiment would increase the robustness of the results and provide insight into their economic effects. For example, reallocating the marketing budget of the company analysed by Wiesel et al. (2011) in a field experiment resulted in a substantial increase in baseline profit. Field experiments could also provide opportunities to develop and empirically examine more refined marketing campaign strategies based on the direct and indirect effect calculations. For example, one could investigate the amount and timing of marketing expenditure required to achieve a sustained shock to enquiries.

Second, we did not have access to a financial performance metric, such as sales. Optimizing budget allocation based on a financial performance metric should result in greater revenue for the business, whereas increased enquiries may not necessarily convert into sales. Adding sales information to our model and comparing the effects of media channels on it with their effects on enquiries would be interesting, particularly if certain media channels were driving sales more than enquiries. Such insights could help companies better focus and optimize their advertising so that more enquiries convert into sales.

Finally, incorporating prior customer information (e.g. attitudinal metrics) into the analysis, would be an interesting area of future research. Marketers often use customer information, such as surveys and loyalty program data, to design marketing initiatives, before assessing their outcomes using marketing performance metrics. A Bayesian model, such as a Bayesian Vector Autoregression (BVAR), could replicate this process in statistical form by incorporating customer information into its prior distribution. This is alluded to by Pauwels et al. (2016b), who suggest that future research with BVAR models could examine the use of more complicated prior distributions and use MCMC methods to estimate the posterior distribution.

Chapter 4 Improving Demand Forecasts with Advertising and Calendar-Based Variables

Demand forecasting is an important part of an organization's strategic and operational planning. Improving demand forecasts can help organizations run more efficiently and effectively, which can in turn increase their profitability (Fildes et al. 2019). This chapter investigates whether including advertising spend and calendar-based variables in time-series models can improve demand forecasts. We find that using relatively simple models, augmented with exogenous variables, can produce more accurate forecasts than more complex pure time-series models. This research also examines how the relative performance of modelling approaches to generating coherent forecasts of total demand changes over different forecast horizons and data sizes. The results show that when less data are available, an aggregate forecasting approach is preferable, while a bottom-up or ensemble approach becomes more effective as more data are collected. As a result of these findings, we develop an extension to time-series cross validation that reduces its sensitivity to the number of observations used in the initial training subset. This new technique, termed repeated time-series cross validation (RTSCV), provides a more comprehensive way to assess time-series model performance.

4.1 Introduction

Demand forecasting is a critical task for many businesses as it assists them with operational decision-making. Demand forecasts are key variables in managerial decisions, such as pricing, promotions, staffing and supply chain management, across a wide variety of industries, including retail, tourism and utilities. In the short-term, accurate forecasts can help to reduce the waste of perishable goods (Huber and Stuckenschmidt 2020) and lead to increased sales revenue and customer satisfaction (Fildes et al. 2019). In the longer-term, demand forecasts can be useful in developing infrastructure, such as roads and hotels, to support expected future tourism (Ghalehkhondabi et al. 2019). From a research perspective, the problem of demand forecasting is also often used to benchmark and test forecasting models, such as in the tourism forecasting competition described by Athanasopoulos et al. (2011).

Over time, there has been considerable development in demand forecasting methods. The survey papers of Ghalehkhondabi et al. (2019), Weron (2014) and Fildes et al. (2019) cover these in more detail for the tourism, electricity and retail industries respectively. Makridakis et al. (2020) also provide a broader non-systematic review of forecasting in social settings. Most approaches to demand forecasting can be classified into one of two categories, statistical or machine learning. Statistical approaches cover more traditional time-series models, such as ARIMA or ETS models, which are often based on a regression framework. Machine learning approaches refer to non-parametric and non-linear techniques that are more data-driven than their statistical counterparts (Ghalehkhondabi et al. 2019). Examples include support vector machines and artificial neural networks. There is also a strong interest in ensemble approaches, which seek to combine the forecasts of multiple models together to generate a final forecast (Hyndman and Athanasopoulos 2018). This could involve simply taking the average of predictions from base-forecasting models, or refer to more complicated meta-learning approaches, such as those of Ma and Fildes (2021) and Montero-Manso et al. (2020). While these more complicated approaches can produce more accurate results than simpler models, they are often more difficult to understand and implement in practice.

All forecasting methods are only as reliable as the data used to build them. Since demand forecasting is inherently a time-series problem, most demand forecasting methods focus on using past values of demand to predict future ones. Including exogenous predictors, as well as the history of the time series being forecasted, is a seemingly logical strategy for further improving model accuracy through the incorporation of additional relevant information into the model. In some situations, exogenous variables have been shown to increase model accuracy. For example, Taieb and Hyndman (2014) find that including current and past weather temperatures can improve electricity demand forecasts, with the final model placing fifth in the Load Forecasting track of the Kaggle Global Energy Forecasting Competition 2012. This makes sense as hot and cold weather temperatures often lead to the use of air conditioning for cooling or heating, which increases electricity demand. However, in other cases, the inclusion of exogenous variables is less helpful. Athanasopoulos et al. (2011) compare the performance of time-series models with and without exogenous variables for tourism data, finding that the pure time-series approaches produce more accurate forecasts than the ones with exogenous variables.

Makridakis et al. (2020) outline two key conditions for including exogenous predictors in time-series models. First, the predictors need to be forecasted themselves or known in advance of the time series being forecasted. If accurate forecasts of the predictors cannot be produced, it will be more difficult for them to improve a pure time-series model. Second, there must be a strong relationship between the exogenous variables and the forecast variable that is likely to continue into the future (Makridakis et al. 2020). If the relationship between the predictors and forecast variable is weak, it is unlikely to add enough value to the model to be useful. Additionally, if the relationship breaks down over time, including exogenous variables in the model is likely to reduce the accuracy of its predictions. Time-varying parameter models can help address this issue by allowing the exogenous variables' parameters to change over time. Indeed, the time-varying parameter models with exogenous variables performed better than the fixed parameters models in Athanasopoulos et al. (2011), but still not as well as the pure time-series models. Finally, the resources required for collecting, processing and using exogenous variables in demand forecasting models is another issue that needs to be considered when practically implementing a model.

Overall, there are questions surrounding the usefulness of exogenous variables in time-series forecasting. More specifically, Makridakis et al. (2020) call for more empirical research into understanding the value of collecting exogenous data and determining the domains in which their inclusion in forecasting models are likely to practically improve performance. This chapter addresses this call through the following research question, which is derived from RQ3:

1. To what extent can advertising spend and calendar-based indicator variables improve demand forecasts across different forecast horizons?

Once calculated, demand forecasts can often be organized into a hierarchical structure based on a characteristic of interest, such as geographic location or product type. The collection of time series that forms the hierarchy is called a hierarchical time series (Hyndman and Athanasopoulos 2018). It is often necessary or desirable to produce forecasts for different levels within the hierarchy. For example, if a retailer desires to predict weekly sales over an entire country, it might also want to make forecasts for sales in individual states, as well as the actual retail stores within them. When producing forecasts at different levels in a hierarchy, the forecasts are usually required to be coherent, or add up in a way that's consistent with the underlying data (Hyndman and Athanasopoulos 2018). For example, when individual store sales forecasts in a particular state are added up, they should equal the state's predicted sales total. Furthermore, when the state totals are added up, they should equal the country's predicted sales total. Ensuring that demand forecasts within a hierarchy are coherent is important so that there is consistency in planning and execution across a business.

One method for generating coherent forecasts is the bottom-up approach, which involves forecasting each bottom-level time series and adding up the results to produce forecasts for the higher levels. Since forecasting is done at the most disaggregated level, no information is lost due to aggregation. However, bottom-level time series can be noisy, particularly when the data are highly disaggregated. Additionally, by forecasting each bottom-level series independently, the bottom-up approach ignores relationships between time series (Wickramasuriya et al. 2019). Another way to forecast hierarchical time series is the top-down approach, which involves forecasting the top-level time series and breaking these down into forecasts for the lower-level time series using proportions. However, this approach loses information due to aggregation and does not produce unbiased coherent forecasts (Hyndman et al. 2011). More advanced approaches focus on forecasting all levels of the hierarchical time series and then using a regression model to optimally combine and reconcile the forecasts so that they are coherent (Hyndman et al. 2011; Wickramasuriya et al. 2019).

While there have been several advances in approaches to generating coherent forecasts, there is an absence of research into how approaches perform across different forecast horizons, as well as how this might change as more data are collected and used in the model estimation process. For example, when less data are available, it may be better to forecast overall demand directly, rather than adopting the bottom-up approach, which may require more data to be effective. Additionally, more complicated approaches may perform better on smaller forecast horizons, while simpler ones may be more accurate for larger forecast horizons due to more variance in the forecasting process. This chapter addresses these gaps in the literature through the second and third research questions, which are derived from RQ4:

2. How does the performance of bottom-up and aggregate approaches to demand forecasting change across forecast horizons?

3. How does the performance of bottom-up and aggregate approaches to demand forecasting change as the amount of data used in model estimation grows?

Overall, this chapter makes three main contributions to the extant demand forecasting literature. First, we address the call of Makridakis et al. (2020) for more empirical research into how exogenous variables can practically improve demand forecasts by examining how the inclusion of advertising spend and calendar-based variables affects demand forecasts across different forecast horizons. We compare the performance of a range of time-series models and provide insight into whether simpler models with advertising spend and calendar-based exogenous variables can outperform more complex pure time-series models. Since more complex time-series models are often harder to understand and implement in practice, it is useful to investigate if comparative results can be achieved with simpler models augmented by exogenous variables. Analyzing different forecast horizons is important as relative model performance may change over time due to the additional uncertainty introduced in the forecasting process. Thus, if a company desires a demand forecast thirty days into the future, they may prefer to use a different model to the one they might use if they were generating forecasts one day in advance. Calendar-based exogenous variables, such as weekend dummy variables, are strong candidates for model inclusion as they are known in advance and are efficient to collect. The relationship between advertising and demand is well-documented in marketing literature, with advertising variables being shown to help improve sales forecasts (Hanssens et al. 2014). Since advertising budgets are often set in advance, reliable forecasts of advertising expenditure can likely be used as exogenous variables in demand forecasting.

Our second contribution is to explore the performance of bottom-up and aggregate approaches to forecasting overall demand as the forecast horizon and amount of data used in model estimation changes. This is important because the most accurate approach to forecasting demand may change as more data are collected and the forecast horizon increases. Thus, companies may be able to produce more accurate demand forecasts by switching their forecasting approach as the conditions change, rather than adopting a one-size fits all approach. When there is less data available, it is expected that an aggregate forecasting approach will perform better than a bottom-up approach due to greater uncertainty in the forecasting process. However, as more data are collected, this approach will become less effective as information-loss due to data aggregation grows.

Additionally, implementations of bottom-up forecasting often use the same forecasting model for each bottom-level time series. By allowing the chosen model and exogenous variables to vary for each bottom-level time series, we investigate whether a customized approach to model fitting can produce more accurate forecasts.

Third, this chapter makes a methodological contribution to the literature by developing an extension to time-series cross validation that reduces the effect of the number of data points used in the initial training subset on overall model performance. Since model performance can change depending upon the amount of data used to fit the model, choosing a small data size for the initial model in time-series cross validation can result in models that perform better on larger datasets appearing less effective than they actually are given the data available for training. Our extension to time-series cross validation addresses this issue by giving greater weight to models built on more training data, which is more realistic since the final model will likely be fitted using all available data. However, the extension can also consider the errors from a large number of train and test subsets, which helps to obtain a robust estimate of overall model performance. We term this new technique repeated time-series cross validation (RTSCV) as it is analogous to repeated k-fold cross validation, an existing model evaluation technique.

The rest of this chapter is structured as follows. First, the data and methodology used to address the research questions are presented. The methodology was broken into two stages, an individual time-series modelling stage and a hierarchical modelling stage, with each stage addressing separate research questions. The results for each methodological stage are then discussed, along with the proposed new model validation technique for time-series models, repeated time-series cross validation. Finally, key findings and conclusions are provided, along with future research directions.

4.2 Data

The data for this research comes from an Australian travel brand, which conducts online and offline advertising to assist in the sale of travel packages. Online advertising refers to Facebook, paid search, display and video advertising, while offline advertising refers to TV, magazine and outdoor advertising. Since the travel brand sells high involvement products, consumers tend to have discussions with a sales consultant before making a purchase. Thus, advertising is primarily evaluated by the number of consumer enquiries to the travel brand. An enquiry can be made online through the completion of a form on the travel brand's website or offline via a phone call to the company's call centre. Enquiry and advertising data were collected daily between 01 April 2018 and 29 April 2019, resulting in 394 observations. Descriptive statistics for the data are shown in Table 4-1, with total enquiries and total advertising expenditure referring to the sum of online and offline enquiries and advertising expenditure respectively.

Variable	**Min**	**Max**	**25% Quartile**	**75% Quartile**	**Mean**	**Median**	**SD**
Online Enquiries	7.00	235.00	37.00	73.75	58.62	51.00	33.02
Offline Enquiries	0.00	393.00	66.25	258.75	153.67	136.00	105.42
Total Enquiries	11.00	563.00	125.00	307.75	212.30	197.50	117.66
Online Advertising Expenditure	379.48	17857.18	4261.63	7809.68	6231.82	5956.61	2606.35
Offline Advertising Expenditure	0.00	52145.96	82.57	5802.35	4283.01	2147.50	6320.68
Total Advertising Expenditure	379.48	60479.14	6077.63	12535.32	10514.83	8526.69	7129.69

Table 4-1 - Descriptive Statistics for Daily Enquiries & Advertising (AUD) Data

This chapter focuses on the prediction of daily enquiry numbers, which are shown in Figure 4-1.

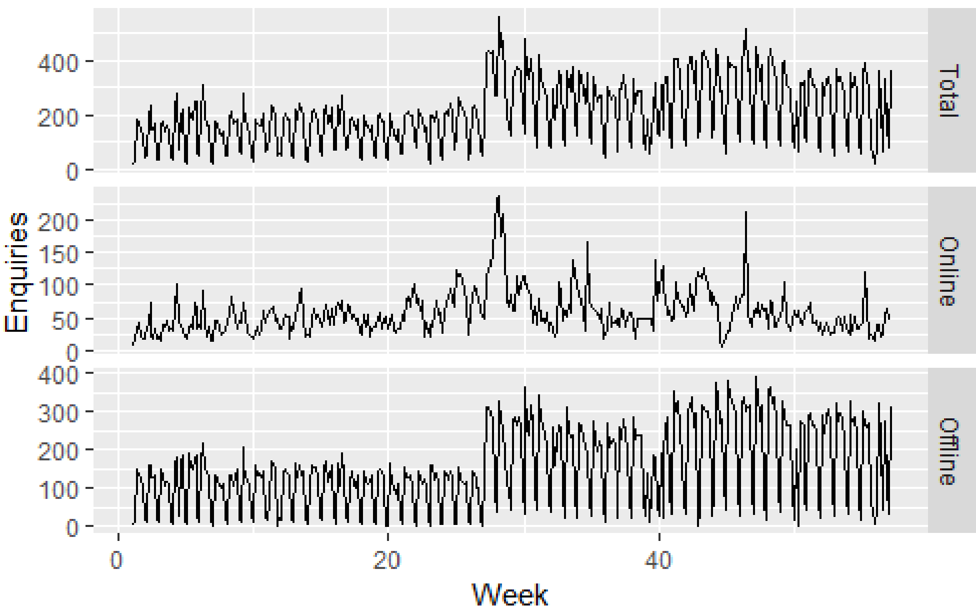


Figure 4-1 - Online, Offline & Total Daily Enquiries by Week

As can be seen in Table 4-1 and Figure 4-1, the travel brand receives more offline than online enquiries, with an average of 153.67 offline enquiries per day compared to 58.62 online enquiries. A marked increase in daily enquiries can be seen in Figure 4-1 following a rebrand of the company that went live on 01 October 2018. Figure 4-1 also shows clear weekly seasonality, particularly in offline and total enquiries. This is due to the call centre having reduced hours on weekends and public holidays. Further evidence of this phenomenon is depicted in Figure 4-2, which subsets enquiries by the different days of the week.

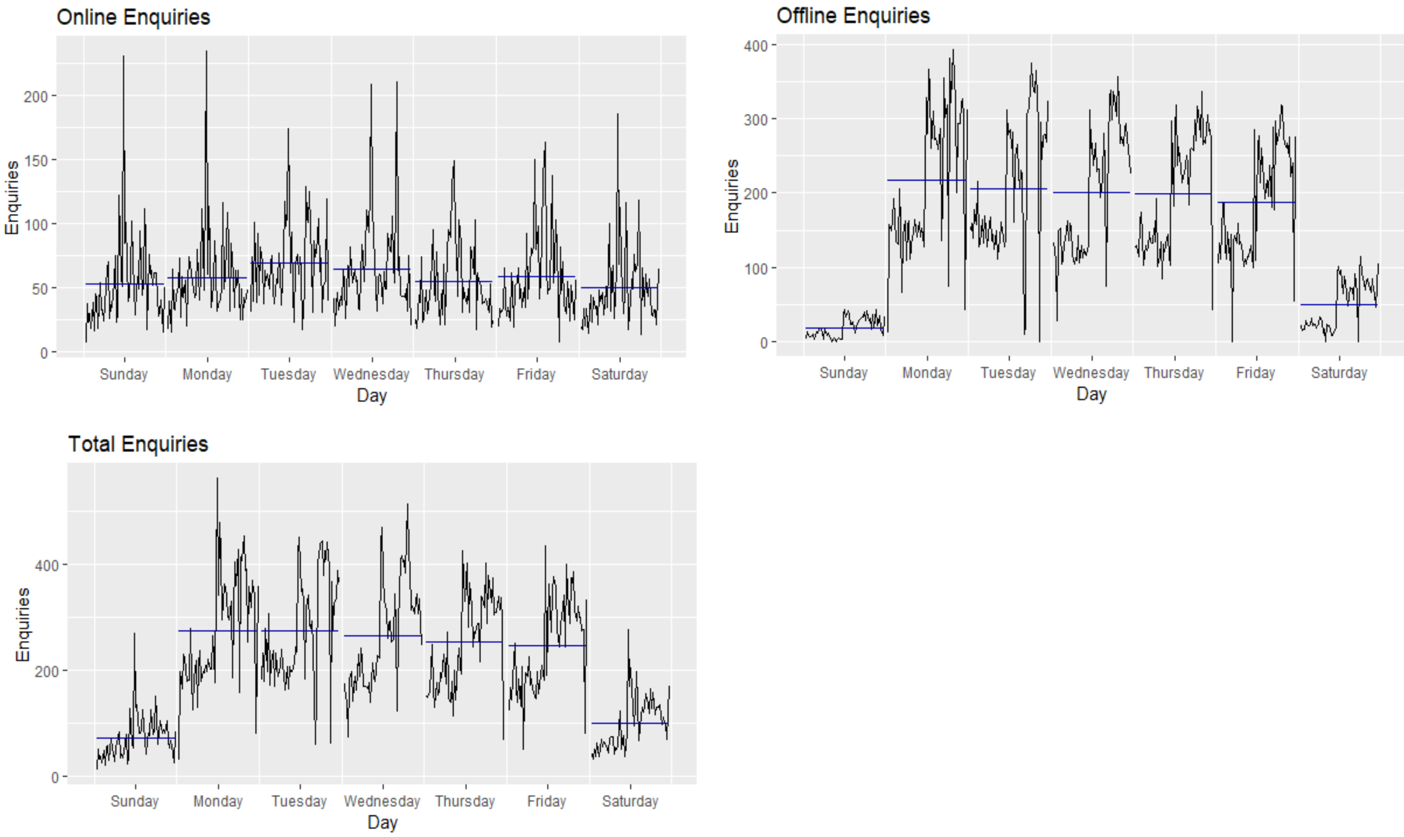


Figure 4-2 – Daily Enquiries by Day of the Week

These calendar-based effects led to the creation of three exogenous variables that were considered in the modelling of enquiries. First, an after rebrand dummy variable that captures the company rebrand from 01 October 2018 onwards. Second, a weekend dummy variable to capture the reduced enquiry numbers on Saturdays and Sundays. Third, a day off dummy variable that accounts for public holidays in addition to weekends. Public holidays are defined as Christmas Eve, Christmas Day, Boxing Day, New Year's Eve, New Year's Day, Easter (Friday-Monday), Anzac Day, Australia Day and the Queen's Birthday[3]. These holidays were chosen as they affect all of Australia. While Christmas Eve and New Year’s Eve are not strictly holidays, many Australians have these days off or work reduced hours, meaning that enquiry numbers may behave differently on these days. The breakdown of the calendar-based variables are shown in Table 4-2.

Dummy Variable Name	**Condition**	**Number of True Values**	**Number of False Values**
After Rebrand	On or after 01 October 2018	211	183
Weekend	Saturday or Sunday	113	281
Day Off	Saturday, Sunday or Public Holiday	126	268

Table 4-2 – Calendar-Based Exogenous Variables Tested in Demand Forecasting Models

In addition to calendar-based effects, the travel brand’s daily online and offline or overall advertising expenditure totals were considered as exogenous variables in the modelling process. Advertising expenditure totals are used to produce more parsimonious models and because it is easier to forecast expenditure totals accurately. Online and offline advertising expenditure totals are considered as an alternative to total advertising expenditure since online and offline enquiries may respond differently to advertising expenditure in these channels. Figure 4-3 confirms this, suggesting that there is a relatively strong positive relationship between online advertising expenditure and online enquiries.

[3] Western Australia’s Queen’s Birthday holiday is not included as it occurs on a different day to the rest of Australia.

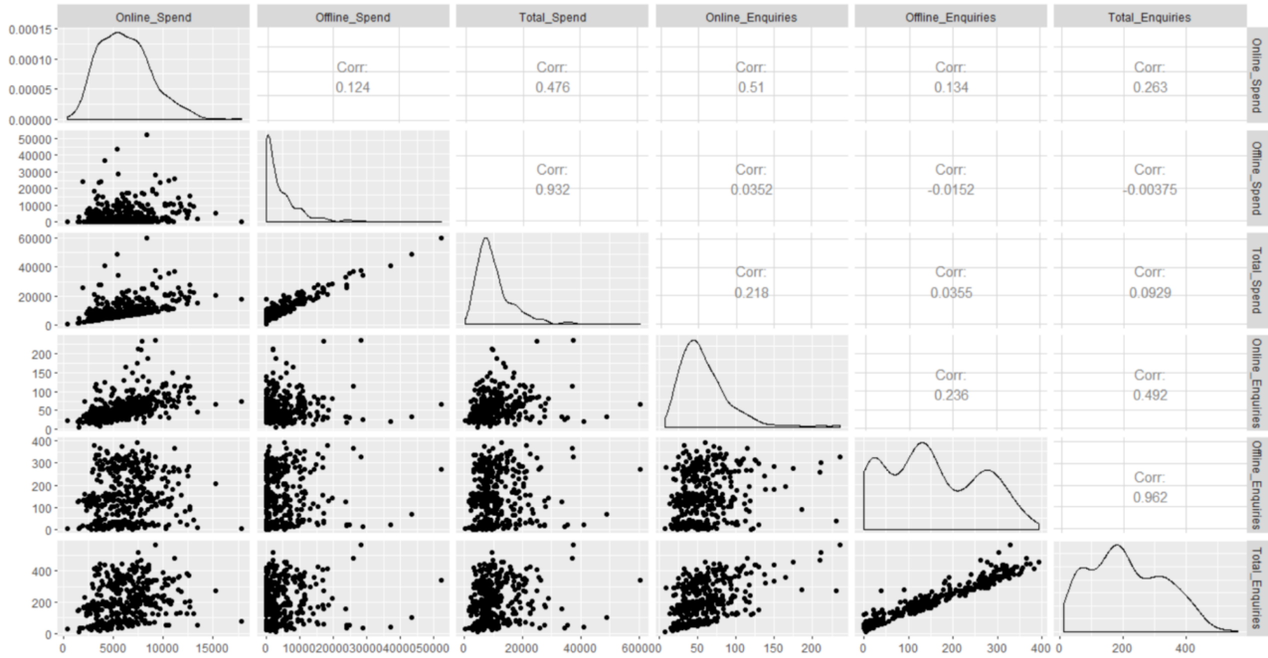


Figure 4-3 - Correlations between Daily Advertising Expenditure & Enquiry Variables

The advertising expenditure variables were operationalized using a $\ln(X_t + 1)$ transformation to capture the diminishing returns and multiplicative aspects of the relationship between advertising and demand. Adding one before the log transformation was necessary as some days have zero offline advertising spend.

4.3 Methodology

This research explores the effects of exogenous variables and data size on demand forecasts for the travel brand's daily total enquiries. The analysis consisted of two main methodological stages: individual time-series modelling and hierarchical modelling. In the first stage, candidate models were constructed for online enquiries, offline enquiries and total enquiries, with the best-performing model being chosen for each time series. The hierarchical modelling stage involved taking these models and using them to build forecasts of total enquiries. First, bottom-up forecasts of total enquiries were created by adding the forecasts of the best-performing models for online and offline enquiries. Second, ensemble forecasts for total enquiries were constructed by averaging the bottom-up and total enquiries model forecasts. Finally, the bottom-up, ensemble and aggregate forecasts for total enquiries were compared to reveal the best model for forecasting demand. The modelling process is depicted in Figure 4-4.

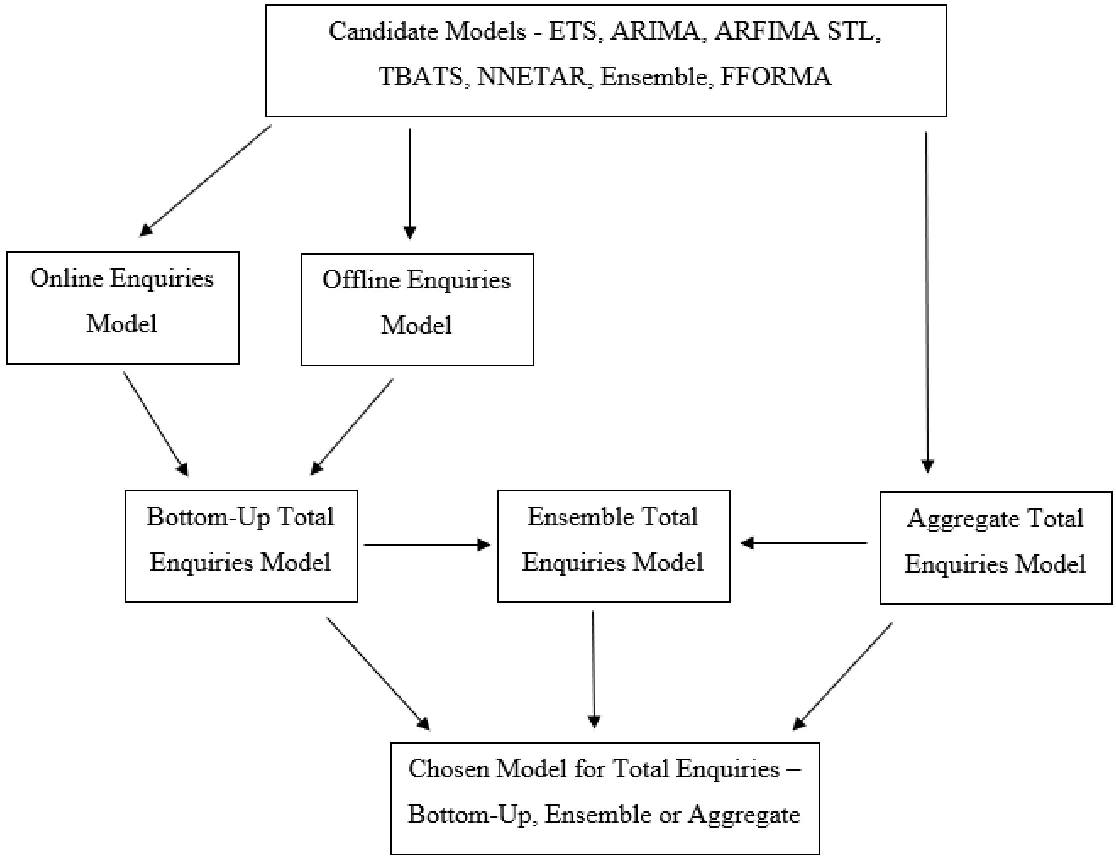


Figure 4-4 – Demand Forecasting Modelling Process

In the individual time-series modelling stage, eight candidate models were considered for the online, offline and total enquiries time series. These models were chosen as they cover a diverse range of approaches to time-series forecasting and have been shown to be successful in other applications, as demonstrated by many of them being discussed in the forecasting book by Hyndman and Athanasopoulos (2018). Details about these models are provided in Table 4-3.

Model	Description	Reference
ETS	Error, Trend and Seasonality; Forecasts are weighted averages of past observations, with the weights decreasing exponentially as observations become older; Can handle time series with trend and (or) seasonal components; Errors can be additive or multiplicative depending on whether seasonal variations are constant or proportional to the time series	Winters (1960), Hyndman and Athanasopoulos (2018)

Model	Description	Reference
ARIMA	Autoregressive Integrated Moving Average; ARIMA and ETS models are two most popular methods for time-series forecasting; Focuses on using autocorrelations in data to predict future values; Forecasts are a linear combination of previous values (AR terms) or forecast errors (MA terms); Can incorporate exogenous predictors; Models selected using the algorithm presented by Hyndman and Khandakar (2008)	Box et al. (2015), Hyndman and Khandakar (2008)
ARFIMA	Autoregressive Fractionally Integrated Moving Average; ARIMA model but with fractional differencing; Useful when the autoregressive relationship between time-series values decays slower than usual (data has long-memory); Differencing parameter is restricted to between 0 and 0.5 to ensure a stationary model is produced; ARIMA part of model selected using the Hyndman and Khandakar (2008) algorithm and parameters are estimated using the Haslett and Raftery (1989) algorithm	Haslett and Raftery (1989)
STL	Seasonal and Trend Decomposition using LOESS; Decomposes time series into seasonal, trend and remainder components using local regression (LOESS); ETS model is fitted to remainder component	Cleveland et al. (1990), Hyndman and Athanasopoulos (2018)
TBATS	Trigonometric Box-Cox transform, ARMA errors, Trend, and Seasonal components; Allows seasonality to change over time; Models data using Fourier terms with an exponential smoothing state space model and a Box-Cox transformation	De Livera et al. (2011)
NNETAR	Neural Network Autoregression; Feed-forward neural network with one hidden layer; Past values of time series are used as predictors, along with any exogenous variables provided; Fits 20 neural networks with different random starting weights and averages their results to produce forecasts	Hyndman and Athanasopoulos (2018)
Ensemble	Forecasts are average of forecasts of ETS, ARIMA, ARFIMA, STL, TBATS and NNETAR models; Ensembles often produce more accurate forecasts than those of individual models; Combining forecasts helps to capture the strengths of different models while reducing the influence of errors in individual models	Clemen (1989)
FFORMA	Feature-Based Forecast Model Averaging; Weighted ensemble approach to prediction; Automated method for obtaining weighted forecast combinations using features of the time series being predicted; Created versions of model with and without a Box-Cox transformation to the data before model fitting; Have reported version without transformation as it produced more accurate results; Used default settings as shown to perform well in paper behind the model; Objective parameter was set to model averaging as purpose of analysis was to produce forecasts; Trained model on 2973 datasets from the M3 forecasting competition[4] before applying to enquiries data	Montero-Manso et al. (2020)

Table 4-3 - Summary of Candidate Models for Demand Forecasting

[4] The datasets from the M3 forecasting competition can be found in the *Mcomp* R package. See https://cran.r-project.org/web/packages/Mcomp/index.html for more details.

Each of the models in Table 4-3 were fitted to the online, offline and total enquiries time series. For the ARIMA and NNETAR models, which can support exogenous predictors, best subsets regression was used to investigate how including different combinations of the advertising spend (total, online and offline) and calendar-based (weekend, day off and after rebrand) variables affected model performance. A selection of these exogenous predictor variations are detailed in Table 4-4. No lag adjustments or inclusions were made to the advertising expenditure variables, meaning that the enquiries for a certain day were predicted using that day's advertising expenditure. This is because the travel brand's advertising campaigns are planned thirty days in advance, meaning that daily advertising expenditure totals are known in advance. While there is likely an endogenous relationship between advertising expenditure and enquiries, no endogeneity corrections were made as the focus of this analysis is purely on prediction, not explanation. Papies et al. (2017) note that if the goal of a linear model is prediction, it is generally considered better to not correct for endogeneity as the correction will adjust the model parameters away from the best fitting model. Thus, from a model prediction perspective, the cure for endogeneity can be worse than the problem.

	Exogenous Variables - Online Enquiries	**Exogenous Variables - Offline & Total Enquiries**
Model Variation 1 (baseline - ARIMA or NNETAR)	None	None
Model Variation 2 (Total_Xreg)	Total Advertising Spend, Weekend and After Rebrand	Total Advertising Spend, Day Off and After Rebrand
Model Variation 3 (All_Xreg)	Online Advertising Spend, Offline Advertising Spend, Weekend and After Rebrand	Online Advertising Spend, Offline Advertising Spend, Day Off and After Rebrand
Model Variation 4 (No_Xreg)	Weekend and After Rebrand	Day Off and After Rebrand
Model Variation 5 (Extra_Xreg)	Online Advertising Spend, Offline Advertising Spend, Day Off and After Rebrand	Weekend and After Rebrand

Table 4-4 - ARIMA & NNETAR Model Exogenous Variable Combinations

The model variations in Table 4-4 provide insight into the effects of the exogenous variables upon performance without overcrowding the analysis. Different variations are shown for the online enquiries time series compared to the offline and total enquiries time series as different sets of exogenous variables performed better on each time series. More specifically, it was found that using the weekend dummy variable instead of the day off dummy variable produced more accurate results for the online enquiries model than the offline and total enquiries models. In other words, it is better to not account for public holidays when modelling online enquiries. However, it is important to include them when modelling offline or total enquiries. This is likely because online enquiries are made via the travel brand's website, which is still live on public holidays. Offline enquiries, on the other hand, are made via a phone call to a call centre, which has reduced hours on public holidays. Since offline enquiries make up a large portion of total enquiries, this effect is still felt when modelling total enquiries. The best-performing variations of the ARIMA and NNETAR models for each enquiries time series were subsequently compared to the other fitted models to find the best-performing model for each time series.

The performance of a statistical model is often assessed using k-fold cross validation, which involves randomly dividing a dataset into *k* folds (groups), building a model on *k-1* of the folds and evaluating the model's ability to predict the data with the remaining fold. The model building and evaluation process is repeated *k* times so that all folds are used as the test dataset once. After this, the model evaluation scores are typically averaged across all of the folds to produce an overall summary of the model's performance. Due to the serial correlation present in time-series data, the application of cross validation to time-series models is not always clear. Bergmeir et al. (2018) show that k-fold cross validation can be applied to time-series models which have uncorrelated errors and use previous values of the time series as predictors. However, the technique is difficult to apply to other time-series models, including ones with exogenous variables. An alternative often used by practitioners is out-of-sample evaluation, where a dataset is broken into two subsets, one used to train the model and the other used to test the model's performance. However, this can result in a model assessment that is only based on a small number of predictions on a particular section of the data. This can lead to misleading conclusions about the model's generalizability and performance, particularly in small datasets.

Another option is to use time-series cross validation, which extends the concept of out-of-sample evaluation by incorporating one of the key benefits of cross validation, the use of multiple train and test subsets (Hyndman and Athanasopoulos 2018). First, a model is fitted on the first *X* data points and used to predict *h* steps ahead of these observations, where *h* is the desired forecast horizon. A new model is then fitted on the first *X + 1* data points and used to predict the data point *h* steps ahead of these observations. This process is repeated until the end of the dataset, with model performance being calculated by averaging over the results of the test subsets. Figure 4-5 and Figure 4-6 from Hyndman and Athanasopoulos (2018) show example train (blue) and test (red) subsets for 1-step and 4-step ahead time-series cross validation.

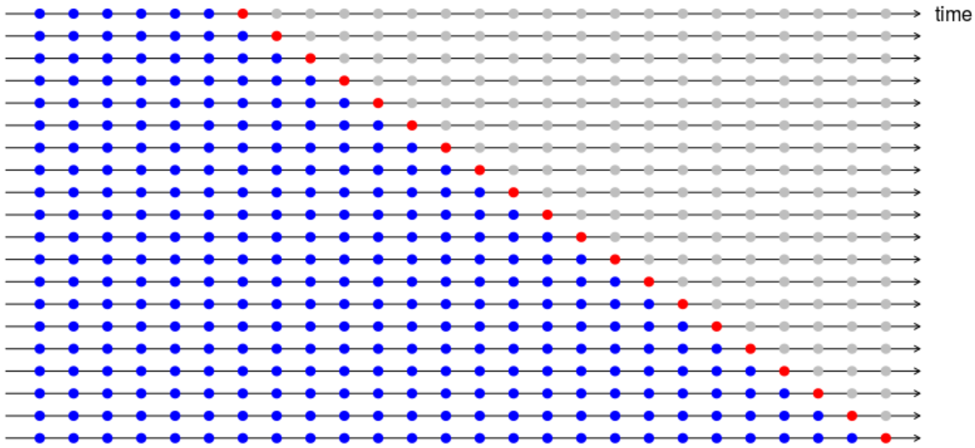


Figure 4-5 - Example Train/Test Subsets for 1-Step Time-Series Cross Validation

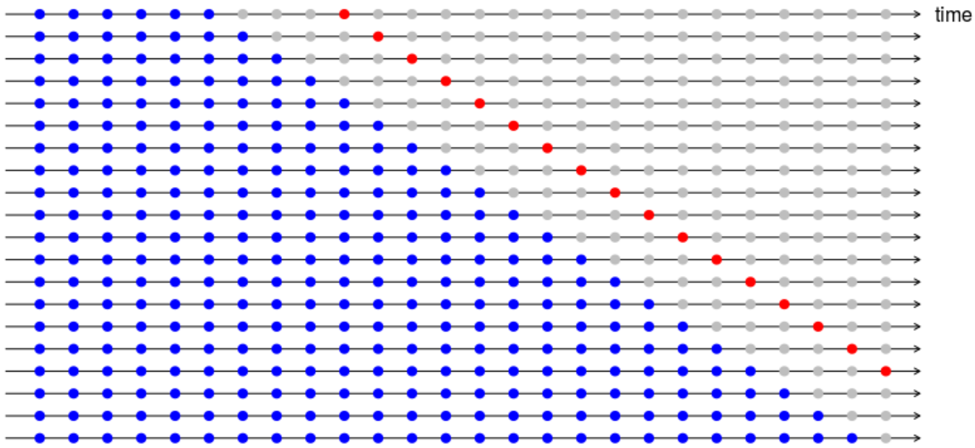


Figure 4-6 - Example Train/Test Subsets for 4-Step Time-Series Cross Validation

By using multiple train and test subsets, time-series cross validation provides more insight into a model's generalizability and performance. Another benefit of the procedure is that it replicates practical constraints by only training models with observations that occurred prior to those in the test sets. The refitting of models as the time series grows and more data becomes available is also analogous to what often occurs in practice. Time-series cross validation can be conducted on an expanding or rolling window basis. Expanding window refers to when the origin of the training sets is fixed and more data are included in each subset, while rolling window refers to when the same number of data points is included in each subset. The other key parameter that needs to be chosen is *X*, or the number of data points used to fit the initial model. Since unreliable forecasts are produced when a model is trained on a small dataset, the earliest observations should not be included in the test subsets (Hyndman and Athanasopoulos 2018). However, if *X* is set too large, there will not be enough train and test subsets get an accurate assessment of model performance.

In this research, time-series cross validation was performed on an expanding window basis with training sets containing the first 250-393 data points of the 394 observation dataset. The expanding window approach was chosen to maximize the number of data points used in model estimation, thereby producing more accurate and robust results. A rolling window basis was also considered, with no material difference to the results. As a robustness check, values of *X* other than 250, namely 200, 280, 300 and 330, were also considered. This was done to see if model choices and performance were sensitive to the size of the first train subset. In the individual time-series modelling stage, the results were not materially affected by different values of *X*. Thus, only results for $X = 250$ are reported. Higher values of *X* were considered as the value of the after rebrand variable does not change until after the first 184 data points. Thus, some models are unable to be estimated unless they contain more than 184 data points. To investigate model performance over different forecast horizons, forecasts up to thirty periods (days) ahead in the future were produced for each train and test set. Thirty periods was chosen as the maximum forecast horizon to provide insight into expected enquiries over the next month and because advertising expenditure totals are finalized thirty days in advance. For some models built using larger numbers of data points, it was not possible to forecast the full thirty periods in the future as there were no actual values to compare with predictions. For example, for models built using more than 364 data points, there are no actual values to compare against $h = 30$ forecasts as there are only 394 observations in the total dataset. Thus, more errors were collected for smaller values of *h*, with the final model fitted on 393 observations only producing a prediction for $h = 1$.

The time-series cross validation produced errors for each model over the different train and test subsets for forecasts from 1-step ahead up to 30-steps ahead. Each model's performance was assessed by calculating the RMSE and MAE for each *h*-step forecast across all available test subsets. This produced summary measures of each model's ability to predict at each forecast horizon from $h = 1$ to $h = 30$. Both RMSE and MAE were calculated to see if the models differed in their abilities to predict the mean or median respectively. Individual RMSE values across all forecasts were also calculated for each test subset that contained thirty periods (i.e. models estimated using up to 364 data points). These values summarize a model's overall ability to predict the next thirty days of enquiries and provide insight into how this might change as more data are collected.

The modelling process was operationalized in the R programming language, with models being fitted using the *forecast* and *fforma* packages[5]. Unless otherwise specified in Table 4-3, a Box-Cox transformation on each dataset was performed prior to model fitting, with the parameter being chosen using the Guerrero (1993) method. This was done to help avoid unrealistic, negative forecasts. Whenever a Box-Cox transformation is performed on a dataset before modelling, the fitted model's back-transformed predictions are the median of the forecast distribution (Hyndman and Athanasopoulos 2018). However, if the forecasts of the online and offline enquiries models are to be added together to produce total enquiries forecasts, the mean of the forecast distribution is required (Hyndman and Athanasopoulos 2018). Additionally, since enquiries are positively skewed, mean forecasts produce more conservative estimates of enquiries, which are more helpful for management to work with when planning. As a result, a bias adjustment is made to each back-transformed point forecast so that the final point forecast is the mean of the forecast distribution. For a given Box-Cox transformation parameter λ, the back-transformed mean of the forecast distribution is:

$$y_t = \begin{cases} e^{w_t}\left[1 + \frac{\sigma_h^2}{2}\right] & if \lambda = 0; \\ (\lambda w_t + 1)^{1/\lambda}\left[1 + \frac{\sigma_h^2(1-\lambda)}{2(\lambda w_t + 1)^2}\right] & otherwise; \end{cases} \tag{4-1}$$

where σ_h^2 is the h-step forecast variance, which is directly related to the size of the difference between the mean and median of the forecast distribution (Hyndman and Athanasopoulos 2018).

4.4 Individual Time-Series Modelling

In the individual time-series modelling stage, eight candidate models were fitted to online, offline and total enquiries to find the best-performing model for each time series.

---

[5] See https://github.com/robjhyndman/forecast and https://github.com/pmontman/fforma for more information about the *forecast* and *fforma* packages respectively.

Model performance was assessed using RMSE and MAE measures calculated through time-series cross validation comprising models built on the first 250-393 data points of the 394 observation dataset. The results for each enquiries dataset consist of three sections. The first two sections describe the results of the ARIMA and NNETAR models. The best-performing ARIMA and NNETAR models are then compared against the other candidate models in the third section to find the best-performing model for the time series.

4.4.1 Online Enquiries Modelling

4.4.1.1 ARIMA Model Choice

The results of the ARIMA model variations for online enquiries are presented in Table 4-5 and Figure 4-7 to Figure 4-8. As can be seen in Figure 4-7 and Figure 4-8, the model variations with exogenous variables are more accurate than the baseline ARIMA model across all forecast horizons. This suggests that including the calendar-based and advertising variables in the ARIMA model improves its predictive performance. Interestingly, there was not much difference between the model with total advertising expenditure (Total_Xreg) and the one without any advertising variables (No_Xreg). However, when total advertising expenditure was split into online and offline totals (All_Xreg and Extra_Xreg), more accurate forecasts were produced. This suggests that there can be value in splitting aggregate advertising budgets into more granular totals to capture additional information that would otherwise be lost due to aggregation. In this case, the extra value is likely due to the strong relationship between online advertising spend and online enquiries. This makes sense as online advertising drives consumers to the travel brand's website, where consumers can make online enquiries. Figure 4-7 and Figure 4-8 also show that the online and offline spend model with the weekend dummy variable (All_Xreg) outperforms the model with the day off dummy variable (Extra_Xreg), particularly at later forecast horizons. Thus, it seems that including public holidays in the weekend dummy variable does not result in additional forecast accuracy for online enquiries. As stated earlier, this is likely because online enquiries can still be made on public holidays as they are made via the travel brand's website. Overall, the ARIMA model with the online spend, offline spend, weekend and after rebrand exogenous variables (All_Xreg) was chosen as the best-performing model. Another advantage of this model is that while its forecast accuracy reduces over the first ten *h*-steps, it stabilizes thereafter.

	RMSE for h = 1-30			MAE for h = 1-30		
	Mean	**Median**	**Sum**	**Mean**	**Median**	**Sum**
ARIMA	33.86	34.05	1015.82	25.06	25.35	751.69
Total_Xreg	30.28	30.65	908.28	22.83	22.86	684.99
All_Xreg	28.51	29.06	855.20	20.51	21.14	615.40
No_Xreg	29.98	30.52	899.28	22.62	22.56	678.71
Extra_Xreg	29.29	29.98	878.55	21.03	21.39	630.87

Table 4-5 – Performance Metrics (best shaded) - ARIMA Models - Online Enquiries

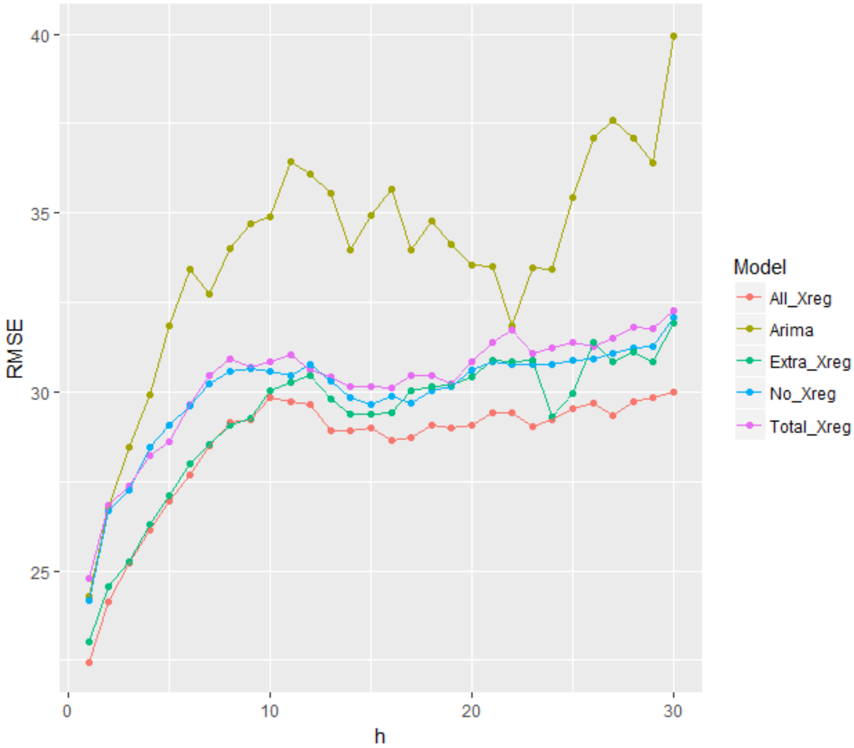


Figure 4-7 - RMSE for h = 1-30 – ARIMA Models - Online Enquiries

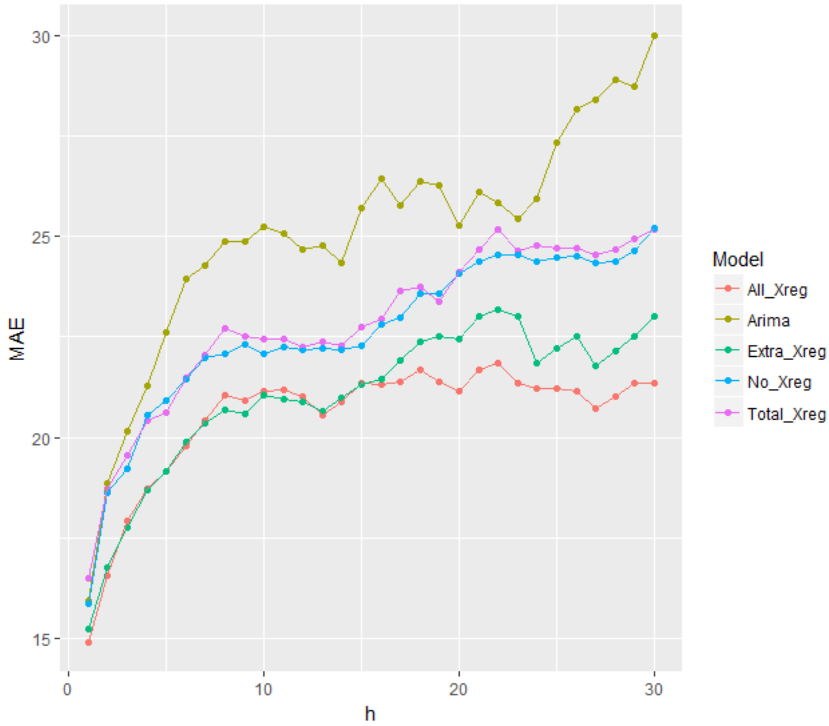


Figure 4-8 - MAE for h = 1-30 – ARIMA Models - Online Enquiries

4.4.1.2 NNETAR Model Choice

Table 4-6 and Figure 4-9 to Figure 4-10 describe the performance of the NNETAR model variations for online enquiries. Overall, the NNETAR models did not perform as well the ARIMA models. This is possibly because there are not enough data for the neural networks to produce accurate forecasts. As with the ARIMA model variations, the models with separate variables for online and offline advertising expenditure had lower RMSE and MAE values than the other variations. The accuracy of these models also worsened over the first ten *h*-step forecasts, before stabilizing over the remaining *h*-steps. The best-performing NNETAR model is the one with the online spend, offline spend, day off and after rebrand exogenous variables (Extra_Xreg). However, its performance is quite similar to the model variation with the weekend instead of the day off dummy variable (All_Xreg). For robustness, both model variations were compared against the other candidate models (only Extra_Xreg is reported), with there being no difference in overall model choice.

	RMSE for h = 1-30			**MAE for h = 1-30**		
	Mean	**Median**	**Sum**	**Mean**	**Median**	**Sum**
NNAR	35.31	35.48	1059.39	25.35	25.48	760.59
Total_Xreg	37.62	39.29	1128.55	27.53	28.03	825.87
All_Xreg	32.97	33.39	988.98	23.16	23.29	694.90
No_Xreg	37.06	38.37	1111.68	27.48	28.71	824.44

	RMSE for h = 1-30			MAE for h = 1-30		
	Mean	**Median**	**Sum**	**Mean**	**Median**	**Sum**
Extra_Xreg	32.78	33.22	983.36	23.00	23.41	689.85

Table 4-6 - Performance Metrics (best shaded) - NNETAR Models - Online Enquiries

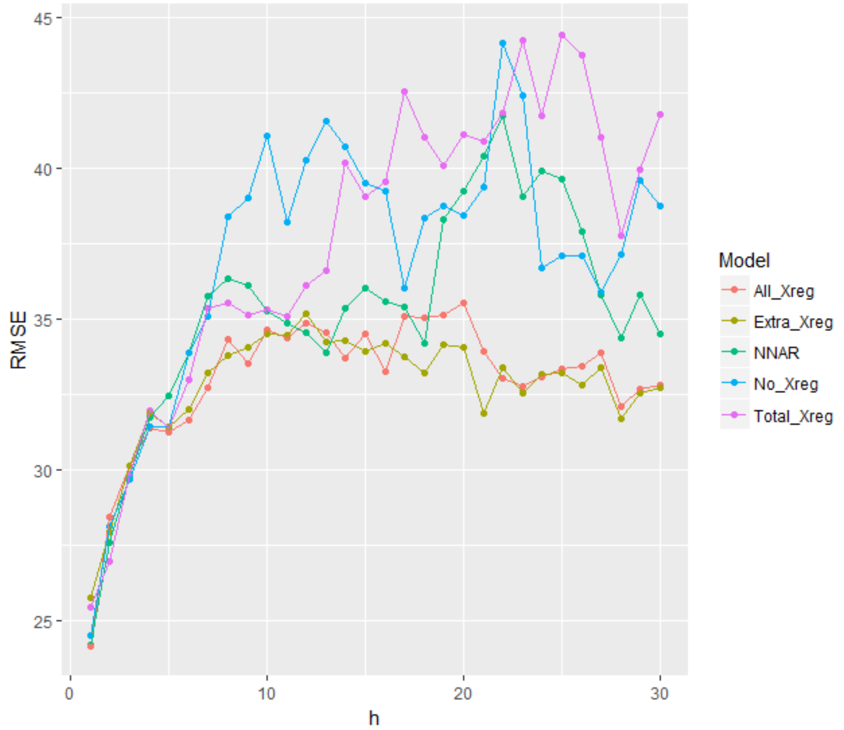


Figure 4-9 – RMSE for h = 1-30 – NNETAR Models - Online Enquiries

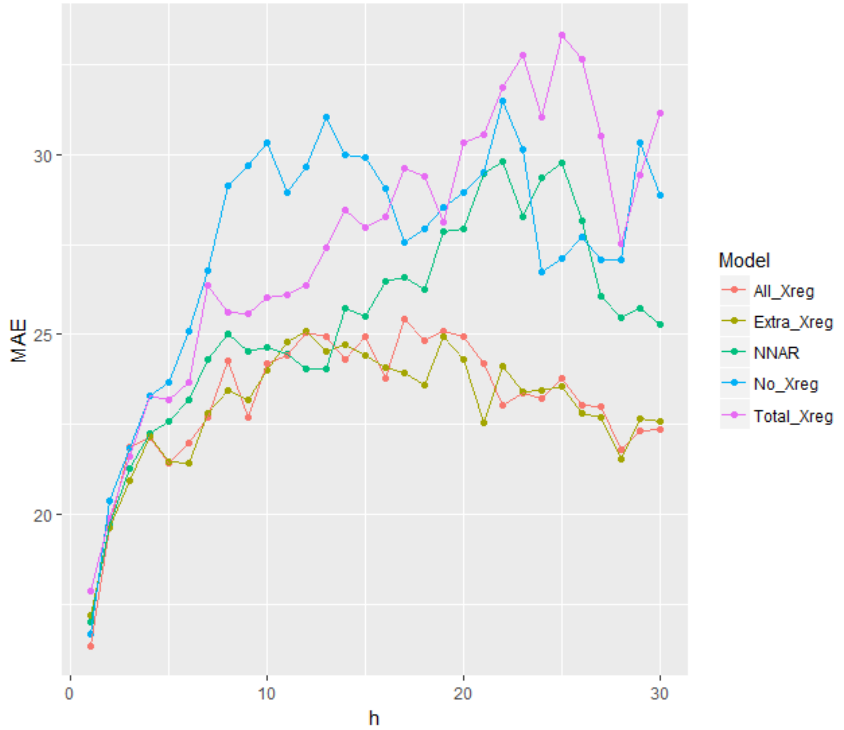


Figure 4-10 - MAE for h = 1-30 – NNETAR Models – Online Enquiries

4.4.1.3 Overall Model Choice

The performance of the eight candidate models fitted to the online enquiries time series is shown in Table 4-7 and Figure 4-11 to Figure 4-13. As can be seen in Figure 4-11 and Figure 4-12, while the RMSE and MAE values of the models are relatively close together for the early *h*-step forecasts, they diverge dramatically as the forecast horizon lengthens. In particular, the accuracy of the models with an exponential smoothing component, namely the ETS, STL and TBATS models, deteriorate rapidly the further into the future a forecast is made. This is likely because the errors in predictions at early forecast horizons compound as they are used to make predictions at later forecast horizons. The Ensemble and FFORMA approaches seem to be adversely affected by these models, with the simpler forecast averaging approach outperforming the more complicated FFORMA method. However, the FFORMA method did not incorporate the exogenous variables, which may explain its less accurate performance.

The online enquiries dataset seems to be more suited to the autoregressive style models, with the ARIMA, ARFIMA and NNETAR models having more accurate and stable predictive performance. The strong performance of the ARFIMA model suggests that the time series may exhibit long-memory characteristics. While the NNETAR model has higher RMSE and MAE values than the ARIMA and ARFIMA models, Figure 4-13 shows that its RMSE for predictions across each test subset improves as more data are included in model training. Thus, as more data are collected, the NNETAR model's performance may improve relative to the other methods. Overall, the ARIMA model with the online spend, offline spend, weekend and after rebrand exogenous variables produced the most accurate predictions across all forecast horizons. This shows that a relatively simple model, augmented with exogenous predictors, can outperform more complex methods, particularly at later forecast horizons. The ARIMA model appears to capture key relationships in the online enquiries time series that persist over time. While the predictions of the other models are comparable at smaller forecast horizons, they become increasingly noisy as the forecast horizon lengths. This is potentially due to the additional complexity of some of the other models, such as TBATS and FFORMA. These models tend to be more focused on eliminating bias from the data. While they can often produce more accurate next step forecasts, they are more likely to overfit the data. This becomes apparent when longer term forecasts are made and the short-term relationships that they detect in the data disappear, resulting in less accurate forecasts.

	RMSE for h = 1-30			MAE for h = 1-30			RMSE for Test Subsets (trained on 250-364 obs.)		
	Mean	**Median**	**Sum**	**Mean**	**Median**	**Sum**	**Mean**	**Median**	**Sum**
ETS	42.43	41.95	1273.02	33.42	32.72	1002.50	40.87	37.06	4699.54
ARIMA	28.51	29.06	855.20	20.51	21.14	615.40	27.80	28.83	3196.88
ARFIMA	29.58	30.04	887.32	22.19	22.68	665.66	29.40	29.00	3380.57
STL	41.61	41.30	1248.38	32.48	31.81	974.40	40.34	37.35	4639.59
TBATS	37.34	36.19	1120.06	31.14	30.47	934.31	37.92	38.53	4360.42
NNETAR	32.78	33.22	983.36	23.00	23.41	689.85	31.43	32.13	3614.26
Ensemble	30.19	30.37	905.55	23.46	23.32	703.80	30.14	29.43	3465.93
FFORMA	34.94	35.02	1048.09	27.30	27.07	819.11	34.42	31.57	3958.78

Table 4-7 - Performance Metrics (best shaded) – All Models - Online Enquiries

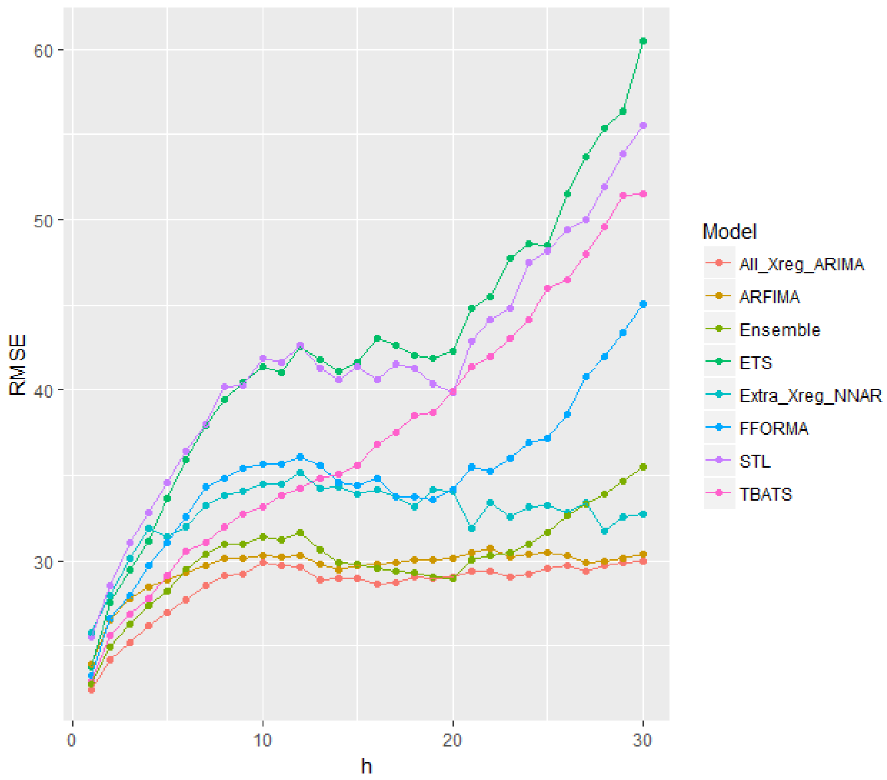


Figure 4-11 - RMSE for h = 1-30 – All Models – Online Enquiries

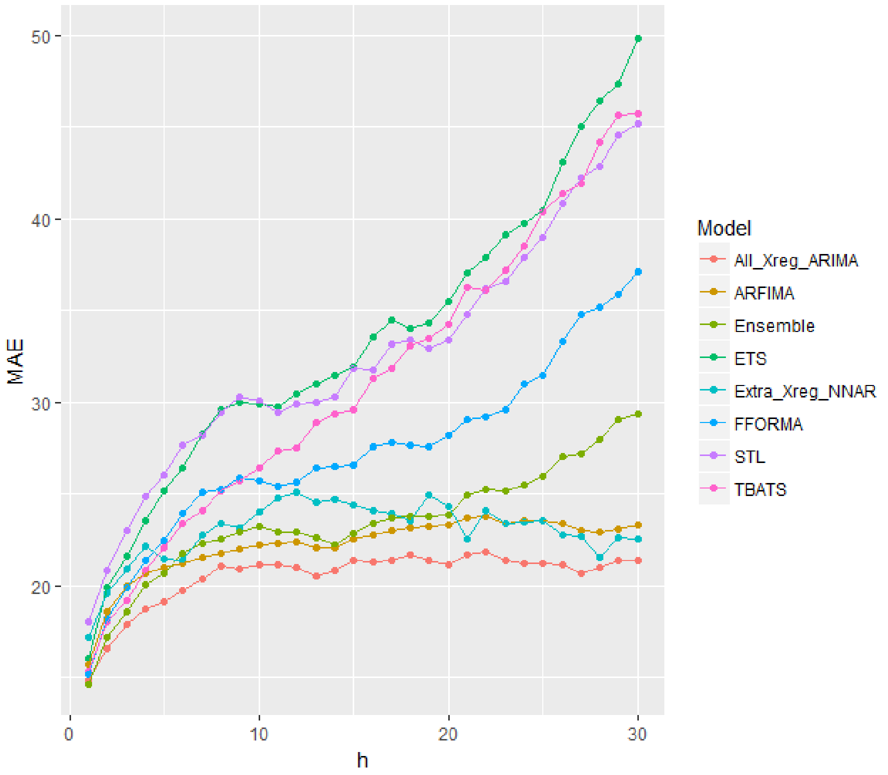


Figure 4-12 - MAE for h = 1-30 – All Models – Online Enquiries

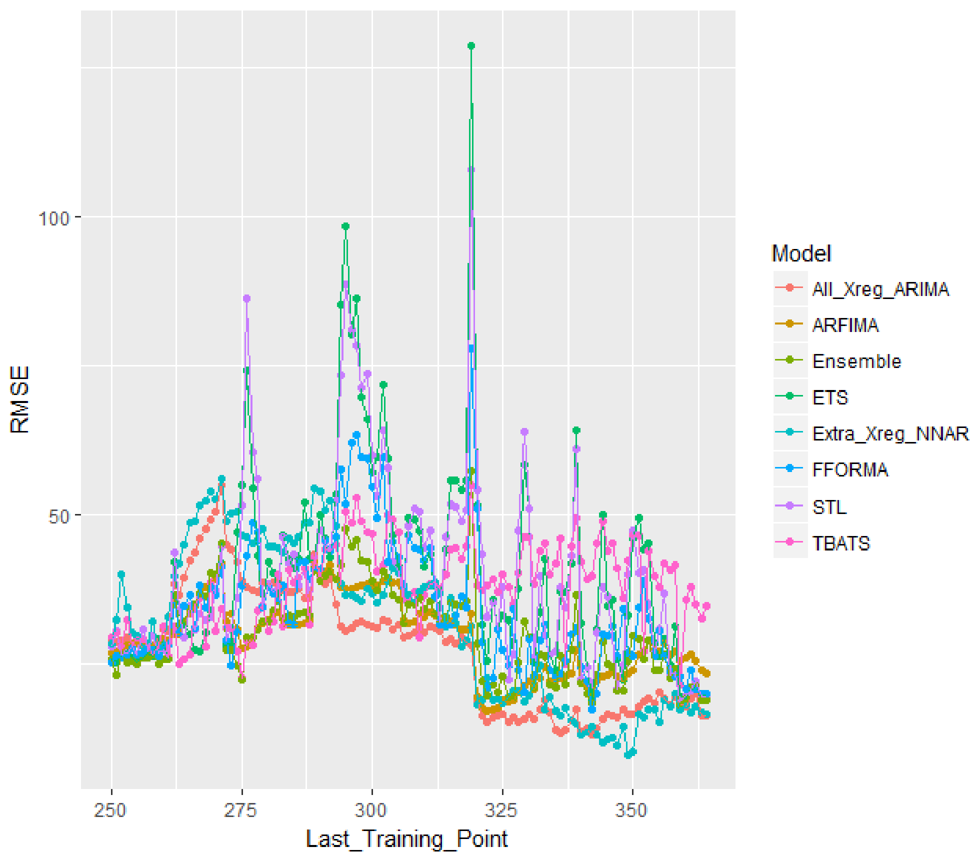


Figure 4-13 - RMSE for Test Subsets – All Models – Online Enquiries

4.4.2 Offline Enquiries Modelling

4.4.2.1 ARIMA Model Choice

The performance of the ARIMA model variations for offline enquiries is reported in Table 4-8 and Figure 4-14 to Figure 4-15. There is a clear separation between the models with exogenous variables and the baseline ARIMA model, suggesting that including them in the model improves forecast accuracy. Additionally, there is a large improvement between the models with the day off dummy variable (Total_Xreg, All_Xreg and No_Xreg) and the one with the weekend dummy variable (Extra_Xreg). This shows that accounting for public holidays when forecasting offline enquiries is quite important. Impressively, forecast accuracy is consistent across all forecast horizons for these models, with there being only a small reduction in accuracy as the forecast horizon increases. The best-performing ARIMA model variation across all forecast horizons is the one with the total advertising expenditure, day off and after rebrand exogenous variables. Thus, including both advertising and calendar-based exogenous predictors produces the most accurate predictions. However, accounting for calendar-based effects has a greater effect on forecast accuracy than the inclusion of advertising effects. Unlike online enquiries, using total advertising expenditure produces more accurate forecasts than splitting it into online and offline totals. This is potentially because online enquiries are particularly driven by online advertising expenditure, while offline enquiries seem to benefit from advertising spend more generally.

	RMSE for h = 1-30			**MAE for h = 1-30**		
	Mean	**Median**	**Sum**	**Mean**	**Median**	**Sum**
ARIMA	81.00	80.87	2430.12	52.92	52.97	1587.59
Total_Xreg	47.63	47.77	1428.96	33.31	33.62	999.30
All_Xreg	49.70	49.61	1491.00	34.34	34.56	1030.06
No_Xreg	49.44	49.38	1483.05	34.71	34.61	1041.39
Extra_Xreg	73.29	73.42	2198.70	47.74	46.91	1432.20

Table 4-8 - Performance Metrics (best shaded) – ARIMA Models - Offline Enquiries

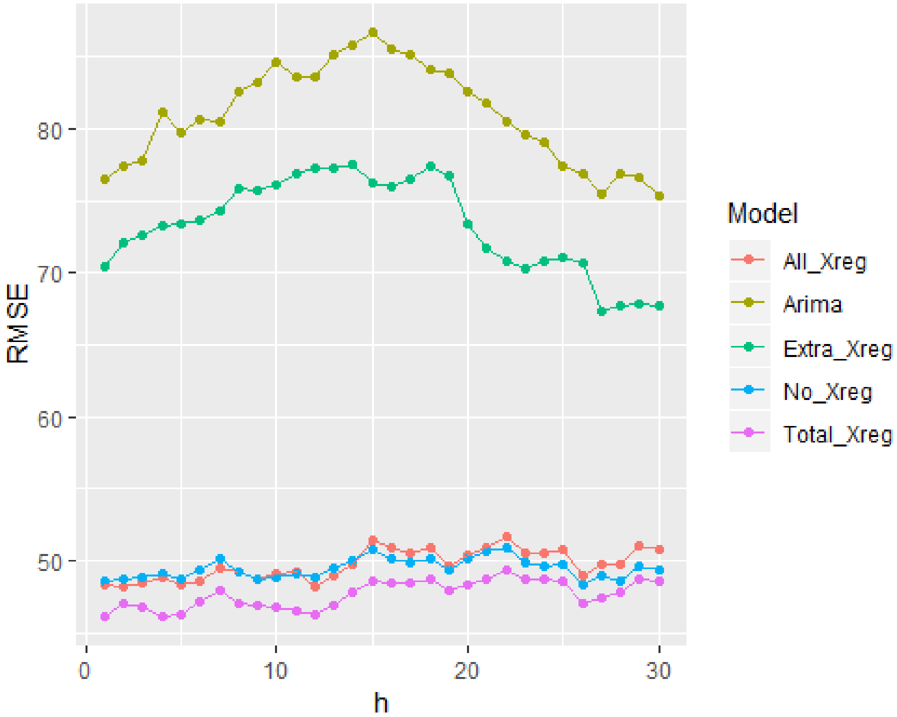


Figure 4-14 - RMSE for h = 1-30 – ARIMA Models – Offline Enquiries

Figure 4-15 - MAE for h = 1-30 – ARIMA Models – Offline Enquiries

4.4.2.2 NNETAR Model Choice

As can be seen in Table 4-9 and Figure 4-16 to Figure 4-17, the NNETAR models do not perform as well as the ARIMA models at predicting offline enquiries. This is similar to the results of modelling the online enquiries, with it being possible that there are not enough data for the advantages of the neural network to come through. As with the ARIMA models, the NNETAR models that accounted for public holidays via the day off dummy variable (Total_Xreg, All_Xreg and No_Xreg) produced more accurate forecasts than the others (NNETAR and Extra_Xreg). The individual RMSE and MAE values for the NNETAR models with the day off dummy variable are closer than the ARIMA models, with it being more difficult to see a clear outperformer in Figure 4-16 and Figure 4-17. However, Table 4-9 shows that the NNETAR model with the total spend, day off and after rebrand exogenous variables has the lowest average RMSE and MAE values.

	RMSE for h = 1-30			**MAE for h = 1-30**		
	Mean	**Median**	**Sum**	**Mean**	**Median**	**Sum**
NNETAR	115.87	113.71	3476.01	85.89	85.78	2576.82
Total_Xreg	78.53	77.19	2355.83	56.76	57.05	1702.87
All_Xreg	80.36	79.06	2410.95	59.30	58.64	1779.07
No_Xreg	81.64	81.44	2449.29	57.78	57.44	1733.33
Extra_Xreg	101.01	101.23	3030.16	72.20	71.92	2166.08

Table 4-9 - Performance Metrics (best shaded) – NNETAR Models - Offline Enquiries

Figure 4-16 - RMSE for h = 1-30 – NNETAR Models – Offline Enquiries

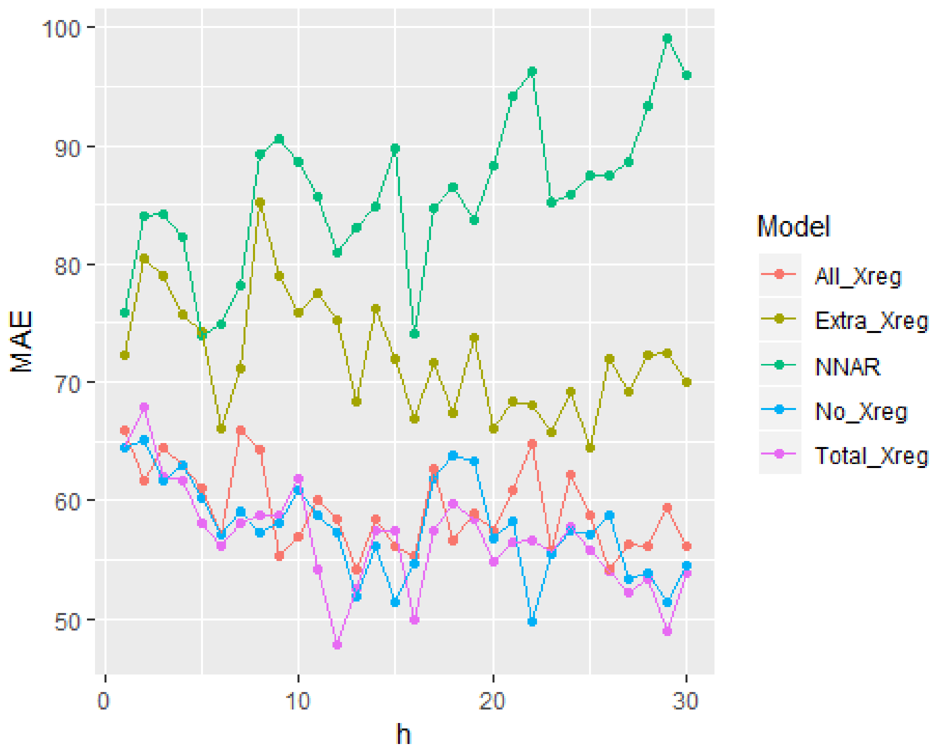


Figure 4-17 - MAE for h = 1-30 – NNETAR Models – Offline Enquiries

4.4.2.3 Overall Model Choice

The results of fitting the eight candidate models to the offline enquiries time series are presented in Table 4-10 and Figure 4-18 to Figure 4-20. Overall, the accuracy of the predictions by the models used to fit offline enquiries is less than it was for online enquiries. Thus, the online enquiries time series appears to be more predictable than the offline enquiries time series. This is potentially because online enquiries are more controlled as they are recorded electronically via the travel brand's website. Since offline enquiries are made via a call centre and involve manual recording, the potential for data entry error, such as an operator forgetting to log a call, is higher. Additionally, there is greater volume and variance in the offline enquiries time series, making it harder for a statistical model to capture all of the information contained within it. Interestingly, the ARFIMA model has the highest RMSE and MAE values across all forecast horizons, despite being one of the best-performing models for online enquiries. The ARFIMA model's relatively poor performance indicates that there are no long-memory characteristics in the offline enquiries time series and that it is more strongly affected by seasonal factors not captured by the model.

The best-performing model is the ARIMA model with the total advertising expenditure, day off and after rebrand exogenous variables. This model outperforms all of the others by a considerable margin across all forecast horizons and performance metrics. The results reinforce the previous finding that a relatively simple model augmented with exogenous variables can outperform a more complex time-series model. The ARIMA model does this by bringing in additional information, not contained in the history of the offline enquiries time series, via its exogenous predictors. Given the model's outperformance, the increase in predictive accuracy from the exogenous variables seems to be more pronounced for offline enquiries than it is for online enquiries. This is likely due to offline enquiries being more affected by calendar-based effects, such as reduced call centre hours on weekends, than online enquires. After the ARIMA model, the models with the highest overall predictive accuracy are the Ensemble and FFORMA approaches. Interestingly, the Ensemble method has lower RMSE values across all forecast horizons, while the FFORMA model has lower MAE values. Thus, it appears that in this case the Ensemble method produces more accurate predictions of the mean, while the FFORMA method produces more accurate predictions of the median.

	RMSE for h = 1-30			**MAE for h = 1-30**			**RMSE for Test Subsets (trained on 250-364 obs.)**		
	Mean	**Median**	**Sum**	**Mean**	**Median**	**Sum**	**Mean**	**Median**	**Sum**
ETS	83.52	84.71	2505.46	54.89	56.89	1646.56	76.43	84.42	8788.89
ARIMA	47.63	47.77	1428.96	33.31	33.62	999.30	48.37	49.12	5562.66
ARFIMA	124.10	125.10	3723.73	112.70	113.70	3381.78	123.31	121.83	14181.01
STL	103.32	103.56	3099.53	72.42	71.88	2172.53	96.78	92.39	11130.27
TBATS	81.01	82.56	2430.41	61.26	62.11	1837.91	77.23	76.02	8881.46
NNETAR	78.53	77.19	2355.83	56.76	57.05	1702.87	74.52	71.97	8570.27
Ensemble	69.01	70.02	2070.27	51.07	51.79	1532.22	66.63	65.09	7662.73
FFORMA	75.38	76.71	2261.42	47.74	49.09	1432.32	68.94	71.97	7927.90

Table 4-10 - Performance Metrics (best shaded) – All Models - Offline Enquiries

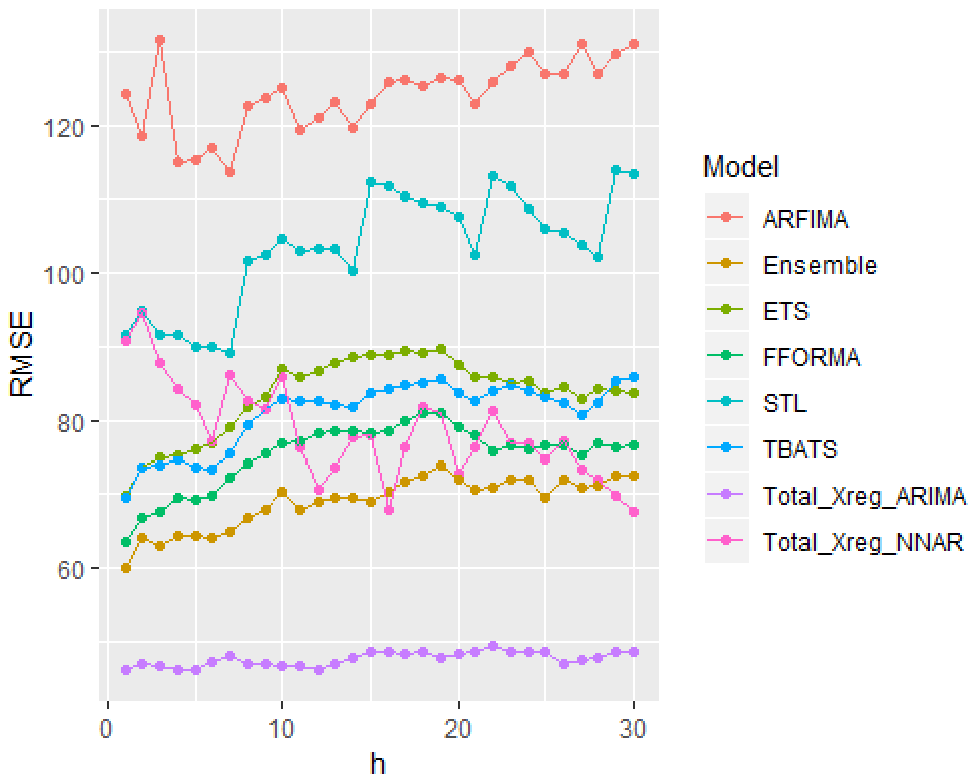


Figure 4-18 - RMSE for h = 1-30 – All Models – Offline Enquiries

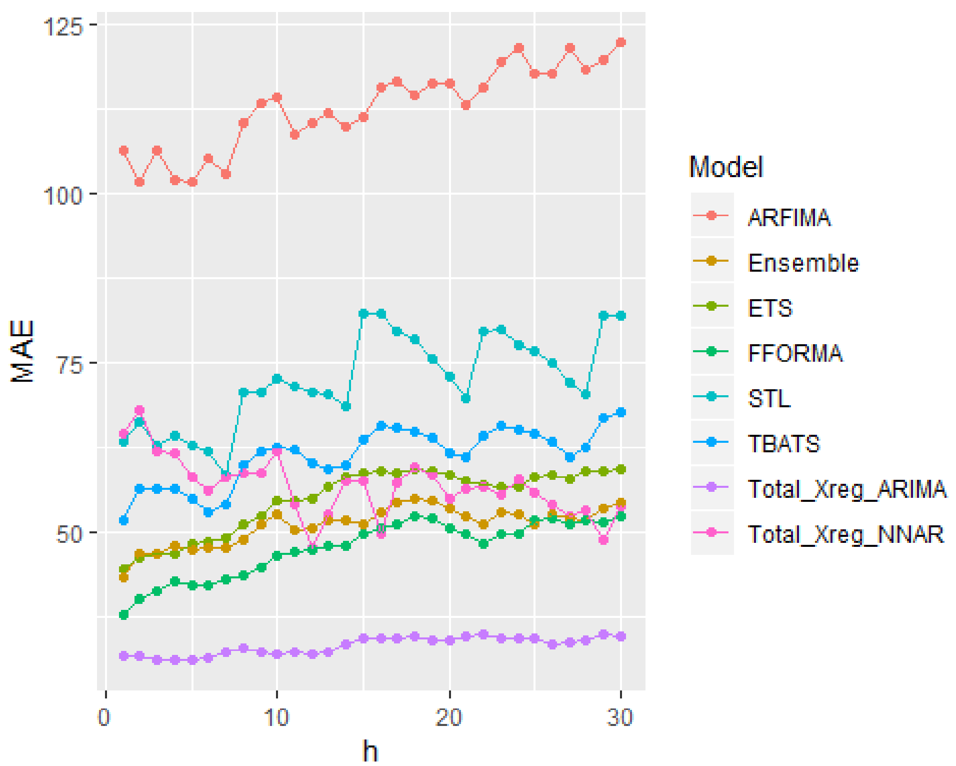


Figure 4-19 - MAE for h = 1-30 – All Models – Offline Enquiries

Figure 4-20 - RMSE for Test Subsets – All Models – Offline Enquiries

4.4.3 Total Enquiries Modelling

4.4.3.1 ARIMA Model Choice

Figure 4-21 and Figure 4-22 show the RMSE and MAE values for each *h*-step forecast of the total enquiries ARIMA model variations, while Table 4-11 summarizes them across all forecast horizons. The results of the ARIMA models mirror those of the offline enquiries, with the model with total advertising spend, day off and after rebrand exogenous variables producing the most accurate forecasts. This is likely because offline enquiries make up a large portion of total enquiries, with the relationships present in the subset time series appearing to translate over to the aggregate total. However, the accuracy of the models for total enquiries is less stable than the models for offline enquiries, gradually decreasing as the forecast horizon lengthens.

	RMSE for h = 1-30			**MAE for h = 1-30**		
	Mean	**Median**	**Sum**	**Mean**	**Median**	**Sum**
ARIMA	86.39	88.48	2591.75	60.09	61.90	1802.78
Total_Xreg	55.95	56.59	1678.65	40.45	41.52	1213.53
All_Xreg	61.60	62.37	1848.06	42.87	44.04	1286.10
No_Xreg	57.03	57.88	1711.00	41.91	42.98	1257.37

	RMSE for h = 1-30			MAE for h = 1-30		
	Mean	**Median**	**Sum**	**Mean**	**Median**	**Sum**
Extra_Xreg	78.72	78.64	2361.64	54.99	55.27	1649.80

Table 4-11 - Performance Metrics (best shaded) – ARIMA Models - Total Enquiries

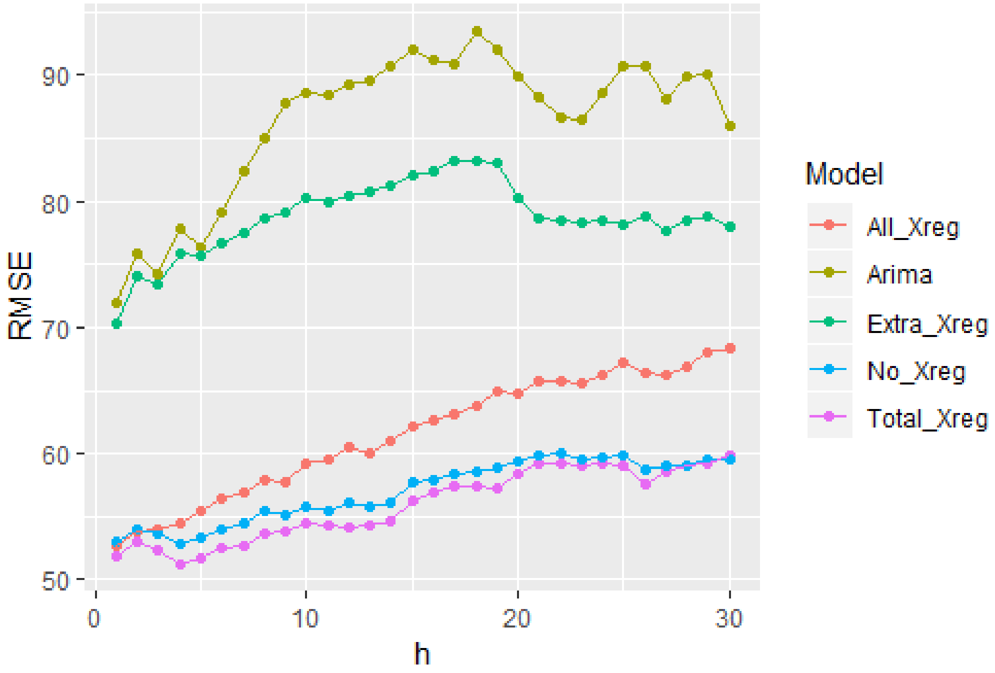


Figure 4-21 - RMSE for h = 1-30 – ARIMA Models – Total Enquiries

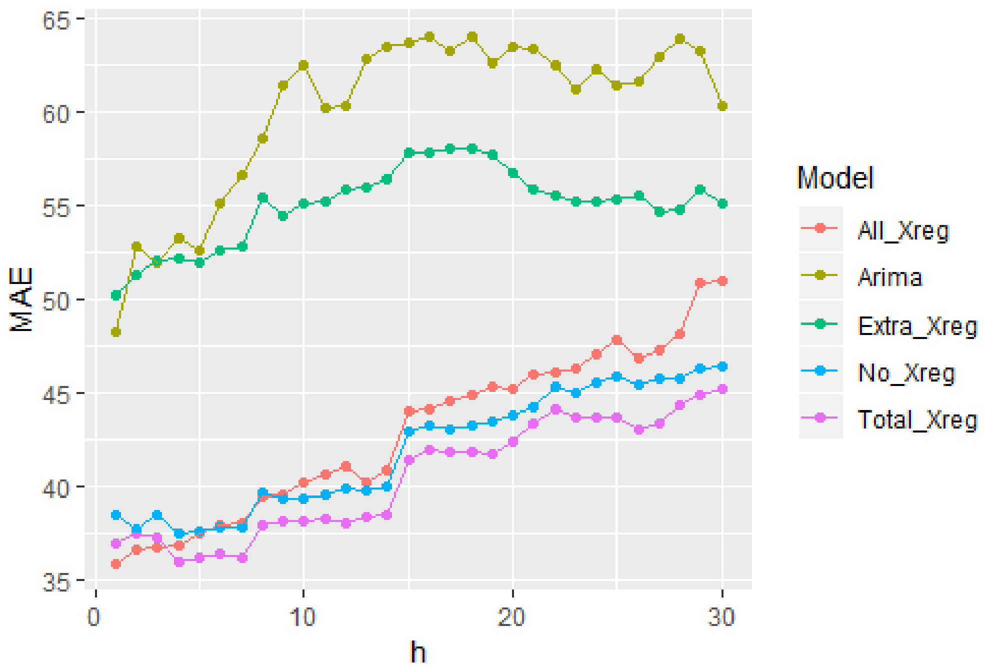


Figure 4-22 - MAE for h = 1-30 – ARIMA Models – Total Enquiries

4.4.3.2 NNETAR Model Choice

The performance of the NNETAR models for total enquiries is shown in Table 4-12 and Figure 4-23 to Figure 4-24. As with online and offline enquiries, the NNETAR models for total enquiries did not perform as well as the ARIMA models. The model variation with the lowest RMSE and MAE values across all forecast horizons is the model with the day off and after rebrand exogenous variables. Interestingly, this model outperformed the model variations with advertising expenditure data. This underscores the relative contribution of the calendar-based variables over the advertising variables to predictive accuracy.

	RMSE for h = 1-30			**MAE for h = 1-30**		
	Mean	**Median**	**Sum**	**Mean**	**Median**	**Sum**
NNETAR	107.05	108.26	3211.64	76.39	76.60	2291.85
Total_Xreg	71.42	70.86	2142.51	52.45	52.62	1573.64
All_Xreg	77.14	77.17	2314.30	56.14	56.51	1684.23
No_Xreg	66.64	66.44	1999.09	48.30	47.89	1448.94
Extra_Xreg	87.99	87.31	2639.55	62.35	62.08	1870.54

Table 4-12 - Performance Metrics (best shaded) – NNETAR Models - Total Enquiries

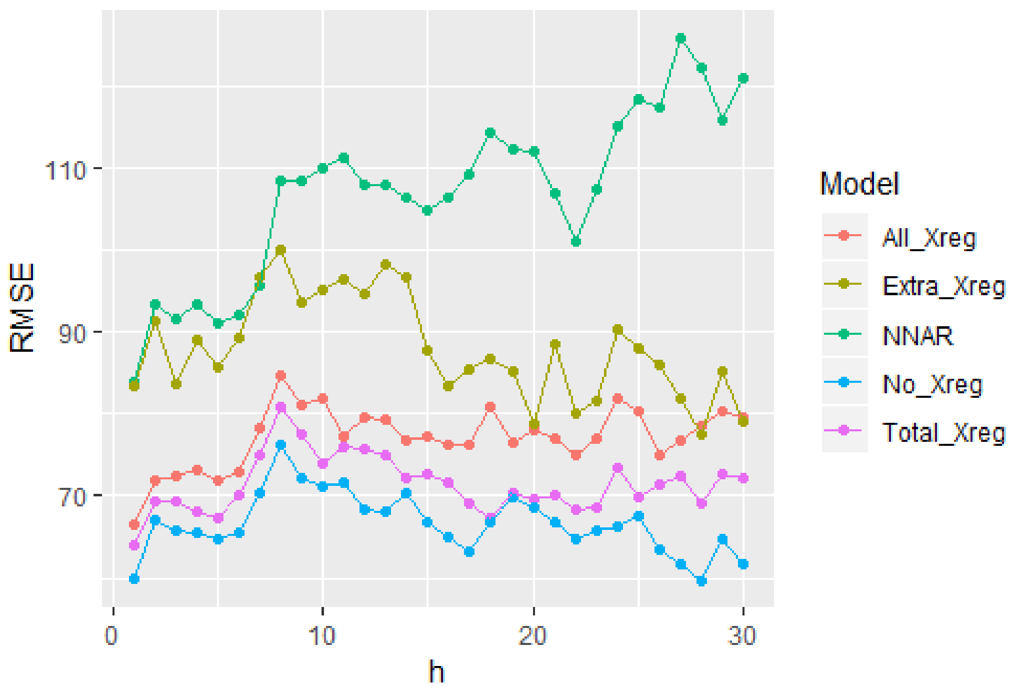


Figure 4-23 - RMSE for h = 1-30 – NNETAR Models – Total Enquiries

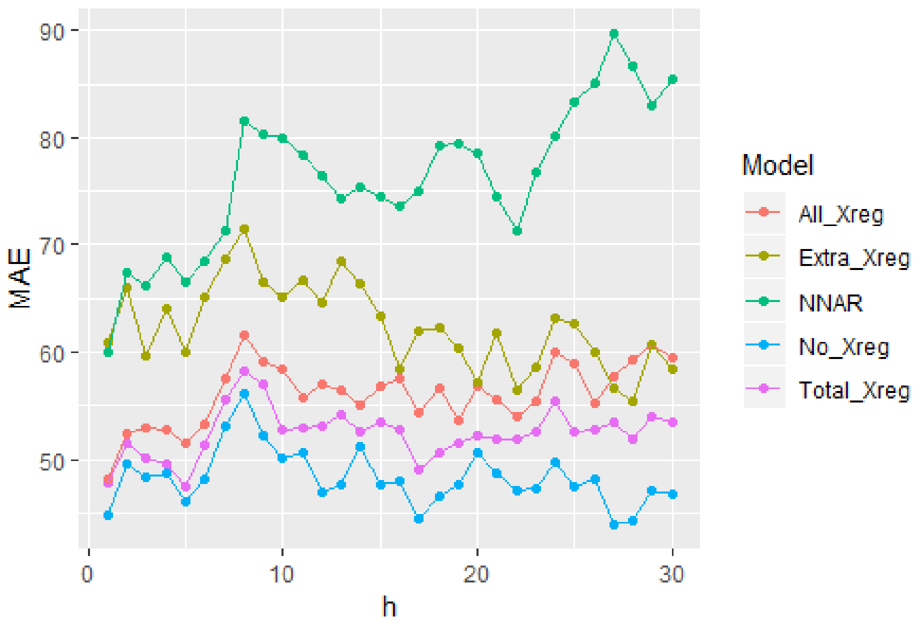


Figure 4-24 - MAE for h = 1-30 – NNETAR Models – Total Enquiries

4.4.3.3 Overall Model Choice

The performance measures of the time-series cross validation for total enquiries are shown in Table 4-13 and Figure 4-25 to Figure 4-27. The results mirror those of the offline enquiries modelling, which is likely a result of offline enquiries being the predominant enquiry channel. The best-performing model across all forecast horizons and performance metrics is the ARIMA model with the total advertising expenditure, day off and after rebrand exogenous variables. Once again, the results show that the inclusion of calendar-based and advertising variables can improve the forecast accuracy of a relatively simple time-series model so that it outperforms more complex pure time-series ones. The NNETAR model has the second highest overall accuracy, performing particularly well in later forecast horizons. This model also contains exogenous predictors, namely the day off and after rebrand variables, reinforcing the potential benefits of including calendar-based effects in time-series models.

	RMSE for h = 1-30			**MAE for h = 1-30**			**RMSE for Test Subsets (trained on 250-364 obs.)**		
	Mean	**Median**	**Sum**	**Mean**	**Median**	**Sum**	**Mean**	**Median**	**Sum**
ETS	87.68	88.35	2630.34	61.90	62.68	1857.09	80.99	79.14	9314.31
ARIMA	55.95	56.59	1678.65	40.45	41.52	1213.53	56.52	55.92	6500.13
ARFIMA	110.16	110.91	3304.68	88.56	89.05	2656.80	107.97	107.00	12416.93
STL	97.22	98.35	2916.48	69.15	69.79	2074.35	90.44	90.30	10401.10
TBATS	81.95	84.01	2458.40	58.16	59.10	1744.76	76.33	80.92	8777.81
NNETAR	66.64	66.44	1999.09	48.30	47.89	1448.94	64.66	63.05	7435.46
Ensemble	70.63	71.06	2119.00	52.46	52.61	1573.95	67.26	66.56	7734.36

	RMSE for h = 1-30			MAE for h = 1-30			RMSE for Test Subsets (trained on 250-364 obs.)		
	Mean	**Median**	**Sum**	**Mean**	**Median**	**Sum**	**Mean**	**Median**	**Sum**
FFORMA	80.25	81.66	2407.48	54.06	55.41	1621.74	74.03	74.88	8513.61

Table 4-13 - Performance Metrics (best shaded) – All Models - Total Enquiries

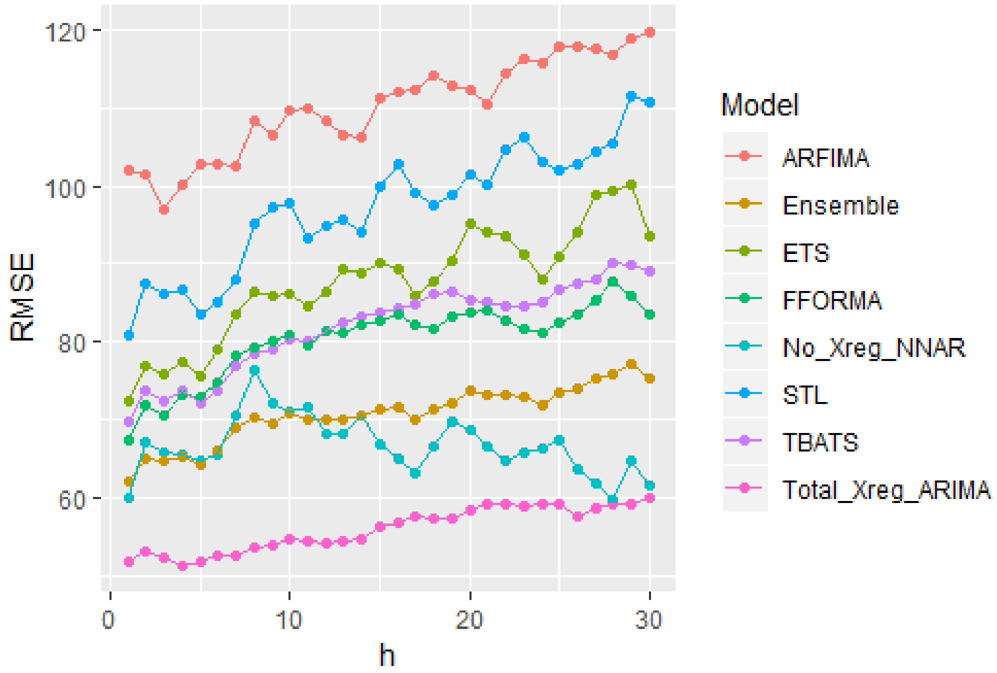


Figure 4-25 - RMSE for h = 1-30 – All Models – Total Enquiries

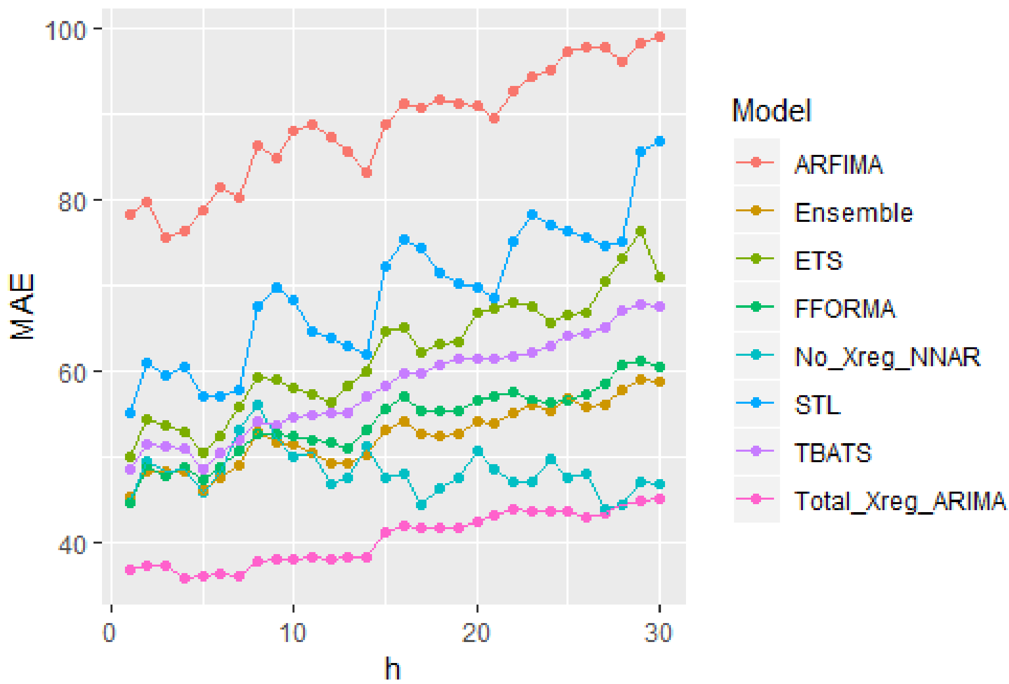


Figure 4-26 - MAE for h = 1-30 – All Models – Total Enquiries

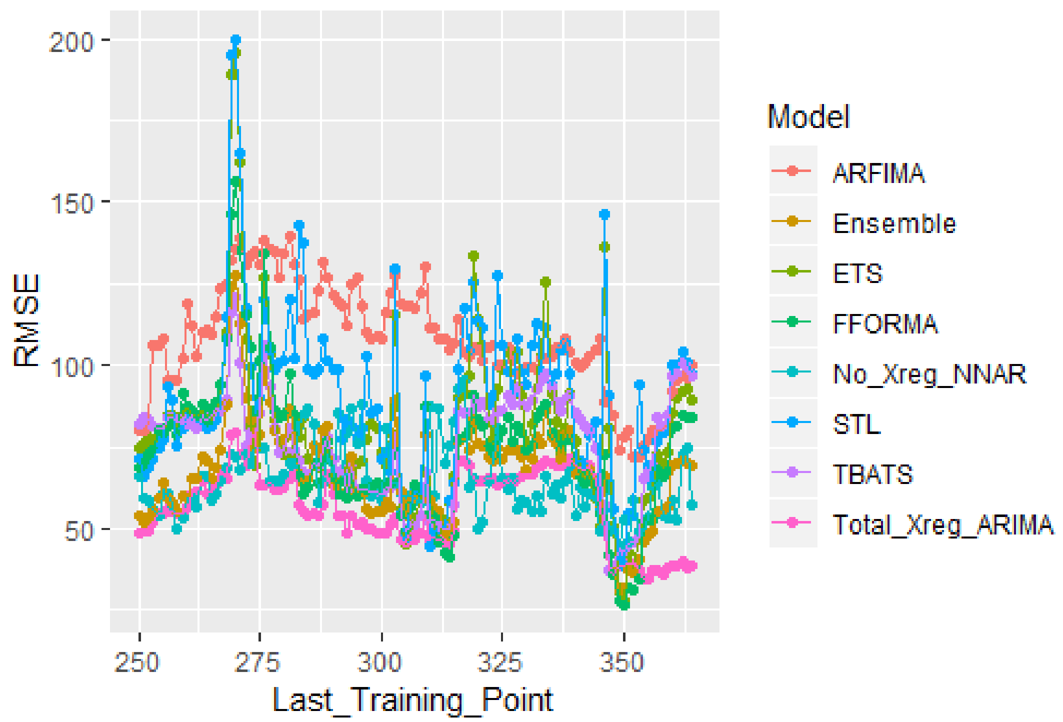


Figure 4-27 - RMSE for Test Subsets – All Models – Total Enquiries

4.4.4 Summary of Chosen Models

In summary, the best-performing models for online, offline and total enquiries are:

- **Online Enquiries** – ARIMA model with online advertising expenditure, offline advertising expenditure, weekend and after rebrand exogenous variables.
- **Offline Enquiries** – ARIMA model with total advertising expenditure, day off and after rebrand exogenous variables.
- **Total Enquiries** - ARIMA model with total advertising expenditure, day off and after rebrand exogenous variables.

For closer examination, the chosen ARIMA models for online, offline and total enquiries were trained on the first 364 observations of the data, leaving a test set consisting of the final 30 observations. The coefficients and in-sample residuals of each fitted model are shown in Table 4-14 to Table 4-16 and Figure 4-28 to Figure 4-30. The mean of each model's residuals appears to be close to zero and there is no significant correlation between residuals. Figure 4-29 and Figure 4-30 show that there are a few outlier points in the offline and total enquiries data. This is reflected in the histograms of their residuals, which are negatively skewed. Since these residuals are not normally distributed, it would be more accurate to create prediction intervals for point forecasts using bootstrapping. Table 4-14 to Table 4-16 report the coefficients for each of the models, with weekly seasonality being detected for each one. The coefficients for the exogenous variables in each model also make sense, with the total advertising spend and after rebrand variables having a positive effect on enquiries, while the weekend and day off variables have a negative effect. Interestingly, online advertising expenditure has a large positive effect on online enquiries, but offline advertising expenditure has a slightly negative effect. This may suggest that on days when offline advertising expenditure is increased, there is less spent on online advertising, which is a stronger driver of online enquiries.

Coefficients - ARIMA(1, 0, 0)(1, 0, 1)[7]			**Performance Metrics**	
Name	**Value**	**Standard Error**	**Metric**	**Value**
ar1	0.8074	0.0519	AIC	375.12
ma1	-0.3811	0.0789	BIC	414.09
sar1	0.8721	0.1190	RMSE	21.22
sma1	-0.7729	0.1472	MAE	14.35
Intercept	-2.0281	0.6282		
Online Spend	0.7378	0.0724		
Offline Spend	-0.0171	0.0090		
Weekend	-0.1317	0.0726		
After Rebrand	0.1896	0.1766		

Table 4-14 - Fitted Model - Online Enquiries

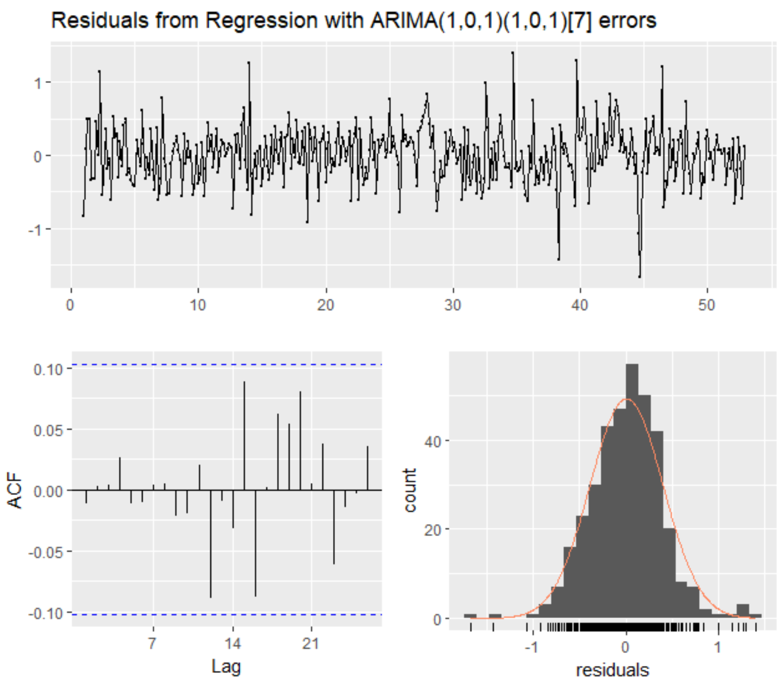


Figure 4-28 - In-Sample Residuals – Online Enquiries

Coefficients - ARIMA(0, 0, 0)(2, 0, 0)[7]			Performance Metrics	
Name	**Value**	**Standard Error**	**Metric**	**Value**
sar1	0.1148	0.0522	AIC	1377.79
sar2	0.0976	0.0519	BIC	1405.07
Intercept	4.2026	1.3394	RMSE	40.96
Total Spend	0.4814	0.1503	MAE	27.55
Day Off	-5.1468	0.2214		
After Rebrand	2.2458	0.2111		

Table 4-15 - Fitted Model – Offline Enquiries

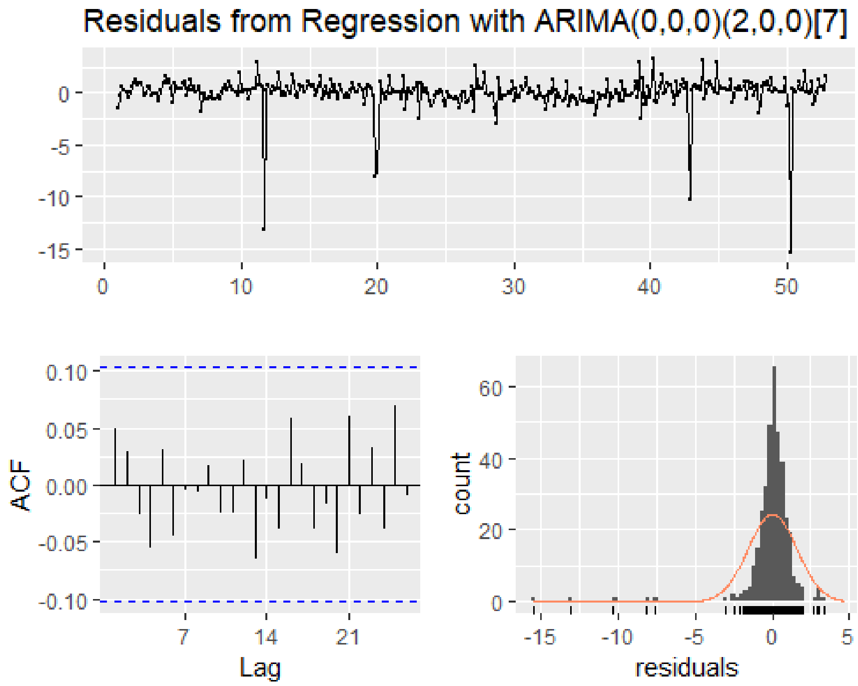


Figure 4-29 - In-Sample Residuals – Offline Enquiries

Coefficients - ARIMA(2, 0, 0)(2, 0, 0)[7]			Performance Metrics	
Name	**Value**	**Standard Error**	**Metric**	**Value**
ar1	0.3037	0.0533	AIC	846.97
ar2	0.1424	0.0536	BIC	882.05
sar1	0.1836	0.0535	RMSE	42.97
sar2	0.1772	0.0539	MAE	30.17
Intercept	6.4997	0.7715		
Total Spend	0.3701	0.0854		
Day Off	-3.4994	0.1245		
After Rebrand	2.1009	0.2116		

Table 4-16 - Fitted Model – Total Enquiries

Figure 4-30 - In-Sample Residuals – Total Enquiries

One common practical implementation question is whether it is better to refit demand forecasting models daily as new data comes in, or to only refit them periodically. To investigate this, one-step ahead forecasts were made by the models for each of the 30 observations in the test set. Each observation was predicted using all relevant prior data to reflect the collection of new data points over time. For each of the models, the RMSE and MAE values of these predictions were calculated and compared to the RMSE and MAE of the *h* = *1* predictions on the last 30 time-series cross validation test sets. The predictions of the time-series cross validation test sets are equivalent to one-step ahead predictions made by models that have been refitted each day as new data becomes available. The RMSE and MAE values are presented in Table 4-17.

Forecast	RMSE	MAE
Online Enquiries – 1-Step	15.76	12.31
Online Enquiries - Refit	15.74	12.34
Offline Enquiries – 1-Step	25.96	20.84
Offline Enquiries - Refit	25.75	20.40
Total Enquiries – 1-Step	40.96	33.17
Total Enquiries - Refit	40.18	32.48

Table 4-17 – Online, Offline & Total Enquiries Out-of-Sample Performance Metrics (best shaded)

The out-of-sample performance measures in Table 4-17 show that the 1-step ahead predictions from the models refitted daily to include new data are more accurate than the ones from the models fitted once on the training set of 364 data points. This suggests that more accurate predictions can be produced by refitting the online, offline and total enquiries models daily to incorporate new data as it becomes available. However, the

RMSE and MAE values for both forecasting approaches are reasonably close, meaning that the improvement in accuracy may not make a material difference in practice.

4.5 Hierarchical Modelling

In the hierarchical modelling stage, the best-performing models from the individual time-series modelling stage were used to test different approaches to forecasting total enquiries. Bottom-up forecasts were created by adding up forecasts from the online and offline enquiries models, while the total enquiries model produced aggregate forecasts. Ensemble forecasts were also created by averaging the bottom-up and aggregate forecasts. The bottom-up, aggregate and ensemble forecasts of total enquiries were assessed using the same techniques as the individual time-series models, namely time-series cross validation with RMSE and MAE measurements for each forecast horizon and test set. To see how the performance of the forecasting approaches changed as the amount of data used in the models increased, time-series cross validation was performed using five different values of X: 200, 250, 280, 300 and 330. Figure 4-31 is a collection of plots showing the RMSE values for each model across h = 1-30 at each value of X. Table 4-18 and Table 4-19 summarize Figure 4-31 by respectively providing the sum and mean of each model's RMSE values across all forecast horizons at each value of X. Since the MAE measurements produced similar results to the RMSE values, they are not reported.

X	Total	Bottom-Up	Ensemble
200	1792.961	2035.413	1861.074
250	1678.645	1737.398	1679.397
280	1609.758	1589.803	1570.030
300	1589.751	1513.990	1523.634
330	1459.368	1315.228	1353.863

Table 4-18 - Hierarchical Modelling - Total RMSE for h = 1-30 (best shaded)

X	Total	Bottom-Up	Ensemble
200	59.77	67.85	62.04
250	55.95	57.91	55.98
280	53.66	52.99	52.33
300	52.99	50.47	50.79
330	48.62	43.84	45.13

Table 4-19 - Hierarchical Modelling - Mean RMSE for h = 1-30 (best shaded)

Figure 4-31 - Hierarchical Modelling – Time-Series Cross Validation Repeats

As can be seen in Figure 4-31 and Table 4-18 to Table 4-19, the aggregate approach to forecasting total enquiries is the most accurate across all forecast horizons when time-series cross validation starts with a model fitted on 200 data points. However, as the number of points used to fit the initial model increases, the results begin to change. At $X = 250$, the performance of the ensemble and aggregate approaches are much closer together. At $X = 280$, the ensemble approach is the best overall when the RMSE values of each approach are added and averaged across the forecast horizons in Table 4-18 and Table 4-19 respectively. When the RMSE is examined across each forecast horizon in Figure 4-31, the ensemble approach outperforms at the early forecast horizons, but the bottom-up approach is more accurate at the later ones. This pattern grows stronger over the $X = 300$ and $X = 330$ time-series cross validation repeats, with the bottom-up approach emerging as the approach with the lowest average and total RMSE across all forecast horizons. It should be noted that the drop in RMSE for the later forecast horizons in the $X = 330$ repeat is likely due to less errors being available towards the end of the dataset. For example, $h = 30$ forecasts are not usable for models fitted on more than 364 data points as there are only 394 observations in total.

The pattern shown in Figure 4-31 and Table 4-18 to Table 4-19 is repeated when RMSE values are calculated across all $h = 1\text{-}30$ forecasts for a particular train and test subset, rather than for a single forecast horizon across all train and test subsets. Calculating RMSE across all forecast horizons for a particular train and test subset provides an overall view of how a particular model performed at predicting the next 30 days of total enquiries. These results are shown in Figure 4-32. In a similar way to the individual forecast horizon analysis, Figure 4-32 shows that when there are less data points used to train a model, the RMSE for the aggregate approach is smaller. As the number of training points increase, the ensemble and bottom-up approaches become more accurate than the aggregate approach.

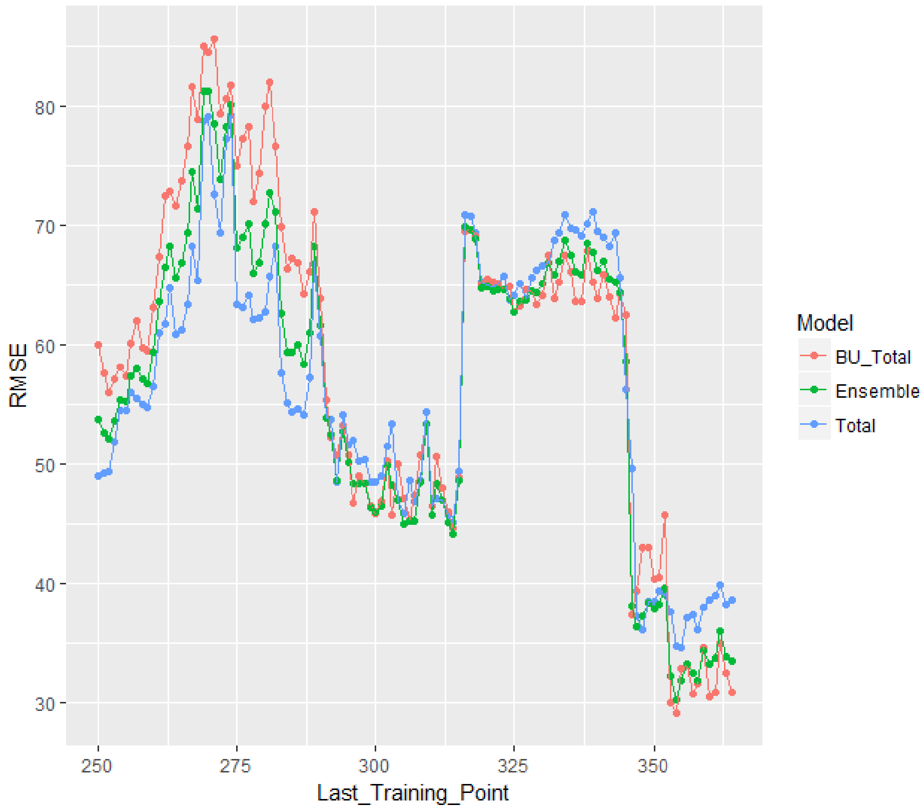


Figure 4-32 - Hierarchical Modelling - RMSE for Test Subsets

The results of the hierarchical modelling are more complicated than the individual time-series modelling, with relative model performance changing as the minimum number of observations used in time-series cross validation increases. For models build on less data, the aggregate approach to forecasting total enquiries is more accurate. However, as more data are used in model training, the bottom-up and ensemble approaches become more accurate. This is likely because the additional data helps reduce the noise in the online and offline enquiries time series and exacerbates the information-loss due to data aggregation in the total enquiries time series. More generally, the results suggest that when a company has a small training dataset, more accurate total demand forecasts will be produced if the top-level time series is forecasted by itself. However, as more data are collected, more accurate total demand forecasts can be produced by switching to a bottom-up or ensemble approach, which better capture the nuances in each bottom-level time series. Since more data becomes available for time-series modelling over time, it is important to periodically repeat the model fitting process to see if model selection changes. While the differences between the models may not be materially different for some applications, they are likely to be important in situations where the accuracy of point forecasts is critical, such as electricity demand or computing resource forecasting.

Another finding that emerges from the hierarchical modelling, and echoed in the individual time-series modelling, is that different modelling approaches can perform better at different forecast horizons. Figure 4-31 shows that the ensemble approach produces more accurate predictions at the early forecast horizons, while the bottom up approach's predictions are more accurate at later forecast horizons. The online enquiries modelling has similar results, with the performance of more complex models, such as TBATS and FFORMA, deteriorating more than the simpler ARIMA model as the forecast horizon lengthens. This suggests that more complex models tend to perform better when making short-term predictions, while more simple models perform relatively better when making long-term predictions. The complexity of models likely works against them in making longer-term forecasts as some of the relationships that they detect likely breakdown over time. In light of this finding, it is important for companies to consider different models for different forecast horizons, rather than assuming that the model that produces the best 1-step ahead predictions will be the most accurate at all forecast horizons.

4.5.1 Development of Repeated Time-Series Cross Validation (RSTCV)

The hierarchical modelling results also highlight a limitation of time-series cross validation, namely its sensitivity to the number of data points used to fit the initial model (X). For example, if the analysis only included the time-series cross validation results of X = 200, the aggregate modelling approach would appear to produce the most accurate predictions. However, when other values of X are examined, it can be seen that the ensemble and bottom up approaches are likely the most accurate for future predictions. This is especially true considering that more data will be collected and become available for training over time. Essentially, as the amount of data available for training a model increases, the time-series model that produces the most accurate forecasts may change. Thus, it can be seen that there is a trade-off when choosing the value of X. Since errors from models built with more training data are likely to be more reflective of future model performance, it is important to set a high value of X. However, if X is set too high, there may only be a small number of train and test subsets to assess model performance. If the errors from these subsets are not representative of the time series, the results are likely to be misleading. Therefore, X needs to be large enough so that there is enough data used in the initial models, but small enough that there are a reasonable number of errors with which to assess model performance.

Since it is difficult to identify an appropriate value of *X*, it is important for practitioners to perform time-series cross validation with several different values of *X* to make sure that they gain a representative understanding of model performance. An alternative to this is to extend the time-series cross validation technique so that it is less sensitive to the number of observations used to fit the initial model. Essentially, an improved technique needs to place more weight on errors from models built with more data, but still consider enough errors so that the results are generalizable. This can be achieved by adapting repeated cross validation, an extension to k-fold cross validation, to a time-series context. Repeated cross validation involves performing k-fold cross validation multiple times and calculating model performance by averaging it over all repeats. Since data are split into folds randomly, running the cross validation procedure multiple times helps reduce the noise from individual runs and produce a more accurate model performance assessment. This same concept can be applied in a time-series context by repeating time-series cross validation multiple times, with each repeat using a different value of *X*, and subsequently averaging the results. More specifically, the steps in repeated time-series cross validation (RTSCV) are:

1. Select the set of values for *X*. For comprehensiveness, the set of values of *X* should be every number from the chosen starting point up until the second last point in the dataset.
2. Perform time-series cross validation with each value of *X*. For each repeat, calculate the desired performance metrics for each forecast horizon (e.g. RMSE at $h = 1$, $h = 2$).
3. Summarize (e.g. average) the performance metrics for each forecast horizon across all repeats.

An example diagram of RTSCV is presented in Figure 4-33.

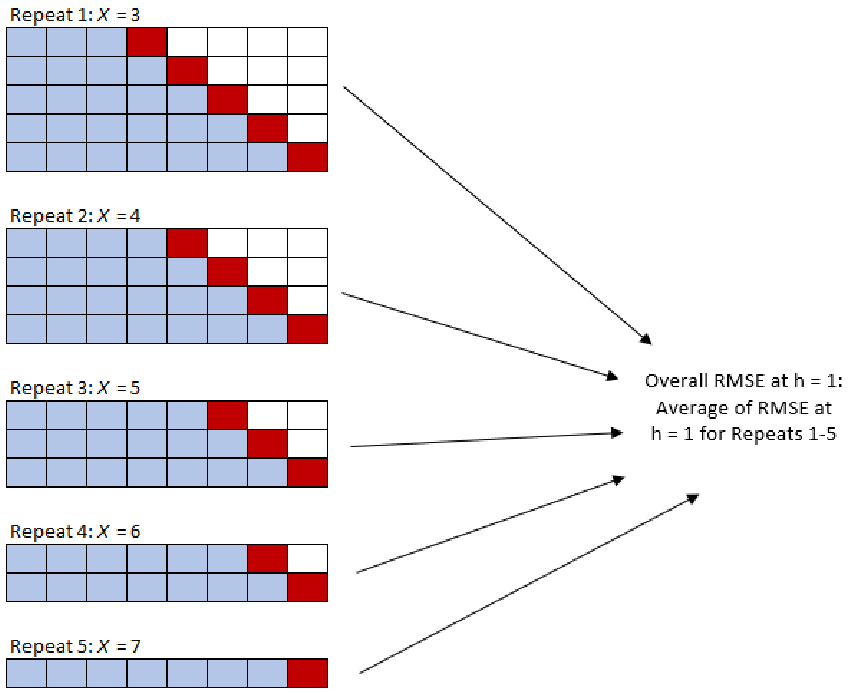


Figure 4-33 - Example Diagram of Repeated Time-Series Cross Validation (RTSCV)

As another example, if the set of X values were 200, 250 and 300, RTSCV would be implemented by performing time-series cross validation three times, one with X = 200, one with X = 250 and one with X = 300. Each repeat would contain RMSE calculations for each h-step forecast. The results would be averaged to produce overall RMSE values for each h-step forecast. For example, the final RMSE value for h = 1 would be:

$$RMSE_{h=1} = {(RMSE_{h=1}^{X=200} + RMSE_{h=1}^{X=250} + RMSE_{h=1}^{X=300})}/{3} \tag{4-2}$$

The train and test splits considered in a repeat of time-series cross validation are a subset of the ones considered in earlier repeats. Thus, the errors from models built on more training data make a greater contribution to the final performance metrics. This is more realistic since all available training data will likely be used to implement a chosen model in practice. However, RTSCV still considers the errors from models built on less training data, which is important to make sure that the results are generalizable. While RTSCV still requires a starting value of X to be chosen, its consideration of many possible values of X smooths out the results of individual runs of time-series cross validation and produces a more accurate picture of model performance. It is suggested that all reasonable values of X are used in RTSCV to ensure that the results are comprehensive. For the data in this analysis, a starting value of X = 200 is likely reasonable since the after rebrand dummy variable does not change until after the first 184 data points. Overall, RTSCV reduces the sensitivity of time-series cross validation to the number of data points used to fit the initial model and helps to capture changes in relative model performance as more data are collected. RTSCV is similar to repeated k-fold cross validation in concept, but different in that repeated k-fold cross validation uses the same data in each repeat, whereas the proposed technique increases the training dataset by one data point each time. This is also analogous to using different test period lengths to test forecast accuracy.

4.5.2 Empirical Evaluation of RSTCV

This research empirically tests RSTCV by using it to assess the performance of the bottom-up, ensemble and aggregate approaches to predicting total enquiries. The set of X values is chosen as the set of integers from 200 to 393 to produce a comprehensive assessment of model performance. As per the original analysis, each repeat of time-series cross validation calculates the RMSE for each h-step forecast from h = 1 to h = 30. The RMSE values for each repeat and forecast horizon are summarized by taking the average of them. The summarized results are presented in Table 4-20 and Figure 4-34. Table 4-20 contains the mean, median and sum of the averaged RMSE values across all forecast horizons. Figure 4-34 presents the averaged RMSE values at each forecast horizon. Table 4-20 focuses on addressing the question of which modelling approach is best across all forecast horizons, while Figure 4-34 shows which approach produces the most accurate predictions at each forecast horizon.

Forecast	**Mean**	**Median**	**Sum**
Aggregate Total	50.91	51.45	1527.18
Bottom Up	49.55	49.86	1486.59
Ensemble	49.20	49.57	1475.92

Table 4-20 - Repeated Time-Series Cross Validation (RTSCV) - Mean RMSE Performance Metrics for h = 1-30 (best shaded)

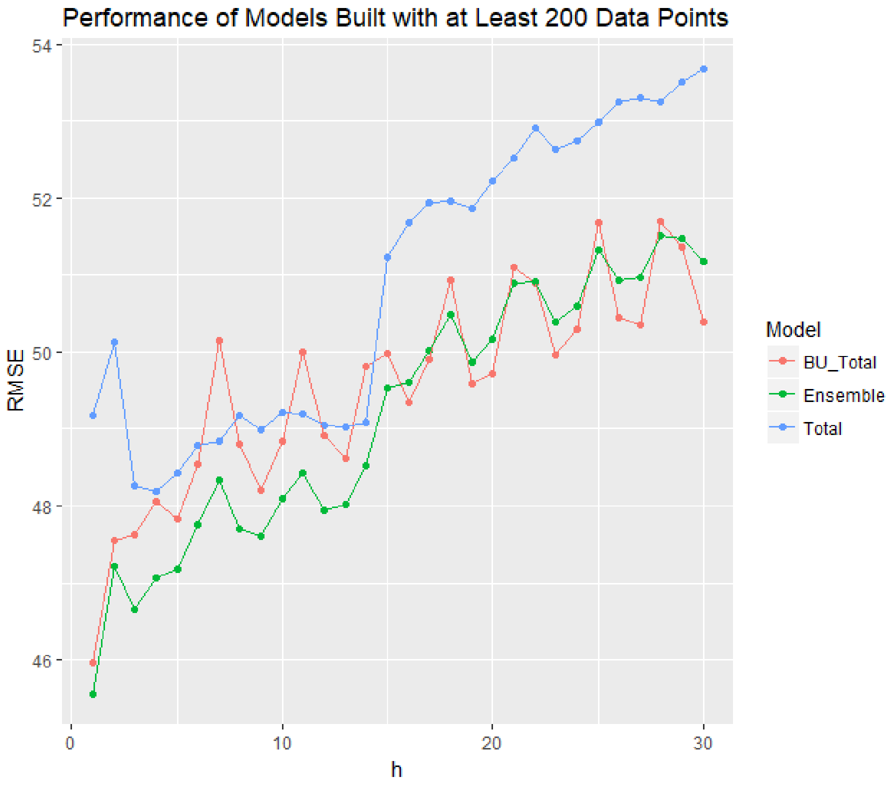


Figure 4-34 - RTSCV - Mean RMSE for h = 1-30

Table 4-20 suggests that the bottom-up and ensemble methods produce the most accurate forecasts when all forecast horizons are considered, with the ensemble method having slightly lower RMSE values on average. Figure 4-34 reveals that the ensemble approach outperforms from h = 1 to h = 15, while the bottom-up becomes more accurate in the later h-step forecasts. Both methods clearly outperform the aggregate modelling approach. These results match up with the findings from the original analysis, which suggest that the ensemble and bottom-up approaches are likely to be the most accurate for future predictions, particularly as more data are collected and used in the model estimation process. However, this trend was identified by one run of RTSCV, as opposed to the several manual runs of normal time-series cross validation that were required in the original analysis. Thus, RTSCV is able to identify the modelling approaches likely to produce the most accurate forecasts as more data are collected, while still considering a large number of errors. The proposed technique offers a more comprehensive way to assess time-series model performance than normal time-series cross validation. It also reduces its sensitivity to the number of data points used to fit the initial model, which can cause misleading results if left unaddressed. While the results of this analysis suggest the most accurate modelling approaches moving forward are the ones that were most accurate towards the end of the dataset, this may not always be the case. Considering all reasonable

X values averages out the results of individual time-series cross validation runs. This helps provide a balance between how model performance changes as more data are collected and the need to have enough errors to accurately assess model performance.

Another advantage of RTSCV is the ability to drill down on the individual time-series cross validation repeats and see how the individual modelling approaches' predictive performances change as *X* increases. This can be achieved by summarizing the RMSE values for each modelling approach in a particular time-series cross validation run and seeing how they compare to each other. As an example, the performance of the individual repeats in the RTSCV results is presented in Figure 4-35. For each time-series cross validation repeat, or value of *X*, a dot point is made for the modelling approach with the lowest sum of RMSE values for $h = 1$ to $h = 30$. Thus, Figure 4-35 shows which modelling approach is best across all forecast horizons for each repeat. As can be seen, there are two changes in the modelling approach with the lowest total RMSE across all forecast horizons, one from the aggregate to the ensemble approach at $X = 251$ and one from the ensemble to the bottom-up approach at $X = 287$. These clear switches provide insight into when individual modelling approaches start outperforming their counterparts. The outperformance of the bottom-up approach as only models with more data are included suggests that it will overtake the ensemble approach as the most accurate overall approach in the future. If required, Figure 4-35 could also be adjusted to provide insight into a particular forecast horizon, such as $h = 1$, rather than considering all forecast horizons.

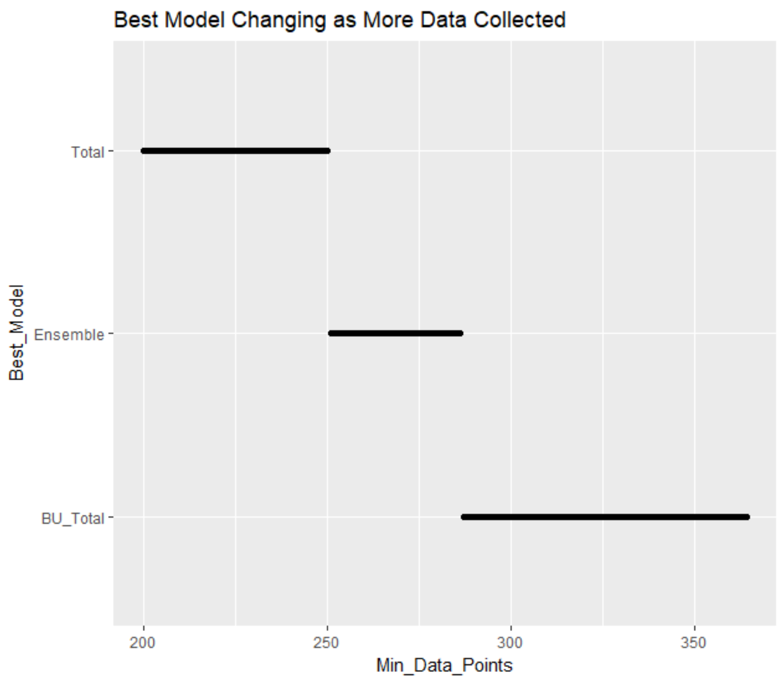


Figure 4-35 - RTSCV - Mean RMSE – Time-Series Cross Validation Repeats

Investigating which modelling approach produces the most accurate forecasts on an individual repeat level offers insight into how the most accurate approach changes over repeats. For example, Figure 4-35 shows that the aggregate modelling approach outperformed the ensemble and bottom-up approaches when models built on less data were included in time-series cross validation. This could potentially provide early warning of when a particular model is starting to outperform its counterparts, before it is reflected in the overall RTSCV results. It could also be used to create generalizations of the number of data points needed for a particular model to become relatively effective at forecasting.

To further investigate the RTSCV technique, two variations of it were examined. First, the RMSE values for each repeat and *h*-step forecast were summarized with the median instead of the mean. Using the median instead of the mean when summarizing performance across repeats could further reduce the influence of individual repeats, potentially producing a more accurate performance assessment. Table 4-21 and Figure 4-36 showcase the results of this variation. Figure 4-36 shows that the ensemble model produces the most accurate forecasts across the early *h*-step forecasts, while the bottom up approach is more accurate in the later ones. Table 4-21 once again shows the bottom up and ensemble models are close on an overall basis, with the bottom up approach this time producing the best performance. Summarizing the RMSE values with the median seems to accentuate the results of the mean variation, with Figure 4-36 in particular showing a sharper distinction between the modelling approaches. Since these results align with the findings from the original analysis, this suggests that using the median to summarize performance across repeats may further reduce the noise of individual repeats and help highlight true model performance.

Forecast	Mean	Median	Sum
Aggregate Total	54.15	55.11	1624.65
Bottom Up	52.11	52.17	1563.38
Ensemble	52.31	52.52	1569.22

Table 4-21 - RTSCV - Median RMSE Performance Metrics for h = 1-30 (best shaded)

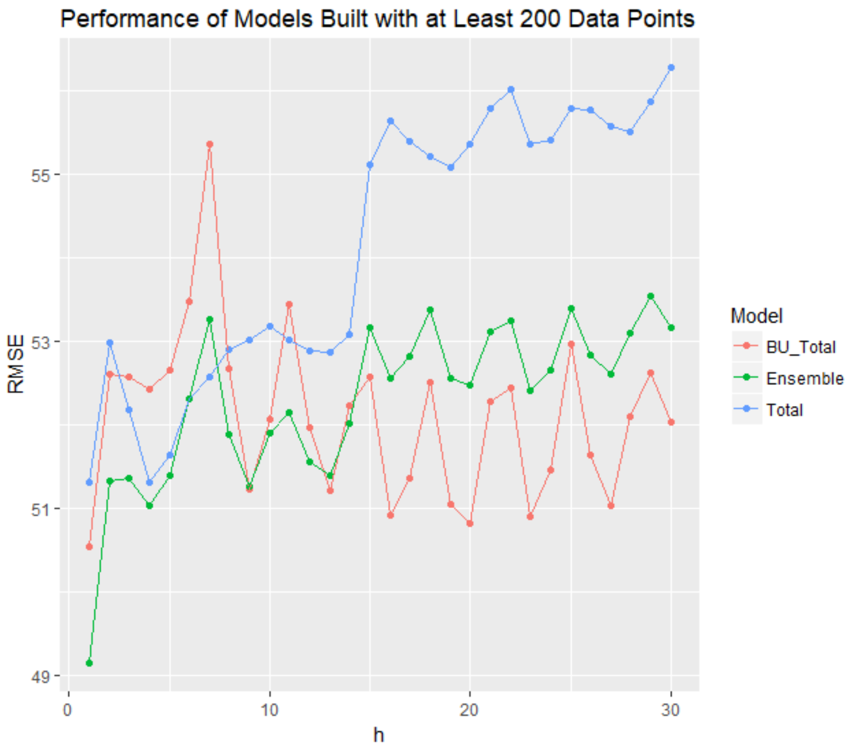


Figure 4-36 - RTSCV - Median RMSE for h = 1-30

The second variation to RTSCV that was examined was varying the performance metric itself. Instead of calculating the RMSE at each *h*-step forecast in each time-series cross validation repeat, the MAE was calculated. The MAE values for each forecast horizon were summarized by taking the average of them. The summarized results are presented in Table 4-22 and Figure 4-37. As can be seen, the MAE variation has similar results to the RMSE variation, with the ensemble approach generally outperforming in the early forecast horizons before being over taken by the bottom up approach in the later forecast horizons. This shows that RTSCV is generalizable and can be applied to multiple performance metrics.

Forecast	Mean	Median	Sum
Aggregate Total	37.58	38.51	1127.45
Bottom Up	35.60	35.92	1068.09
Ensemble	35.56	36.22	1066.77

Table 4-22 - RTSCV - Mean MAE Performance Metrics for h = 1-30 (best shaded)

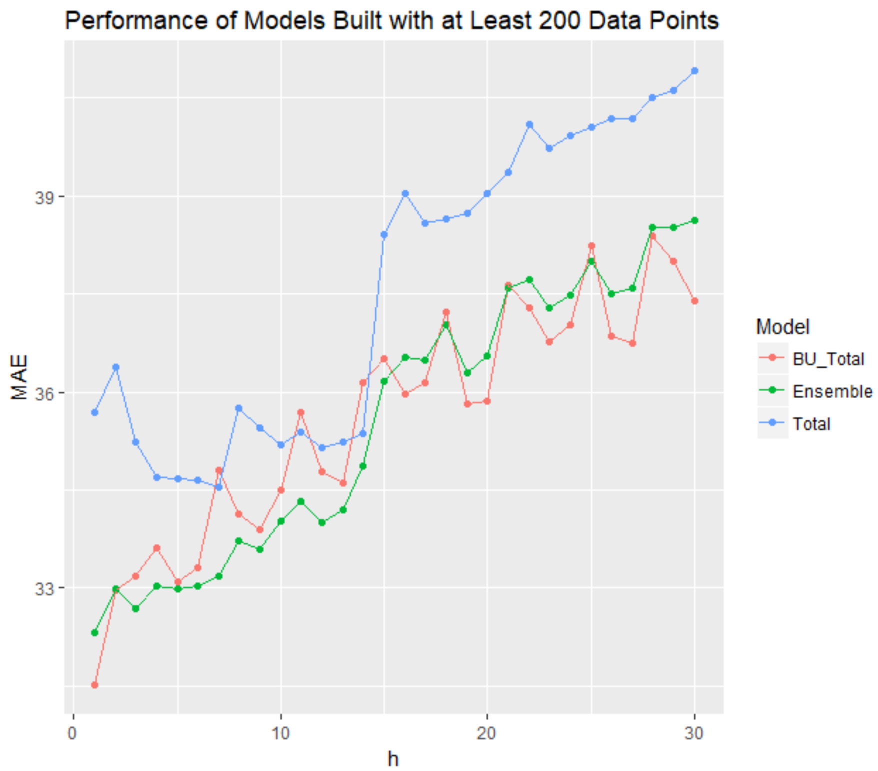


Figure 4-37 - RTSCV - Mean MAE for h = 1-30

4.6 Key Findings & Conclusion

Previous research has produced conflicting results regarding the inclusion of exogenous predictors in time-series models. Some research shows that exogenous variables can improve a model's accuracy by incorporating information not included in the history of the time series being predicted (Taieb and Hyndman 2014). However, other research finds that pure time-series approaches produce more accurate forecasts than ones with exogenous variables (Athanasopoulos et al. 2011). Against this backdrop, Makridakis et al. (2020) call for more research into how exogenous data can improve forecasting performance. We address this call for further research by investigating how including advertising expenditure and calendar-based variables in time-series models can improve their predictive accuracy across different forecast horizons. This chapter also addresses gaps in the literature surrounding the effectiveness of different approaches to generating coherent demand forecasts. In particular, we examine how the accuracy of bottom-up and aggregate approaches to forecasting total demand changes as the forecast horizon and amount of data used in model estimation increases. This provides insights that can help companies generate more accurate demand forecasts, which can improve profitability and productivity.

The first major finding of this chapter is that relatively simple time-series models, augmented with exogenous variables, can outperform more complex pure time-series models. Eight candidate models were considered when choosing the best-performing model for online, offline and total enquiries. For each time series, an ARIMA model with advertising spend and calendar-based predictors produces more accurate forecasts than more complex methods, such as TBATS, NNETAR and FFORMA models. This finding holds across forecast horizons up to thirty periods into the future, with the differences in model accuracy becoming accentuated as the forecast horizon lengthens. More complex models are often more challenging to implement in practice, requiring additional infrastructure and expertise to set up and maintain. Thus, such a finding is useful for practitioners as they may be better off or able to obtain relatively high accuracy with simpler techniques and a few well-chosen exogenous variables. Rather than making models more complicated to obtain higher predictive accuracy, practitioners would do well to consider what extra information can be incorporated into them via exogenous variables. This is particularly important if longer-term forecasts are required, such as forecasts for the next thirty days, which is often the case for inventory planning. While the results may change as more data becomes available to use in model estimation, it is important to note that long time series are often difficult to obtain in practice. Additionally, there are often periodic structural changes in the underlying time series that make it difficult to train a model using the full history of a dataset.

When choosing exogenous variables, it is important to select ones that can be accurately forecast and that have strong relationships with the time series being forecasted (Makridakis et al. 2020). This research finds that advertising expenditure and calendar-based exogenous variables can considerably improve demand forecasts across different forecast horizons. Advertising spend and calendar-based predictors work well as they are relatively easy to collect, can be known in advance and have clear relationships with demand that tend to persist over time. Accounting for calendar-based effects, such as weekends, public holidays and company rebrands, has a greater effect on forecast accuracy than the inclusion of advertising expenditure. This is particularly true when the demand forecasts are made for a brick and mortar location, which are more sensitive to calendar-based effects. For example, the difference between models with and without exogenous variables is greater for offline than online enquiries as offline enquiries are made to a call centre, which is closed on weekends and public holidays. Interestingly, while including advertising expenditure helps improve demand forecasts, breaking total spend into more granular amounts is not always necessary. Only including total advertising expenditure results in more accurate forecasts for offline and total enquiries, while breaking it into online and offline advertising spend works better for online enquiries. This is likely because online advertising has a strong influence on online enquiries. Thus, for some time series, breaking advertising expenditure into more granular amounts produces more accurate forecasts, particularly when the time series has a strong relationship with a granular component of the total. Finally, the results suggest that refitting the time-series models for online, offline and total enquiries daily as more data are collected will produce more accurate daily forecasts than only re-estimating the models periodically.

Another key finding of this research is that different time-series models are better suited to different forecast horizons. The best-performing models for online, offline and total enquiries were used to produce bottom-up, aggregate and ensemble forecasts of total enquiries. The results suggest that the ensemble approach is more accurate for short-term forecasts, while the bottom-up approach outperforms at longer forecast horizons. Thus, it is important for companies to consider model performance at different forecast horizons, rather than assuming a one-size fits all approach. This result indicates that more complex modelling approaches tend to perform better when making short-term predictions, while more simple ones tend to perform relatively better when making longer-term predictions. Such a conclusion is also supported by the accentuated outperformance of the ARIMA models at later forecast horizons in the individual time-series modelling. This is likely because more complex models detect relationships in the data that persist in the short-term, but break down as the forecast horizon lengthens.

The results also show that as more data are collected and used in the model estimation process, the relative accuracy of different modelling approaches can change. More specifically, when a company only has a small training dataset available, forecasting the top-level time series directly will produce more accurate forecasts of total demand. However, as more data becomes available for model training, higher predictive accuracy will be obtained by adopting a bottom-up or ensemble approach to demand forecasting. This conclusion is consistent across all forecast horizons tested in the analysis. When there is less data available, there is not enough information available to accurately capture the relationships in each bottom-level time series. However, as more data are collected, the bottom-level time-series models become more accurate and the aggregation bias in the top-level data becomes more apparent. Thus, the performance of a model can be contingent on the amount of data used to fit it. In light of this, it is important for companies to repeat the model fitting process as more data becomes available for modelling so that any changes in relative model performance can be captured. Additionally, when performing bottom-up forecasting, it is important to choose the best model for each bottom-level time series separately, rather than applying the same model to all of them. While ARIMA models were found to be the most accurate models for both online and offline enquiries in this research, the exogenous variables included in them were different for each time series. Selecting the optimal model for each time series separately will help ensure that the most accurate predictions are produced.

The final major finding of this research is that the results of time-series cross validation are sensitive to the number of observations used to fit the initial model (X). In some cases, this can lead to misleading assessments of model performance, with models that perform well on smaller datasets appearing to outperform models that become better as more data are collected. To address this issue, an extension to time-series cross validation was developed. Analogous to repeated k-fold cross validation, repeated time-series cross validation (RTSCV) involves repeating normal time-series cross validation with different values of X and averaging the results. This helps to smooth out the results of individual repeats and places greater weight on errors from models built with more data, which are more likely to be reflective of future model performance. However, the technique still considers errors from a large number of train and test splits, which is important for obtaining a reliable estimate of model performance.

Overall, this research finds that incorporating advertising and calendar-based predictors into time-series models can improve demand forecasts. While practitioners often tend to focus on developing more complex models only using past values from the time series being predicted, it can be more valuable to consider a simpler approach augmented by exogenous variables. When using a bottom-up approach to forecast total demand, practitioners should adopt a customized approach by performing model selection on each bottom-level time series. Additionally, if predictive accuracy is paramount, forecasters should select an appropriate model for each *h*-step forecast, rather than adopting a one-size fits all approach. Managers should also consider that the relative accuracy of time-series models can change as more data are collected and becomes available for model training. Using repeated time-series cross validation to evaluate models can help address this issue by increasing the weight on errors from models built with more data, while still using a large number of errors to develop a robust assessment of model performance.

4.7 Future Research

There are several opportunities for future research based on the findings of this chapter. First, the analysis could be repeated across a large number of time series, such as the M4 forecasting competition dataset[6], to develop empirical generalizations. For example, how large should a dataset be before there is a switch to a bottom-up or ensemble forecasting approach, and does this vary across forecast horizons. These could help develop rules of thumb for companies to use when creating demand forecasts. Second, additional work could be done to improve the accuracy of the point forecasts. In the hierarchical modelling stage, the state of the art optimal reconciliation approach outlined by Wickramasuriya et al. (2019) could be introduced as another forecasting approach. In the individual time-series modelling stage, the FFORMA model could be adapted to allow for exogenous variables. Given the strong performance of this model in the M4 forecasting competition (Montero-Manso et al. 2020), the addition of exogenous variables is likely to improve its forecasting performance. Investigating the inclusion of lagged advertising effects or AdStock variables capturing historical advertising expenditure, rather than just current advertising spend, may also improve model accuracy. Another way to improve point forecasts could be through the use of bagging techniques, which Bergmeir et al. (2016) show can produce more accurate forecasts when applied to ETS models. Additionally, other neural network model variations, such as LSTM-NNET, could be examined to see they can improve predictive accuracy. Finally, while this analysis has only focused on point forecasts, future research could seek to calculate prediction intervals for the models and examine their width and accuracy. It is often

[6] See https://forecasters.org/resources/time-series-data/m4-competition/ for more information about the M4 forecasting competition.

helpful for businesses to have prediction intervals for demand forecasts to understand the range of values in which demand might actually fall. Thus, investigating a modelling approach's ability to produce accurate and narrow prediction intervals, as well as accurate point forecasts, could be a fruitful avenue for future research.

Chapter 5 Conclusion

Advertising is an essential business function, with organisations investing considerable resources to incentivize consumers to purchase their goods and services. The proliferation of advertising media channels and the devices with which consumers can interact with them has made measuring advertising effectiveness increasingly complex. At the same time, because of the sheer amount of content and competition in the marketplace, measuring the effectiveness of a firm's media channels and optimizing the resources invested in them has become increasingly important. This process can be conducted in many ways, from using simple heuristics, such as last touch, to more complex probabilistic and econometrics techniques, such as VAR models. Models that quantify the effectiveness of a company's advertising media channels can also provide insights into how a company should allocate their financial and human resources to these media channels. The importance of developing models that quantify advertising media effectiveness has been highlighted by the Marketing Science Institute (2018), which lists measuring media efficacy as a research priority for 2018 - 2020. This chapter presents the main conclusions and contributions of this research, along with suggestions for future work.

5.1 Conclusions of this Research

This research leverages consumer and firm activity data to answer important business questions and improve strategic decision-making. More specifically, this research advances the development of statistical models that companies can use to measure advertising media effectiveness. The findings of this research shed light on the data marketing managers should be collecting when developing these models. This research also provides a methodology for connecting the outputs of models that quantify advertising media effectiveness with practical marketing calculations, such as media channel ROI and budget allocation recommendations. The data-driven approach is generalisable to organisations of varying sizes and media channel mixes. It can also accommodate different types of marketing performance metrics, such as sales and enquiries. Altogether, these developments will help researchers and industry professionals better develop and translate advertising media effectiveness models into practice. The outputs from these models can produce insights that enable organisations to better understand the consumer path to purchase and improve the allocation of their marketing resources. For example, visualising the relationships between advertising expenditure, consumer activity metrics and marketing performance metrics in network diagrams can assist marketing managers in communicating advertising results to a non-technical audience. Additionally, the use of marketing analytics to measure advertising media effectiveness increases transparency, visibility and accountability over marketing resources.

This research also shows how advertising expenditure data can be used to improve demand forecasts. Such an application advances the demand forecasting literature and demonstrates the benefits of data-sharing between business functions in an organisation. Furthermore, it highlights how marketing departments and media agencies can harness their data to provide additional value to their clients by using it to address other business problems. Finally, the findings from this research may be applicable to other fields that have problems that involve assessing the effectiveness of multiple factors with carryover and interaction effects, such as healthcare and team sports analytics. The following sections discuss the key findings of this research according to the four main research questions. More detailed references substantiating the key findings are provided in Chapters 2-4.

5.1.1 Research Question 1 (RQ1)

RQ1: *What is the state of the field of research in quantifying advertising media effectiveness over the past decade?*

This research question was addressed by conducting a systematic literature review of research in quantifying advertising media effectiveness over the past decade. The review reveals that approaches to quantifying advertising media effectiveness can be broken down into four stages: data collection, model development, model assessment and resource allocation. Data in advertising media effectiveness models are collected at an individual or aggregate level and can be grouped into five main categories: company information, market conditions, marketing activities, customer information and competitor information. Using individual-level data enables researchers to gain more granular insights and a more detailed understanding of advertising effects. However, it is generally harder to collect and maintain individual-level data due to issues such as accuracy, cost and privacy concerns. Aggregate data are more readily available and accessible for all types of businesses, but do not offer the granularity and potential for personalized targeting that individual-level data do. Thus, aggregate data may be more useful for strategic problems, such as budget allocation between media channels, while individual-level data may be more relevant for tactical decisions, such as customizing a website according to individual consumer characteristics.

Model development approaches in the literature can be categorised according to whether individual-level data or aggregate data are used. Aggregate modelling approaches include persistence modelling and customized modelling, while individual-level modelling approaches include single source modelling and attribution modelling. The strengths and weaknesses of the various modelling approaches are summarized in Table 2-6 (see page 36). The literature review also highlights key findings in Table 2-3 (see page 31) and presents future research recommendations in Table 2-4 (see page 33). The following are the most notable findings and recommendations.

- While heuristic models are widely used in industry due to their simplicity and ease of use, more data-driven advertising media effectiveness models can provide marketers with a more accurate understanding of the effectiveness of their media channels. Additionally, applying the results of advertising media effectiveness models has the potential to improve a company's financial performance (de Haan et al. 2016; Kireyev et al. 2016; Danaher and Dagger 2013; Wiesel et al. 2011; Danaher and Van Heerde 2018; Geyik et al. 2014).
- Advertising media effectiveness models can be used to make general or specific media budget recommendations. Marketing managers do not often implement specific types of optimization solutions, preferring to use interactive tools, such as dashboards, that provide them with a range of options and help them better understand the models behind the numbers.
- When developing models, researchers would benefit from considering all the media channels a company uses, not just the ones that fall under the paid media category.
- Researchers in this field need to take advantage of and incorporate advertising metadata (e.g. position on the website) and consumer activity data (e.g. Facebook likes) into their models. This data can help models create a more accurate picture of the consumer path to purchase and the relationships between media channels.

- Individual-level and aggregate modelling approaches to quantifying advertising media effectiveness are rarely discussed together or compared. These approaches essentially address the same problem from different perspectives, with aggregate modelling approaches coming from a top-down perspective and individual-level modelling approaches coming from a bottom-up perspective.
- Consequently, a key area for future research is reconciling newer, individual-level modelling approaches, such as attribution modelling, with more traditional aggregate modelling approaches, such as persistence modelling. Given their ability to incorporate multiple information sources, dynamic models that use Bayesian techniques may be a useful starting point for this research.

5.1.2 Research Question 2 (RQ2)

RQ2: *What effects do consumer activity metrics and the level of aggregation in marketing performance metrics have on ROI, CPA and budget allocation calculations?*

This research finds that including consumer activity metrics when modelling the effectiveness of media channels enables the indirect relationships between paid media channels and marketing performance metrics to be captured. Excluding these relationships can reduce the accuracy of ROI, CPA and budget allocation calculations. Additionally, disaggregating marketing performance metrics results in the detection of relationships that do not appear when the data are aggregated. This is an example of hierarchical aggregation losing some of the information contained in the underlying data (Hyndman and Athanasopoulos 2018). The research undertaken in addressing RQ2 also yielded the following findings.

- Paid media channels can directly influence both marketing performance metrics and consumer activity metrics, which in turn can directly influence marketing performance metrics.
- Effects of online media channels tend to last for longer than offline media channels. While the effects of online (offline) media channels tend to be larger on online (offline) enquiries, online (offline) media channels do still influence offline (online) enquiries.

- Confirmation of the findings of Srinivasan et al. (2016) in an expanded context with more media channels, a high involvement product and daily data.
- Reinforcing the findings of Dinner et al. (2014), increasing paid search expenditure has a small negative effect on paid search engagement rate. Strategically, marketing managers should be careful not to allocate too much of the marketing budget to paid search advertising. Tactically, media buyers should focus on bidding for relevant and specific keywords, rather than a broad set that potentially generates less engagement.
- This research is the first to include both paid and owned Facebook advertising in a persistence model. The results show that owned media channels can influence other owned media channels, as well as marketing performance metrics. Additionally, the effects of paid and owned media channels on marketing performance metrics differ, with the immediate effects of owned media channels tending to be smaller, but more long-lasting, than those of paid media channels.
- Representing customer-initiated paid media channels, such as paid search, with both a consumer activity metric and firm activity metric in persistence models can help separate the effects of increases in advertising expenditure and consumer engagement on marketing performance. Both paid search advertising expenditure and engagement rate have positive effects on offline enquiries and website page views. Interestingly, the effects of increasing paid search expenditure were larger in the short-term, but the effects of increasing the paid search engagement rate lasted for longer and were larger overall. This highlights that simply increasing advertising expenditure is not necessarily the best way to improve a media channel's effectiveness. It also shows that leveraging different features of a media channel can have different effects on marketing performance metrics.

5.1.3 Research Question 3 (RQ3)

RQ3: *To what extent can advertising spend and calendar-based indicator variables improve demand forecasts across different forecast horizons?*

The third aim of this research is to examine how data used in advertising media effectiveness models can be applied in a different context to help address non-marketing business problems. More specifically, this research investigates how advertising data can be used to improve demand forecasts, which are used by organisations in operational planning. The results show that incorporating advertising and calendar-based predictors into time-series models can considerably improve demand forecasts across forecast horizons up to 30 periods (days) in the future. When predicting a travel brand's daily online, offline and total enquiries, ARIMA models with advertising spend and calendar-based predictors produce more accurate forecasts than more complex methods, such as TBATS, NNETAR and FFORMA models. This suggests that using relatively simple models, augmented with exogenous variables, can produce more accurate forecasts than more complex pure time-series models. Since more complex models are often more difficult to implement in practice, this finding is useful for practitioners as they may be able to obtain relatively high accuracy with simpler techniques and a few well-chosen exogenous variables. Addressing RQ3 also produced the following additional findings.

- Accounting for calendar-based effects, such as weekends, public holidays and company rebrands, has a greater effect on forecast accuracy than the inclusion of advertising expenditure. This is most prevalent in demand forecasts that are dependent on human involvement, such as the travel brand's offline enquiries, which are taken via phone calls to a call centre.
- Breaking advertising expenditure into more granular amounts does not always produce the most accurate forecasts. For offline and total enquiries, more accurate forecasts are produced by models that only include total advertising expenditure as an exogenous predictor. On the other hand, online enquiries models that break total advertising expenditure into online and offline components generate more accurate forecasts than ones that keep advertising expenditure aggregated. This is likely because online advertising has a strong influence on online enquiries. Therefore, breaking advertising expenditure into more granular components may produce more accurate forecasts when the time series has a strong relationship with a particular component.

5.1.4 Research Question 4 (RQ4)

RQ4: *How does the performance of bottom-up and aggregate approaches to demand forecasting change across forecast horizons and data size?*

This research finds that different time-series models are better suited to different forecast horizons. Therefore, if predictive accuracy is important, forecasters should select an appropriate model for each *h*-step forecast, rather than adopting a one-size fits all approach. More generally, the results indicate that more complex modelling approaches tend to perform better when making short-term predictions, while simpler ones tend to perform relatively better when making longer-term predictions. The results also show that as more data are collected and used in the model estimation process, the relative accuracy of different modelling approaches can change. When there is less data available, forecasting the top-level time series directly produces more accurate forecasts of total demand, while a bottom-up or ensemble approach becomes more effective as more data are collected. Additionally, practitioners using a bottom-up approach to forecast total demand should consider different models for each bottom-level time series, rather than using the same model type for each time series.

Addressing RQ4 also revealed that the results of time-series cross validation are sensitive to the number of observations used to fit the initial model. This can lead to misleading assessments of model performance, with models that perform well on smaller datasets appearing to outperform models that become better as more data are collected. In light of this issue, a new model evaluation technique was developed and empirically evaluated. Repeated time-series cross validation (RTSCV) involves repeating normal time-series cross validation with different data sizes for the initial model and averaging the results. Repeated time-series cross validation provides a more comprehensive way to assess time-series model performance by smoothing out the results of individual runs of normal time-series cross validation. It places greater weight on models built with more data, which are more likely to be indicative of future predictive performance. However, it still considers a large number of errors, which are needed to develop a robust assessment of model performance. Overall, repeated time-series cross validation is able to identify the modelling approaches likely to produce the most accurate forecasts as more data are collected.

5.2 Future Research

This section discusses three key opportunities for future research based on the findings of this research. Suggestions for future research have also been made in other sections of the dissertation. First, incorporating prior customer information (e.g. attitudinal metrics) into advertising media effectiveness models is an interesting area of future research. Companies usually have access to data regarding their target audience through mechanisms such as loyalty programs and market research surveys. These data are often used to design marketing strategies, but are not always included in models that assess the effectiveness of media channels (Hanssens and Pauwels 2016). A Bayesian model, such as a Bayesian Vector Autoregression (BVAR), could assist with this by incorporating customer information into its prior distribution. Such research would help marketing managers replicate their decision-making processes in statistical form and leverage existing datasets, such as customer loyalty programs, in the assessment of advertising media effectiveness. This could lead to more accurate and granular assessments of media channel effectiveness, along with the potential to customize marketing budget allocations to different customer segments.

Second, the implementation of a model that quantifies advertising media effectiveness is only the beginning for an organisation. There is relatively little research around maintaining and extending these models so that marketing managers can identify and take advantage of changes in the environment. This is an important issue for practitioners, who need to keep updating and using advertising media effectiveness models into the future. For example, since the marketing environment changes over time, model parameters will need to be periodically updated. This is particularly true if an organisation uses the results of an advertising effectiveness model to adjust their behaviour. Investigating when and how best to incorporate this into existing models, such as using rolling window regressions or Markov regime switching techniques, would be a useful avenue for future research. The results of field experiments could also be used to develop more refined marketing campaign strategies based on calculations of media channel effects. For example, one could investigate the amount and timing of advertising expenditure required to achieve a certain shock to sales.

Finally, further work could be done to expand the use of marketing data in other business functions. For example, accounting for historical advertising expenditure with lagged advertising spend or AdStock variables could improve the accuracy of demand forecasting models. Additionally, investigating the inclusion of consumer activity metrics in these models may result in further gains in predictive performance. Consumer activity metrics could also be used to provide early warning indicators of demand levels, which may assist with inventory planning. Greater data-sharing between business functions will help ensure that there is consistency in planning and execution across the business, as well as benefit the development of advertising media effectiveness models. For example, the data provider for this research, Rapid Media, did not have access to a financial performance metric for the travel brand analysed in this research. Adding a financial performance metric, such as sales, to the advertising media effectiveness model would identify whether there are any differences between the media channels driving sales and enquiries. Such insights could potentially help the travel brand adjust their advertising so that more enquiries convert into sales, thereby resulting in greater revenue for the business.

Overall, this dissertation provides insights into how consumer and firm activity data can be leveraged in multiple contexts to improve strategic decision-making. More specifically, this research advances the development of statistical models that companies can use to measure advertising media effectiveness. Additionally, it provides tools to help industry practitioners better understand the effects of advertising and optimize marketing resources. Finally, it shows how advertising and calendar-based variables can be used to improve demand forecasts, which can result in improved managerial decision-making across a wide variety of business functions and industries.

Bibliography

Abhishek, V., P. Fader, and K. Hosanagar. 2012. Media Exposure through the Funnel: A Model of Multi-Stage Attribution: Heinz College Research.

Albers, S. 2012. Optimizable and implementable aggregate response modeling for marketing decision support. *International Journal of Research in Marketing* 29 (2):111-122.

Anderl, E., I. Becker, F. von Wangenheim, and J. H. Schumann. 2016. Mapping the customer journey: Lessons learned from graph-based online attribution modeling. *International Journal of Research in Marketing* 33 (3):457-474.

Aria, M., and C. Cuccurullo. 2017. bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics* 11 (4):959-975.

Athanasopoulos, G., R. J. Hyndman, H. Song, and D. C. Wu. 2011. The tourism forecasting competition. *International Journal of Forecasting* 27 (3):822-844.

Batra, R., and K. L. Keller. 2016. Integrating Marketing Communications: New Findings, New Lessons, and New Ideas. *Journal of Marketing* 80 (6):122-145.

Bergmeir, C., Rob J. Hyndman, and José M. Benítez. 2016. Bagging exponential smoothing methods using STL decomposition and Box–Cox transformation. *International Journal of Forecasting* 32 (2):303-312.

Bergmeir, C., R. J. Hyndman, and B. Koo. 2018. A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis* 120:70-83.

Bollinger, B., M. A. Cohen, and L. Jiang. 2013. Measuring Asymmetric Persistence and Interaction Effects of Media Exposures Across Platforms: University of British Columbia.

Box, G. E. P., G. M. Jenkins, G. C. Reinsel, and G. M. Ljung. 2015. *Time Series Analysis: Forecasting and Control* 2nd ed. Hoboken, New Jersey: John Wiley & Sons.

Brodersen, K. H., F. Gallusser, J. Koehler, N. Remy, and S. L. Scott. 2015. Inferring causal impact using Bayesian structural time-series models. *Annals of Applied Statistics* 9:247-274.

Chan, D., and M. Perry. 2017. Challenges and Opportunities in Media Mix Modeling: Google Inc.

Clemen, R. T. 1989. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting* 5 (4):559-583.

Cleveland, R. B., W. S. Cleveland, and I. Terpenning. 1990. STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics* 6 (1):3-33.

Dalessandro, B., C. Perlich, O. Stitelman, and F. Provost. 2012. Causally motivated attribution for online advertising. In *Proceedings of the Sixth International Workshop on Data Mining for Online Advertising and Internet Economy*. Beijing, China: ACM, 1-9.

Danaher, P., J., and J. Rossiter, R. . 2011. Comparing perceptions of marketing communication channels. *European Journal of Marketing* 45 (1/2):6-42.

Danaher, P. J. 2017. Advertising Effectiveness and Media Exposure. In *Handbook of Marketing Decision Models*, edited by B. Wierenga and R. van der Lans. Cham: Springer International Publishing, 463-481.

Danaher, P. J., and T. S. Dagger. 2013. Comparing the Relative Effectiveness of Advertising Channels: A Case Study of a Multimedia Blitz Campaign. *Journal of Marketing Research* 50 (4):517-534.

Danaher, P. J., and H. J. Van Heerde. 2018. Delusion in Attribution: Caveats in Using Attribution for Multimedia Budget Allocation. *Journal of Marketing Research* 55 (5):667-685.

de Haan, E., T. Wiesel, and K. Pauwels. 2016. The effectiveness of different forms of online advertising for purchase conversion in a multiple-channel attribution framework. *International Journal of Research in Marketing* 33 (3):491-507.

De Livera, A. M., R. J. Hyndman, and R. D. Snyder. 2011. Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing. *Journal of the American Statistical Association* 106 (496):1513-1527.

Dekimpe, M. G., and D. M. Hanssens. 1995. The Persistence of Marketing Effects on Sales. *Marketing Science* 14 (1):1-21.

———. 1999. Sustained Spending and Persistent Response: A New Look at Long-Term Marketing Profitability. *Journal of Marketing Research* 36 (4):397-412.

Dinner, I. M., H. J. Van Heerde, and S. A. Neslin. 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.

Dorfman, R., and P. O. Steiner. 1954. Optimal Advertising and Optimal Quality. *The American Economic Review* 44 (5):826-836.

Farris, P. W., D. M. Hanssens, J. D. Lenskold, and D. J. Reibstein. 2015. Marketing return on investment: Seeking clarity for concept and measurement. *Applied Marketing Analytics* 1 (3):267-282.

Feit, E. M., P. Wang, E. T. Bradlow, and P. S. Fader. 2013. Fusing Aggregate and Disaggregate Data with an Application to Multiplatform Media Consumption. *Journal of Marketing Research* 50 (3):348-364.

Fildes, R., S. Ma, and S. Kolassa. 2019. Retail forecasting: Research and practice. *International Journal of Forecasting* In Press.

Fischer, M. 2019. Managing advertising campaigns for new product launches: An application at mercedes-benz. *Marketing Science* 38 (2):343-359.

Franses, P. H. 2005. On the Use of Econometric Models for Policy Simulation in Marketing. *Journal of Marketing Research* 42 (1):4-14.

Frison, S., M. G. Dekimpe, C. Croux, and P. De Maeyer. 2014. Billboard and cinema advertising: Missed opportunity or spoiled arms? *International Journal of Research in Marketing* 31 (4):425-433.

Fuller, W. A. 1996. *Introduction to Statistical Time Series*. 2nd ed. New York: John Wiley and Sons.

Gallego, V., P. Suarez-Garcia, P. Angulo, and D. Gomez-Ullate. 2019. Assessing the effect of advertising expenditures upon sales: A Bayesian structural time series model. *Applied Stochastic Models in Business and Industry* 35 (3):479-491.

Geyik, S. C., A. Saxena, and A. Dasdan. 2014. Multi-Touch Attribution Based Budget Allocation in Online Advertising. In *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising*. New York, NY, USA: ACM, 1-9.

Ghalehkhondabi, I., E. Ardjmand, A. Young William, and R. Weckman Gary. 2019. A review of demand forecasting models and methodological developments within tourism and passenger transportation industry. *Journal of Tourism Futures* 5 (1):75-93.

Godfrey, A., K. Seiders, and G. B. Voss. 2011. Enough Is Enough! The Fine Line in Executing Multichannel Relational Communication. *Journal of Marketing* 75 (4):94-109.

Guerrero, V. M. 1993. Time-series analysis supported by power transformations. *Journal of Forecasting* 12 (1):37-48.

Handley, L. *Global ad spend has slowed but 2020 looks set to be a bumper year* 2019 [cited June 2020. Available from https://www.cnbc.com/2019/10/24/global-ad-spend-has-slowed-but-2020-looks-set-to-be-a-bumper-year.html.

Hanssens, D. M., and M. G. Dekimpe. 2017. Models for the Financial-Performance Effects of Marketing. In *Handbook of Marketing Decision Models*, edited by B. Wierenga and R. van der Lans. Cham: Springer International Publishing, 117-142.

Hanssens, D. M., and K. H. Pauwels. 2016. Demonstrating the Value of Marketing. *Journal of Marketing* 80 (6):173-190.

Hanssens, D. M., K. H. Pauwels, S. Srinivasan, M. Vanhuele, and G. Yildirim. 2014. Consumer Attitude Metrics for Guiding Marketing Mix Decisions. *Marketing Science* 33 (4):534-550.

Hanssens, D. M., F. Wang, and X.-P. Zhang. 2016. Performance growth and opportunistic marketing spending. *International Journal of Research in Marketing* 33 (4):711-724.

Haslett, J., and A. E. Raftery. 1989. Space-Time Modelling with Long-Memory Dependence: Assessing Ireland's Wind Power Resource. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 38 (1):1-50.

Hou, J., Y. Zhang, and X. Gu. 2016. Synergy and antagonism in online advertising. In *3rd IEEE/ACM International Conference on Big Data Computing, Applications and Technologies, BDCAT 2016*: ACM, 293-301.

Huber, J., and H. Stuckenschmidt. 2020. Daily retail demand forecasting using machine learning with emphasis on calendric special days. *International Journal of Forecasting* 36 (4):1420-1438.

Hyndman, R. J., R. A. Ahmed, G. Athanasopoulos, and H. L. Shang. 2011. Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis* 55 (9):2579-2589.

Hyndman, R. J., and G. Athanasopoulos. 2018. *Forecasting: principles and practice*. 2nd ed. Melbourne, Australia: OTexts.

Hyndman, R. J., and Y. Khandakar. 2008. Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software* 27 (3):1-22.

Jayawardane, C. H., S. K. Halgamuge, and U. Kayande. 2015. Attributing Conversion Credit in an Online Environment: An Analysis and Classification. In *3rd International Symposium on Computational and Business Intelligence* Bali, Indonesia, 68-73.

Ji, W., and X. Wang. 2017. Additional multi-touch attribution for online advertising. In *31st AAAI Conference on Artificial Intelligence, AAAI 2017*: AAAI Press, 1360-1366.

Ji, W., X. Wang, and D. Zhang. 2016. A probabilistic multi-touch attribution model for online advertising. In *25th ACM International Conference on Information and Knowledge Management, CIKM 2016*: ACM, 1373-1382.

Jin, Y., Y. Wang, Y. Sun, D. Chan, and J. Koehler. 2017. Bayesian Methods for Media Mix Modeling with Carryover and Shape Effects: Google Inc.

Joo, M., K. C. Wilbur, B. Cowgill, and Y. Zhu. 2014. Television Advertising and Online Search. *Management Science* 60 (1):56-73.

Joo, M., K. C. Wilbur, and Y. Zhu. 2016. Effects of TV advertising on keyword search. *International Journal of Research in Marketing* 33 (3):508-523.

Kaatz, C., C. Brock, and L. Figura. 2019. Are you still online or are you already mobile? - Predicting the path to successful conversions across different devices. *Journal of Retailing and Consumer Services* 50:10-21.

Kakalejcik, L., J. Bucko, and M. Vejacka. 2019. Differences in Buyer Journey between High- and Low-Value Customers of E-Commerce Business. *Journal of Theoretical and Applied Electronic Commerce Research* 14 (2):47-58.

Kannan, P. K., and H. A. Li. 2017. Digital marketing: A framework, review and research agenda. *International Journal of Research in Marketing* 34 (1):22-45.

Kannan, P. K., W. Reinartz, and P. C. Verhoef. 2016. The path to purchase and attribution modeling: Introduction to special section. *International Journal of Research in Marketing* 33 (3):449-456.

Kireyev, P., K. Pauwels, and S. Gupta. 2016. Do display ads influence search? Attribution and dynamics in online advertising. *International Journal of Research in Marketing* 33 (3):475-490.

Kolsarici, C., and D. Vakratsas. 2018. Synergistic, Antagonistic, and Asymmetric Media Interactions. *Journal of Advertising* 47 (3):282-300.

Kumar, A., R. Bezawada, R. Rishika, R. Janakiraman, and P. K. Kannan. 2016. From Social to Sale: The Effects of Firm-Generated Content in Social Media on Customer Behavior. *Journal of Marketing* 80 (1):7-25.

Kumar, V., J. B. Choi, and M. Greene. 2017. Synergistic effects of social media and traditional marketing on brand sales: capturing the time-varying effects. *Journal of the Academy of Marketing Science* 45 (2):268-288.

Kumar, V., and S. Gupta. 2016. Conceptualizing the Evolution and Future of Advertising. *Journal of Advertising* 45 (3):302-317.

Leeflang, P. S. H., T. H. A. Bijmolt, J. van Doorn, D. M. Hanssens, H. J. van Heerde, P. C. Verhoef, and J. E. Wieringa. 2009. Creating lift versus building the base: Current trends in marketing dynamics. *International Journal of Research in Marketing* 26 (1):13-20.

Li, H. A., and P. K. Kannan. 2014. Attributing conversions in a multichannel online marketing environment: An empirical model and a field experiment. *Journal of Marketing Research* 51 (1):40-56.

Li, Y., Y. Xie, and Z. Zheng. 2019. Modeling Multichannel Advertising Attribution across Competitors. *MIS Quarterly* 43 (1):263-286.

Liaukonyte, J., T. Teixeira, and K. C. Wilbur. 2015. Television Advertising and Online Shopping. *Marketing Science* 34 (3):311-330.

Lin, C., S. Venkataraman, and S. D. Jap. 2013. Media Multiplexing Behavior: Implications for Targeting and Media Planning. *Marketing Science* 32 (2):310-324.

Liu-Thompkins, Y. 2019. A Decade of Online Advertising Research: What We Learned and What We Need to Know. *Journal of Advertising* 48 (1):1-13.

Lovett, M. J., R. Peres, and L. Xu. 2019. Can your advertising really buy earned impressions? The effect of brand advertising on word of mouth. *Quantitative Marketing and Economics* 17 (3):215-255.

Lütkepohl, H. 2007. *New Introduction to Multiple Time Series Analysis*. 2nd ed. Berlin: Springer.

Ma, S., and R. Fildes. 2021. Retail sales forecasting with meta-learning. *European Journal of Operational Research* 288 (1):111-128.

Makridakis, S., R. J. Hyndman, and F. Petropoulos. 2020. Forecasting in social settings: The state of the art. *International Journal of Forecasting* 36 (1):15-28.

Mantrala, M., and V. Kanuri. 2016. Marketing Optimization Methods. In *Handbook of Marketing Analytics with Applications in Marketing, Public Policy, and Litigation*, edited by N. Mizik and D. M. Hanssens. Cheltenham: Edward Elgar Publishing, 324-370.

Marketing Science Institute. 2018. *Research Priorities 2018-2020*. Cambridge: Marketing Science Institute.

McCarthy, E. J. 1960. *Basic Marketing - A Managerial Approach*. Homeword, Illinois: Richard D. Irwin, Inc.

Montero-Manso, P., G. Athanasopoulos, R. J. Hyndman, and T. S. Talagala. 2020. FFORMA: Feature-based forecast model averaging. *International Journal of Forecasting* 36 (1):86-92.

Naik, P. A., and K. Peters. 2009. A Hierarchical Marketing Communications Model of Online and Offline Media Synergies. *Journal of Interactive Marketing* 23 (4):288-299.

Naik, P. A., and K. Raman. 2003. Understanding the Impact of Synergy in Multimedia Communications. *Journal of Marketing Research* 40 (4):375-388.

Nichols, W. 2013. Advertising Analytics 2.0 Advertising touch points no longer work in isolation. If your company doesn't analyze how they interact dynamically, it will be left behind. *Harvard Business Review* 91 (3):60-69.

Nisar, T. M., and M. Yeung. 2018. Attribution Modeling In Digital Advertising An Empirical Investigation Of the Impact of Digital Sales Channels. *Journal of Advertising Research* 58 (4):399-413.

Nottorf, F. 2014. Multi-channel Attribution Modeling on User Journeys. In *E-Business and Telecommunications, Icete 2013*. Berlin: Springer-Verlag Berlin, 107-125.

O'Hara, K. *BMR: Classical VAR* 2018 [cited June 2020. Available from https://www.kthohr.com/bmr_docs_vars_cvar.html.

Osinga, E. C., P. S. H. Leeflang, and J. E. Wieringa. 2010. Early Marketing Matters: A Time-Varying Parameter Approach to Persistence Modeling. *Journal of Marketing Research* 47 (1):173-185.

Palmatier, R. W., M. B. Houston, and J. Hulland. 2018. Review articles: purpose, process, and structure. *Journal of the Academy of Marketing Science* 46 (1):1-5.

Papies, D., P. Ebbes, and H. J. V. Heerde. 2017. Addressing Endogeneity in Marketing Models. In *Advanced Methods for Modeling Markets*, edited by P. S. H. Leeflang, J. E. Wieringa, T. H. A. Bijmolt and K. H. Pauwels: Springer International Publishing, 581-627.

Pauwels, K., Z. Aksehirli, and A. Lackman. 2016a. Like the ad or the brand? Marketing stimulates different electronic word-of-mouth content to drive online and offline performance. *International Journal of Research in Marketing* 33 (3):639-655.

Pauwels, K., C. Demirci, G. Yildirim, and S. Srinivasan. 2016b. The impact of brand familiarity on online and offline media synergy. *International Journal of Research in Marketing* 33 (4):739-753.

Perron, P. 1989. The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis. *Econometrica* 57 (6):1361-1401.

Ren, K., Y. Fang, W. Zhang, S. Liu, J. Li, Y. Zhang, Y. Yu, and J. Wang. 2018. Learning multi-touch conversion attribution with dual-attention mechanisms for online advertising. Paper read at 27th ACM International Conference on Information and Knowledge Management, CIKM 2018.

Saboo, A. R., V. Kumar, and I. Park. 2016. Using Big Data to Model Time-Varying Effects for Marketing Resource (Re)Allocation. *MIS Quarterly* 40 (4):911-939.

Shao, X., and L. Li. 2011. Data-driven multi-touch attribution models. Paper read at 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, at San Diego, US.

Sinha, R., S. Saini, and N. Anadhavelu. 2014. Estimating the incremental effects of interactions for marketing attribution. Paper read at 2014 1st IEEE International Conference on Behavioral, Economic, Socio-Cultural Computing, BESC 2014.

Song, Y., N. Sahoo, S. Srinivasan, and C. Dellarocas. 2016. Uncovering Characteristic Paths to Purchase of Consumers: Boston University Questrom School of Business.

Srinivasan, S., O. J. Rutz, and K. Pauwels. 2016. Paths to and off purchase: quantifying the impact of traditional marketing and online consumer activity. *Journal of the Academy of Marketing Science* 44 (4):440-453.

Srinivasan, S., M. Vanhuele, and K. Pauwels. 2010. Mind-Set Metrics in Market Response Models: An Integrative Approach. *Journal of Marketing Research* 47 (4):672-684.

Taieb, S. B., and R. J. Hyndman. 2014. A gradient boosting approach to the Kaggle load forecasting competition. *International Journal of Forecasting* 30 (2):382-394.

Tinyakova, V. I., V. V. Davnis, Y. B. Lavrinenko, and L. A. Shishkina. 2018. Probability models for assessing effectiveness of advertising channels in the internet environment. *Journal of Social Sciences Research* 2018 (Special Issue 5):209-215.

Verhoef, P. C., P. K. Kannan, and J. J. Inman. 2015. From Multi-Channel Retailing to Omni-Channel Retailing. *Journal of Retailing* 91 (2):174-181.

Wang, Y., Y. Jin, Y. Sun, D. Chan, and J. Koehler. 2017. A Hierarchical Bayesian Approach to Improve Media Mix Models Using Category Data: Google Inc.

Wedel, M., and P. K. Kannan. 2016. Marketing Analytics for Data-Rich Environments. *Journal of Marketing* 80 (6):97-121.

Weron, R. 2014. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting* 30 (4):1030-1081.

West, D., J. B. Ford, and P. W. Farris. 2014. How Corporate Cultures Drive Advertising and Promotion Budgets. *Journal of Advertising Research* 54 (2):149.

Wickramasuriya, S. L., G. Athanasopoulos, and R. J. Hyndman. 2019. Optimal Forecast Reconciliation for Hierarchical and Grouped Time Series Through Trace Minimization. *Journal of the American Statistical Association* 114 (526):804-819.

Wiesel, T., K. Pauwels, and J. Arts. 2011. Marketing's Profit Impact: Quantifying Online and Off-line Funnel Progression. *Marketing Science* 30 (4):604-611.

Winters, P. R. 1960. Forecasting Sales by Exponentially Weighted Moving Averages. *Management Science* 6 (3):324–342.

Wooff, D. A., and J. M. Anderson. 2013. Time-weighted multi-touch attribution and channel relevance in the customer journey to online purchase. *Journal of Statistical Theory and Practice* 9 (2):227-249.

Xu, L., J. A. Duan, and A. Whinston. 2014. Path to Purchase: A Mutually Exciting Point Process Model for Online Advertising and Conversion. *Management Science* 60 (6):1392-1412.

Yadagiri, M. M., S. K. Saini, and R. Sinha. 2015. A non-parametric approach to the multi-channel attribution problem. In *16th International Conference on Web Information Systems Engineering, WISE 2015*, edited by W. Cellary, S.-C. Chen, T. Li, D. Wang, H. Wang, J. Wang and Y. Zhang. Berlin: Springer Verlag, 338-352.

Yin, Z., Y. Li, P. Mazzoleni, and Y. Shen. 2016. Mining Effective Subsequences with Application in Marketing Attribution. In *2016 IEEE 16th International Conference on Data Mining Workshops*. New York: IEEE, 700-707.

Yuvaraj, C. B., B. R. Chandavarkar, V. S. Kumar, and B. S. Sandeep. 2018. Enhanced Last-Touch Interaction Attribution Model in Online Advertising. In *Proceedings of 2018 Ieee Distributed Computing, Vlsi, Electrical Circuits and Robotics*. New York: IEEE, 110-114.

Zantedeschi, D., E. M. Feit, and E. T. Bradlow. 2016. Measuring Multichannel Advertising Response. *Management Science* 63 (8):2706-2728.

Zhang, Y., Y. Wei, and J. Ren. 2014. Multi-touch Attribution in Online Advertising with Survival Theory. In *14th IEEE International Conference on Data Mining, ICDM 2014*, edited by J. Z. Huang, H. Toivonen, R. Kumar, J. Pei and X. Wu: IEEE, 687-696.

Zhao, K. F., S. H. Mahboobi, and S. R. Bagheri. 2019. Revenue-based attribution modeling for online advertising. *International Journal of Market Research* 61 (2):195-209.

Zupic, I., and T. Čater. 2015. Bibliometric Methods in Management and Organization. *Organizational Research Methods* 18 (3):429-472.

Appendices

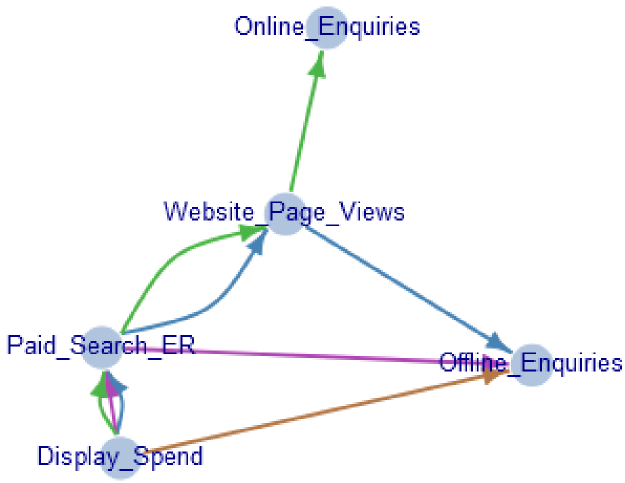
Appendix 1 – Breakdown of Advertising Direct & Indirect Effects (Chapter 3)

This section provides a detailed breakdown of the direct and indirect effects of each media channel on online and offline enquiries for the luxury travel brand discussed in Chapter 3. First, the direct and indirect effects are visualised in a network diagram created using the *igraph* package in R[7]. The network diagram provides marketing managers with a visualisation of the paths to purchase identified in the analysis, which is useful for reporting and strategic decision-making. Details of each path to purchase from the media channel to online or offline enquiries are presented in a table, including the number of variables in the path, the path type and the total effect (IRF) contribution of each path. Two additional visualisations are included. One shows the percentage breakdown of each path over time. For example, if a path has a value of 60% in period one, it means that 60% of the total effect of that path occurred in period one. The other visualisation shows the percentage breakdown of the total effects for each marketing performance metric over time, which is the result of adding up the effects of all of the individual paths. Percentage breakdowns provide insight into the length and concentration of direct and indirect effects. They are also used as some paths are substantially larger than others in raw terms (see path details tables). In practice, these visualisations could be presented in an interactive dashboard for marketing managers to use for reporting and strategic decision-making.

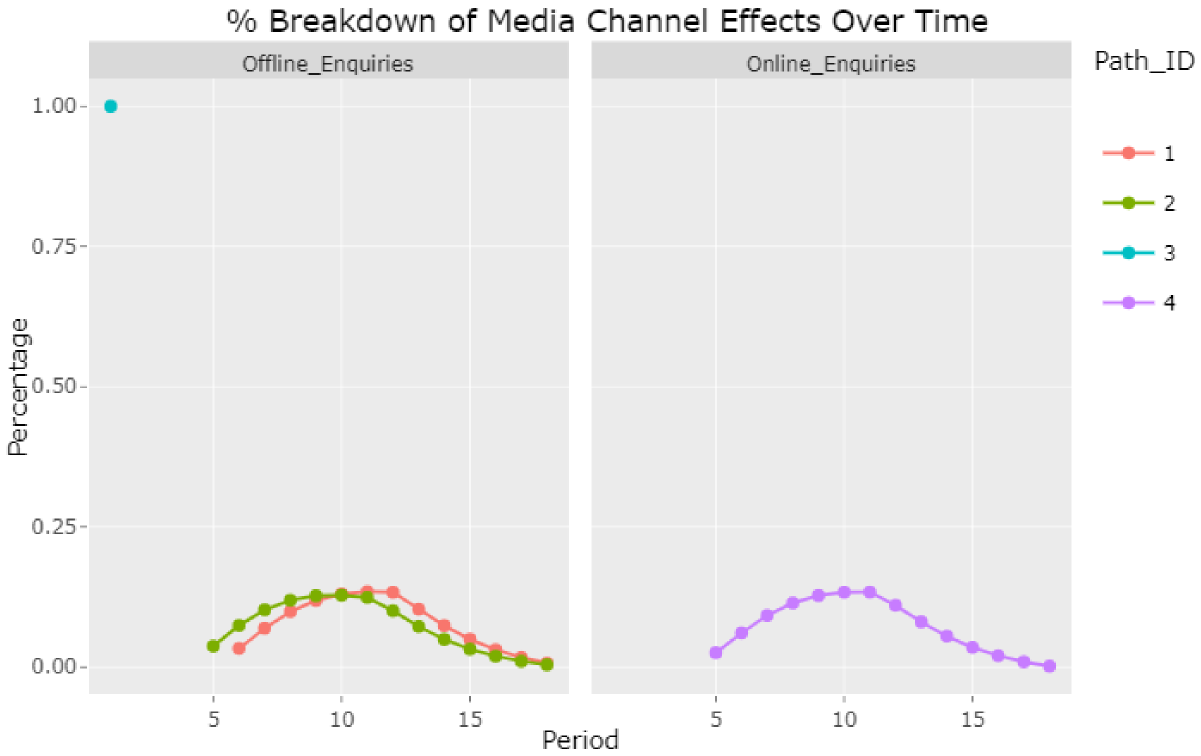
Combining the information in Table 3-4 (see page 5656) with this section can provide further insights into the timing of indirect effects on online and offline enquiries. For example, magazine spend affects website page views in period 2 and website page views affects online enquiries in periods 1 and 2. Therefore, magazine spend affects online enquiries in periods 3 and 4 after the shock to magazine spend. Adding the starting period of a path with the total periods of each direct effect in the path reveals the number of periods the indirect effect lasts.

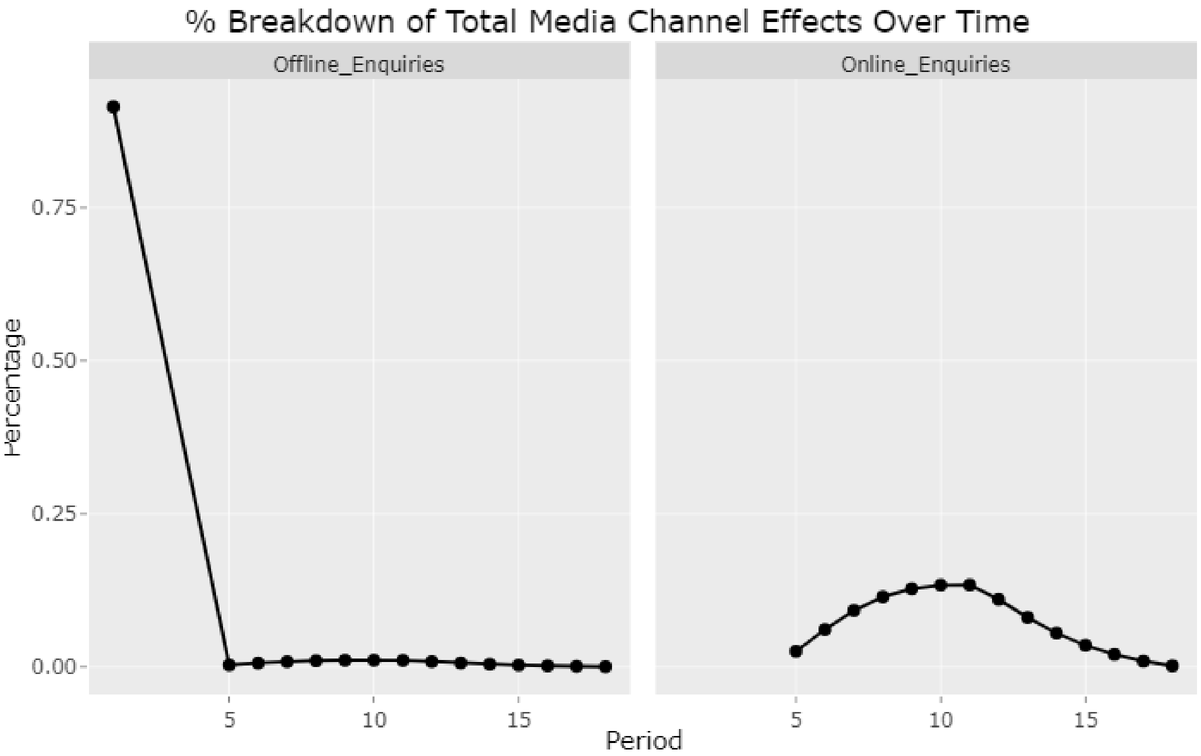
[7] See https://igraph.org/r/ for documentation.

Display Spend

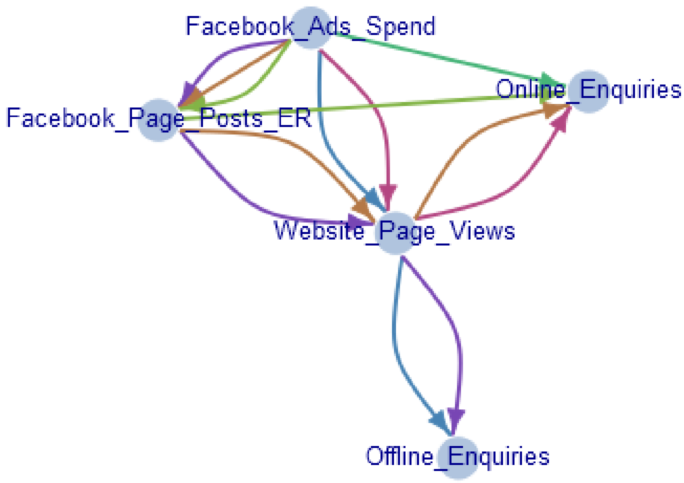


Path	**Path Length**	**Effect Type**	**Total IRF**
3 - Display Spend -> Offline Enquiries (all of the effect occurs in one period)	2	Direct	0.0415
2 - Display Spend -> Paid Search Engagement Rate -> Offline Enquiries	3	Indirect	0.0037
1 - Display Spend -> Paid Search Engagement Rate -> Website Page Views -> Offline Enquiries	4	Indirect	0.0002
4 - Display Spend -> Paid Search Engagement Rate -> Website Page Views -> Online Enquiries	4	Indirect	0.0006

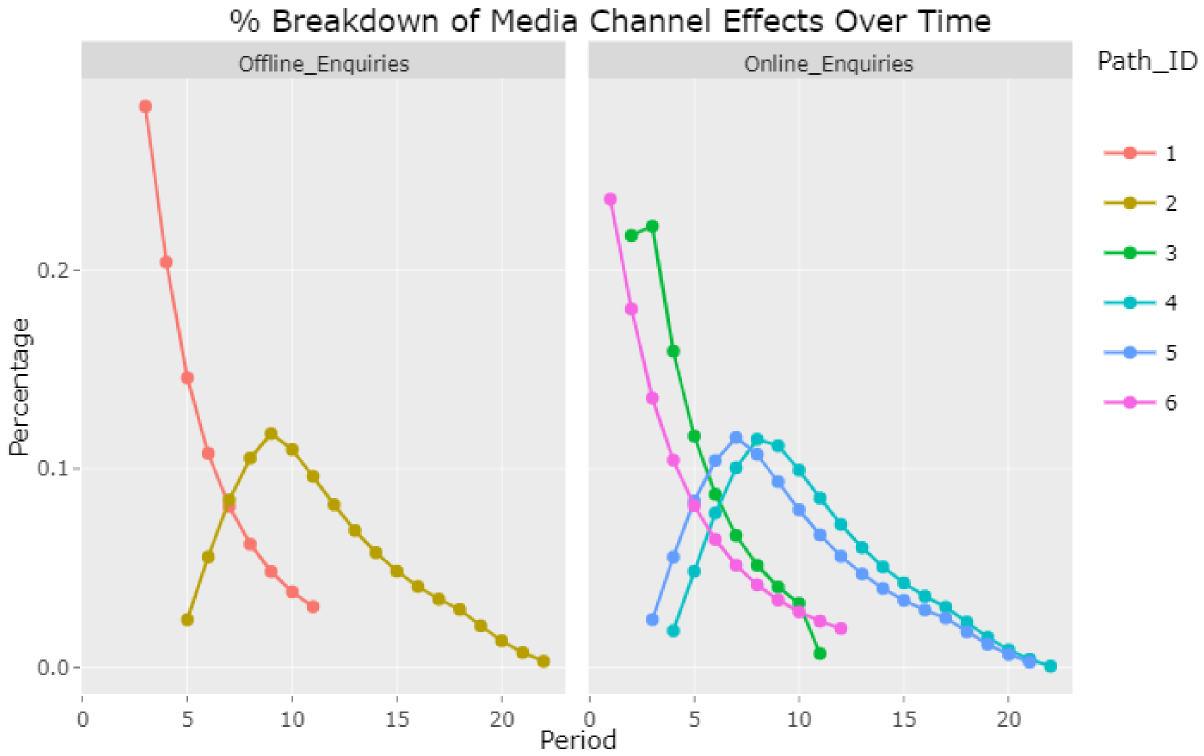

% Breakdown of Media Channel Effects Over Time
Offline_Enquiries
Online_Enquiries
Path_ID
1
2
3
4
Percentage
1.00
0.75
0.50
0.25
0.00
5
10
15
Period

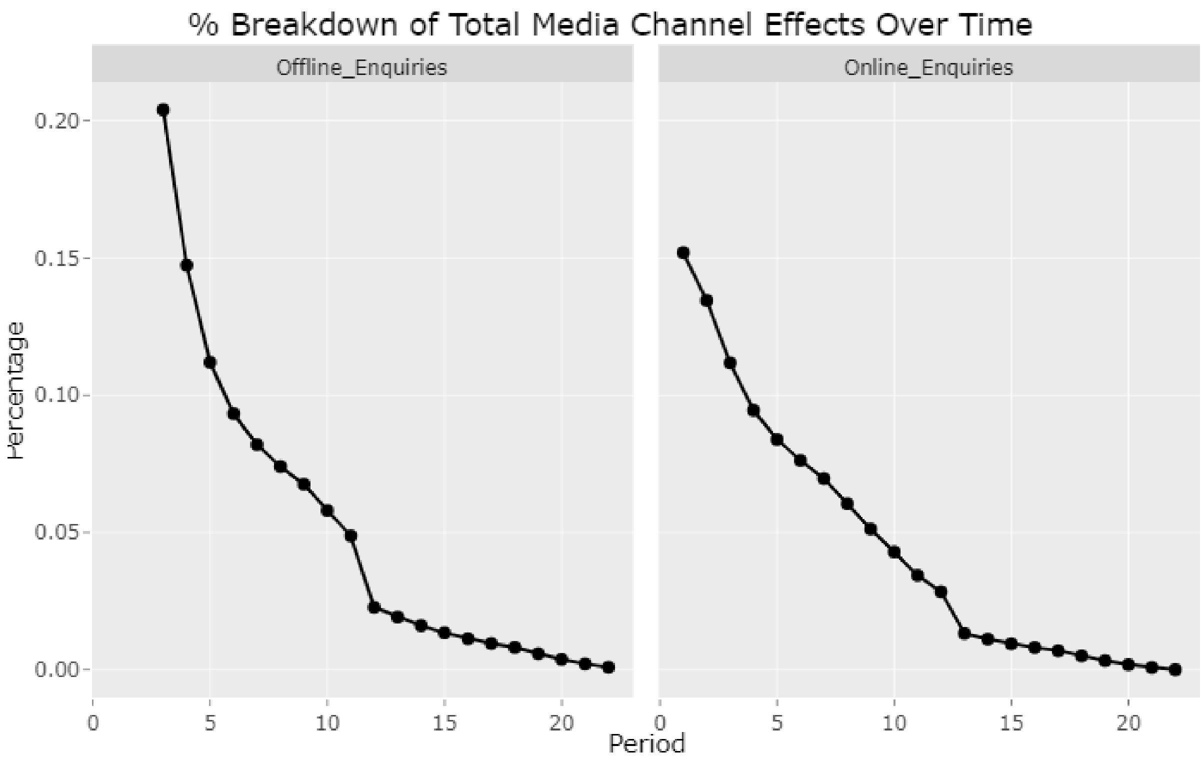

% Breakdown of Total Media Channel Effects Over Time
Offline_Enquiries
Online_Enquiries
Percentage
0.75
0.50
0.25
0.00
5
10
15
Period

Facebook Ads Spend



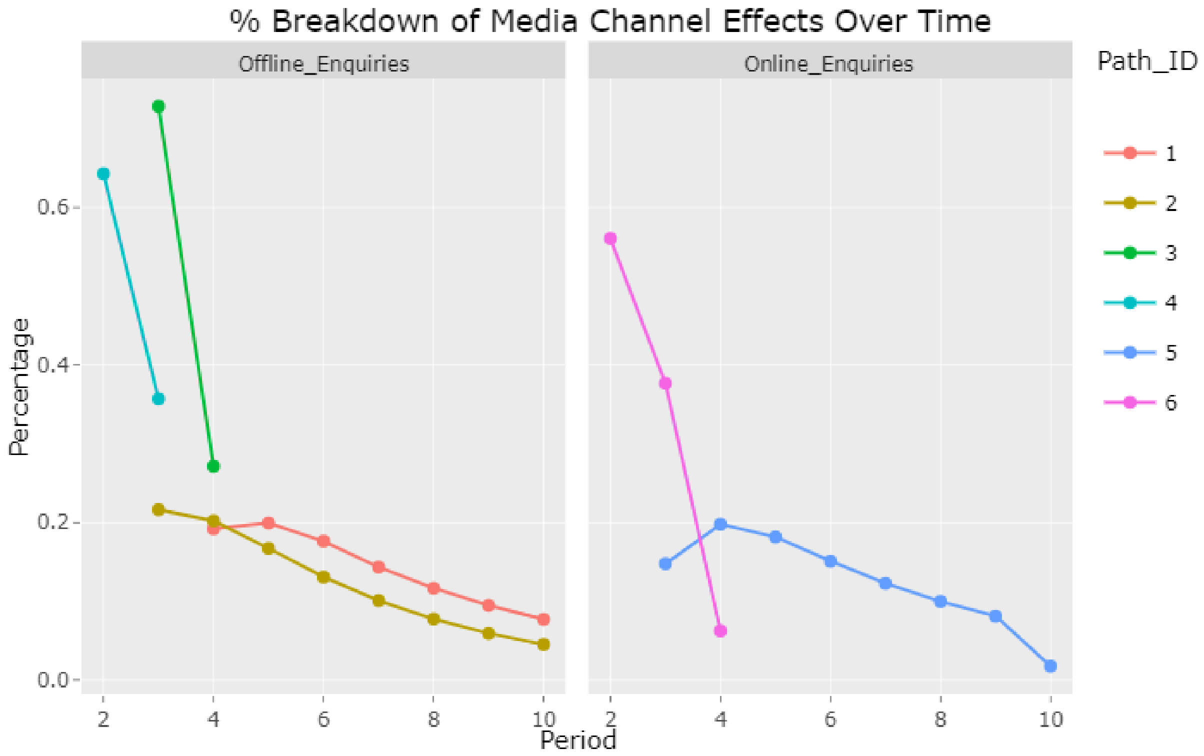
Path	Path Length	Effect Type	Total IRF
6 - Facebook Ads Spend -> Online Enquiries	2	Direct	0.5864
3 - Facebook Ads Spend -> Website Page Views -> Online Enquiries	3	Indirect	0.0761
1 - Facebook Ads Spend -> Website Page Views -> Offline Enquiries	3	Indirect	0.0238
5 - Facebook Ads Spend -> Facebook Page Posts Engagement -> Online Enquiries	3	Indirect	0.2174
2 - Facebook Ads Spend -> Facebook Page Posts Engagement -> Website Page Views -> Offline Enquiries	4	Indirect	0.0092
4 - Facebook Ads Spend -> Facebook Page Posts Engagement -> Website Page Views -> Online Enquiries	4	Indirect	0.0292

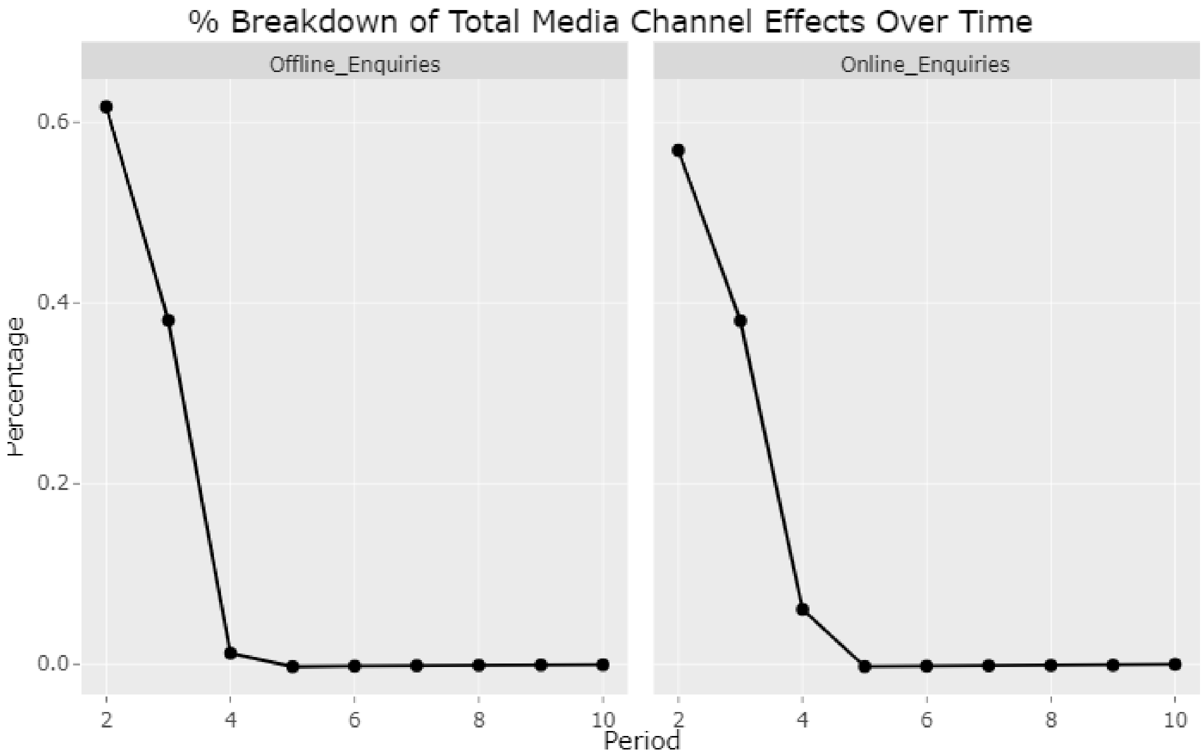

% Breakdown of Media Channel Effects Over Time
Offline_Enquiries
Online_Enquiries
Path_ID
1
2
3
4
5
6
Percentage
0.0
0.1
0.2
0
5
10
15
20
Period


% Breakdown of Total Media Channel Effects Over Time
Offline_Enquiries
Online_Enquiries
Percentage
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
Period

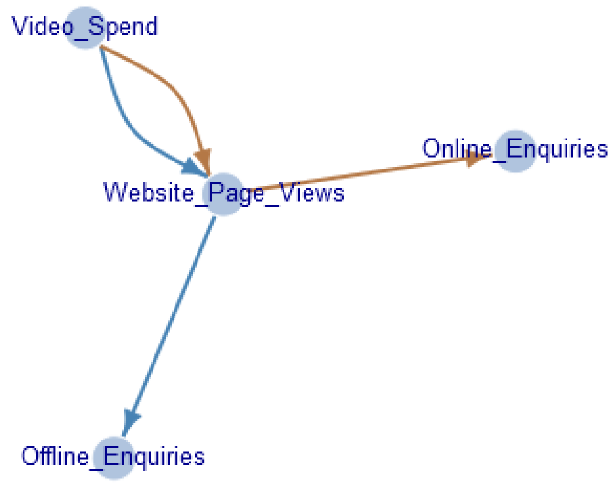
Paid Search Spend

Path	Path Length	Effect Type	Total IRF
4 - Paid Search Spend -> Offline Enquiries	2	Direct	0.08715
3 - Paid Search Spend -> Website Page Views -> Offline Enquiries	3	Indirect	0.0052
6 - Paid Search Spend -> Website Page Views -> Online Enquiries	3	Indirect	0.0165
2 - Paid Search Spend -> Paid Search Engagement Rate -> Offline Enquiries	3	Indirect	-0.0014
1 - Paid Search Spend -> Paid Search Engagement Rate -> Website Page Views -> Offline Enquiries	4	Indirect	-0.0001
5 - Paid Search Spend -> Paid Search Engagement Rate -> Website Page Views -> Online Enquiries	4	Indirect	-0.0002

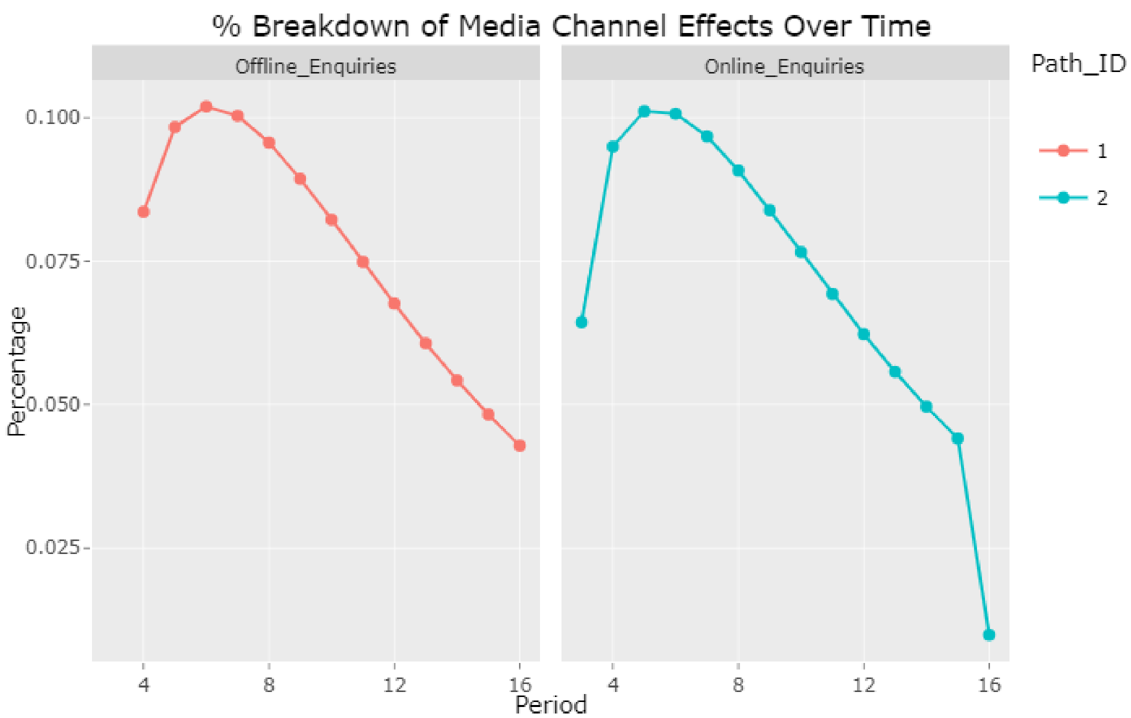

% Breakdown of Media Channel Effects Over Time
Offline_Enquiries
Online_Enquiries
Path_ID
1
2
3
4
5
6
Percentage
0.0
0.2
0.4
0.6
2
4
6
8
10
Period


% Breakdown of Total Media Channel Effects Over Time
Offline_Enquiries
Online_Enquiries
Percentage
0.0
0.2
0.4
0.6
2
4
6
8
10
Period

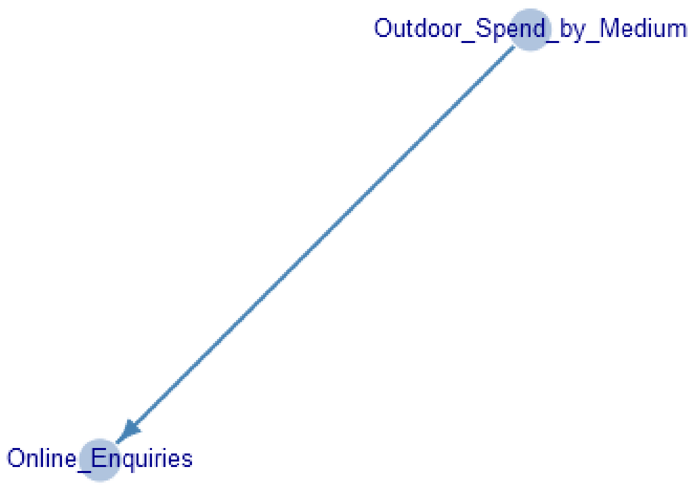
Video Spend



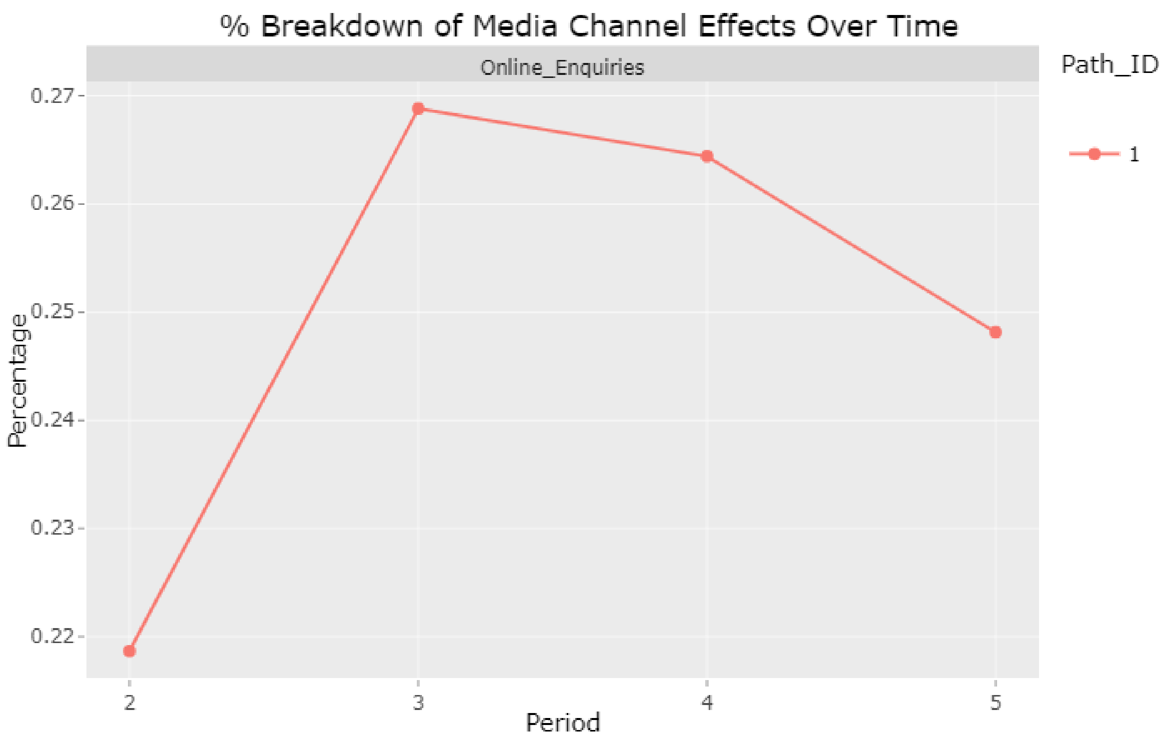
Path	Path Length	Effect Type	Total IRF
2 - Video Spend -> Website Page Views -> Online Enquiries	3	Indirect	0.0466
1 - Video Spend -> Website Page Views -> Offline Enquiries	3	Indirect	0.0146



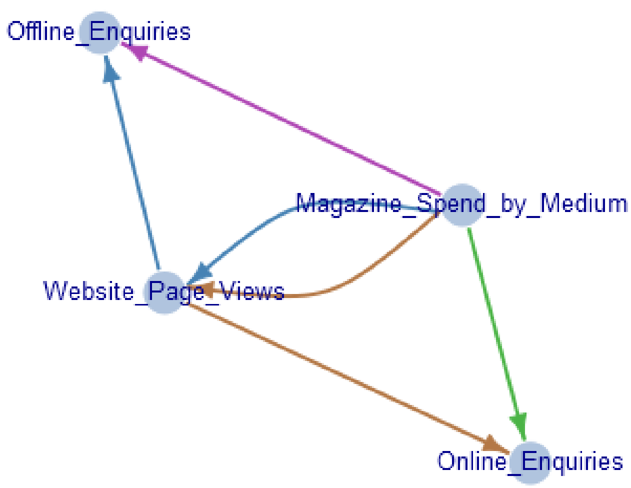
Outdoor Spend



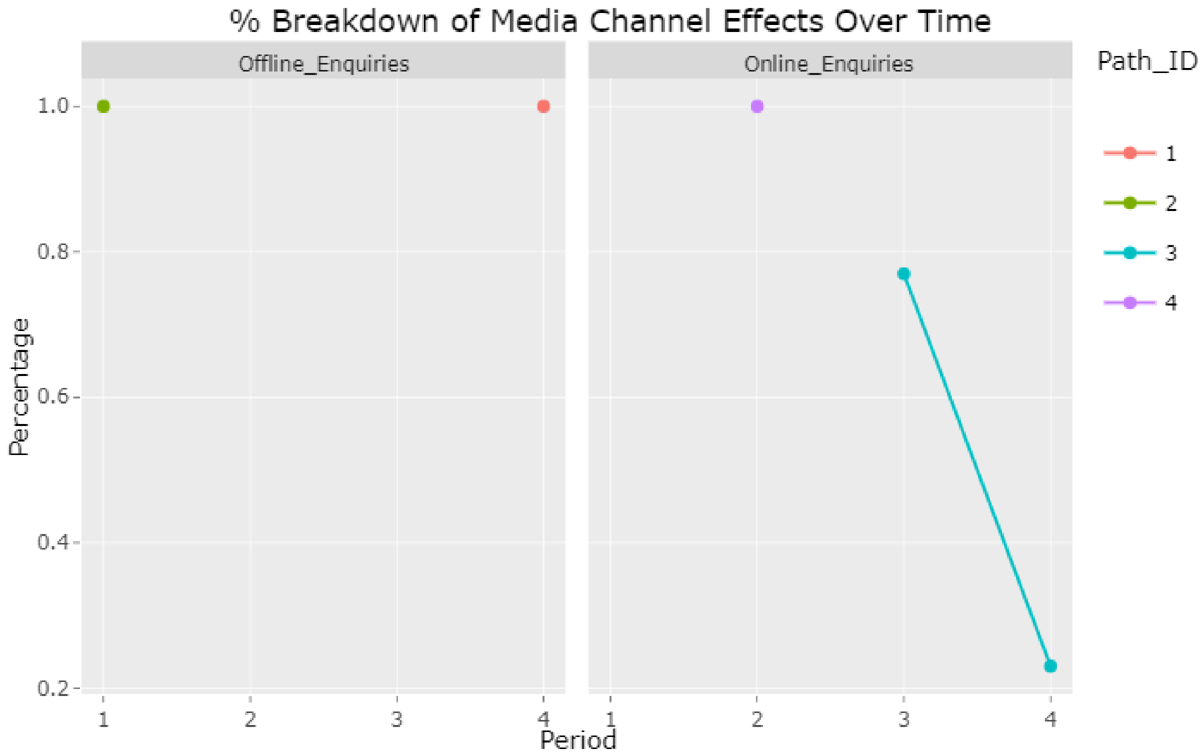
Path	Path Length	Effect Type	Total IRF
1 - Outdoor Spend -> Online Enquiries	2	Direct	0.0907

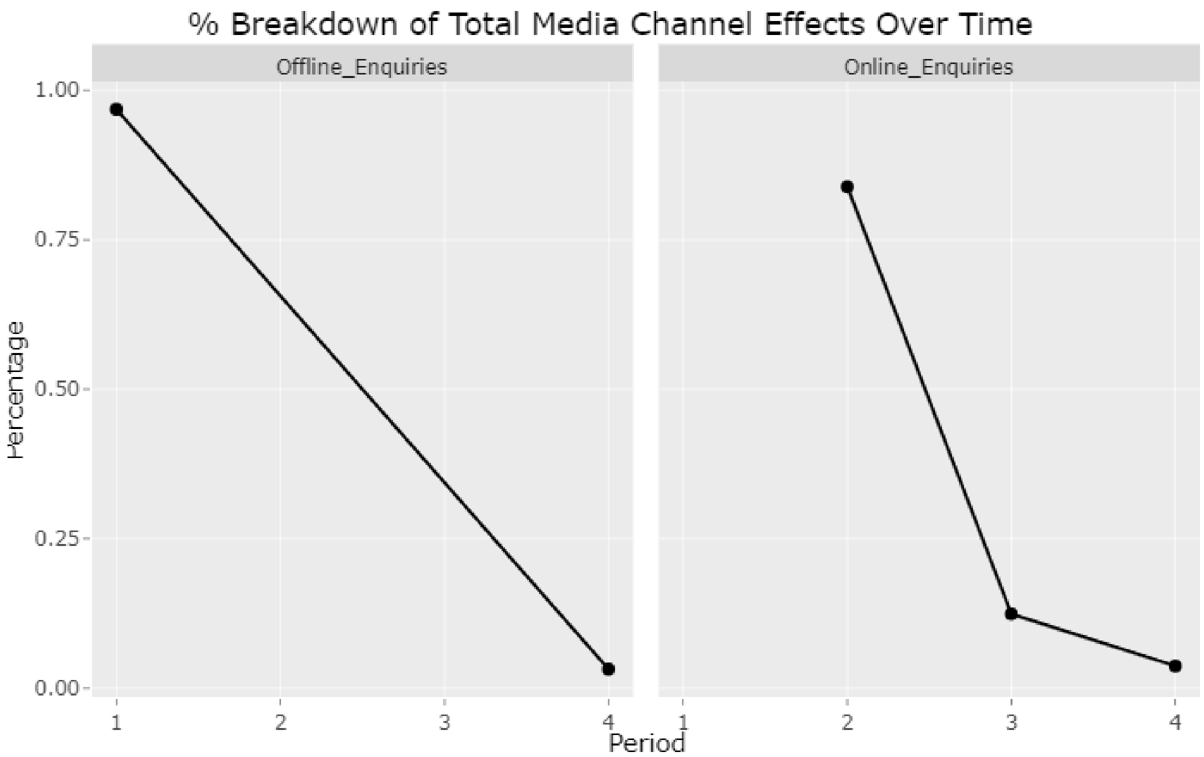


Magazine Spend

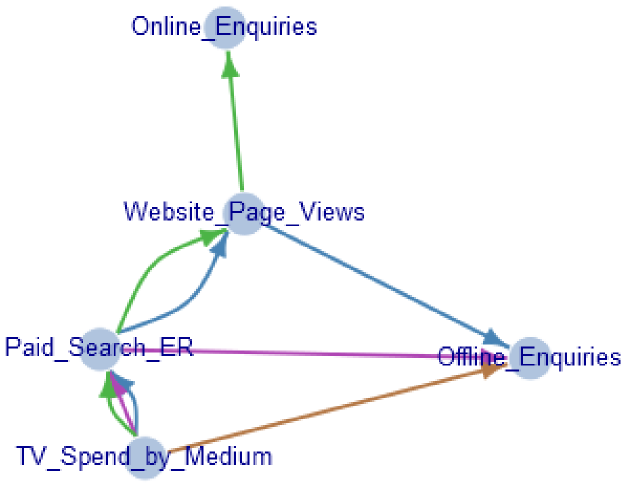


Path	Path Length	Effect Type	Total IRF
4 - Magazine Spend -> Offline Enquiries (all of the effect occurs in one period)	2	Direct	0.0738
2 - Magazine Spend -> Online Enquiries (all of the effect occurs in one period)	2	Direct	0.0404
3 - Magazine Spend -> Website Page Views -> Online Enquiries	3	Indirect	0.0078
1 - Magazine Spend -> Website Page Views -> Offline Enquiries (all of the effect occurs in one period)	3	Indirect	0.0024

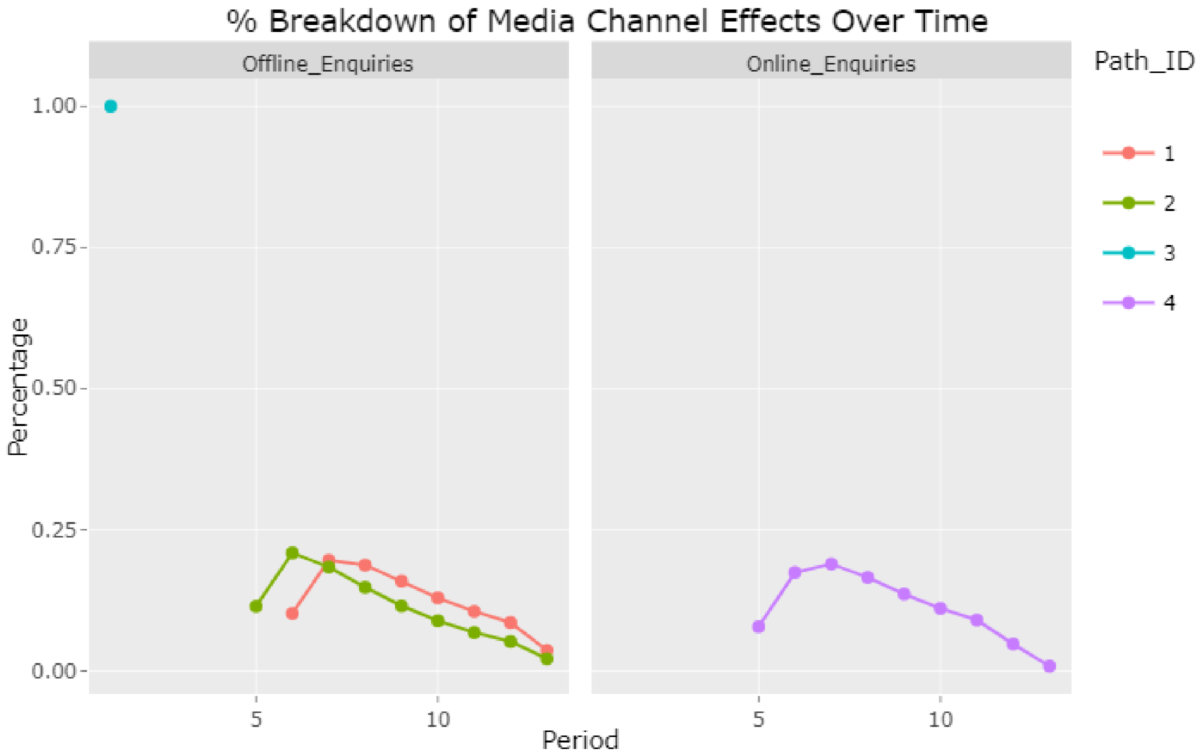

% Breakdown of Media Channel Effects Over Time
Offline_Enquiries
Online_Enquiries
Path_ID
1
2
3
4
Percentage
1.0
0.8
0.6
0.4
0.2
Period
1
2
3
4

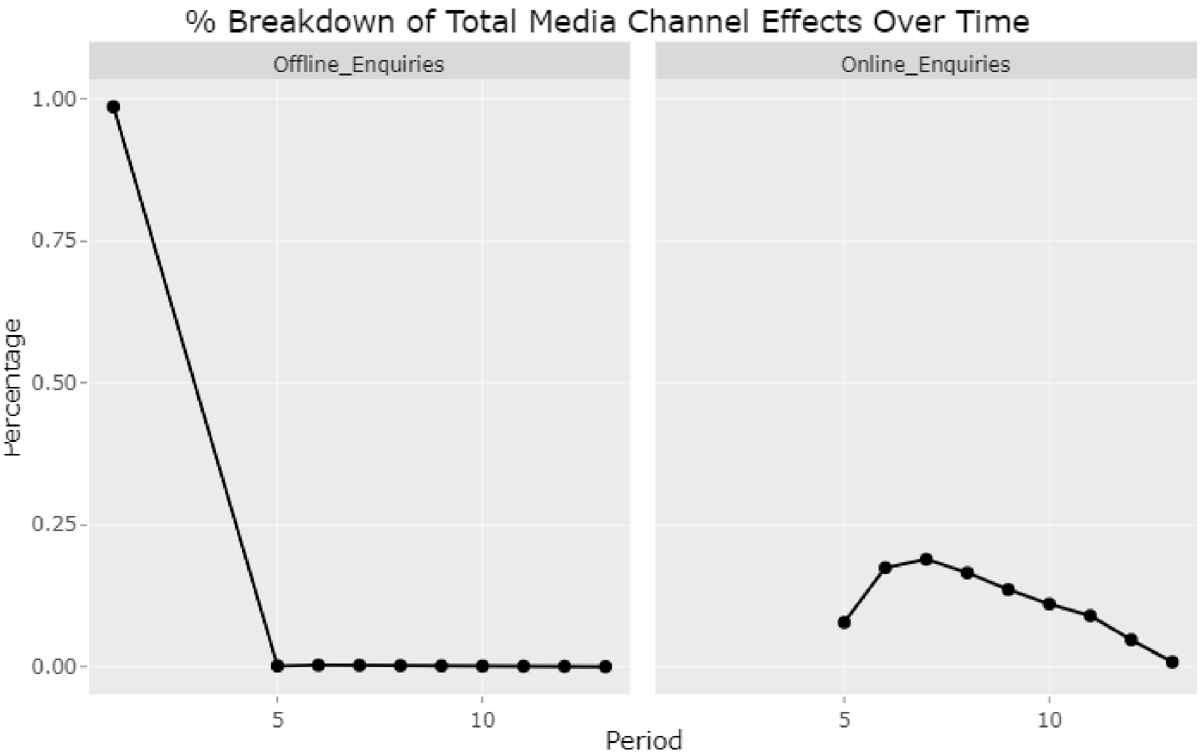

% Breakdown of Total Media Channel Effects Over Time
Offline_Enquiries
Online_Enquiries
Percentage
1.00
0.75
0.50
0.25
0.00
Period
1
2
3
4

TV Spend

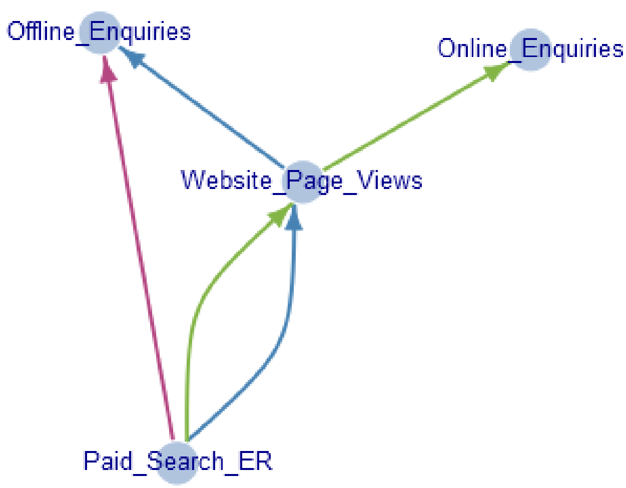


Path	Path Length	Effect Type	Total IRF
3 - TV Spend -> Offline Enquiries (all of the effect occurs in one period)	2	Direct	0.0829
2 - TV Spend -> Paid Search Engagement Rate -> Offline Enquiries	3	Indirect	0.0011
4 - TV Spend -> Paid Search Engagement Rate -> Website Page Views -> Online Enquiries	4	Indirect	0.0002
1 - TV Spend -> Paid Search Engagement Rate -> Website Page Views -> Offline Enquiries	4	Indirect	0.0001

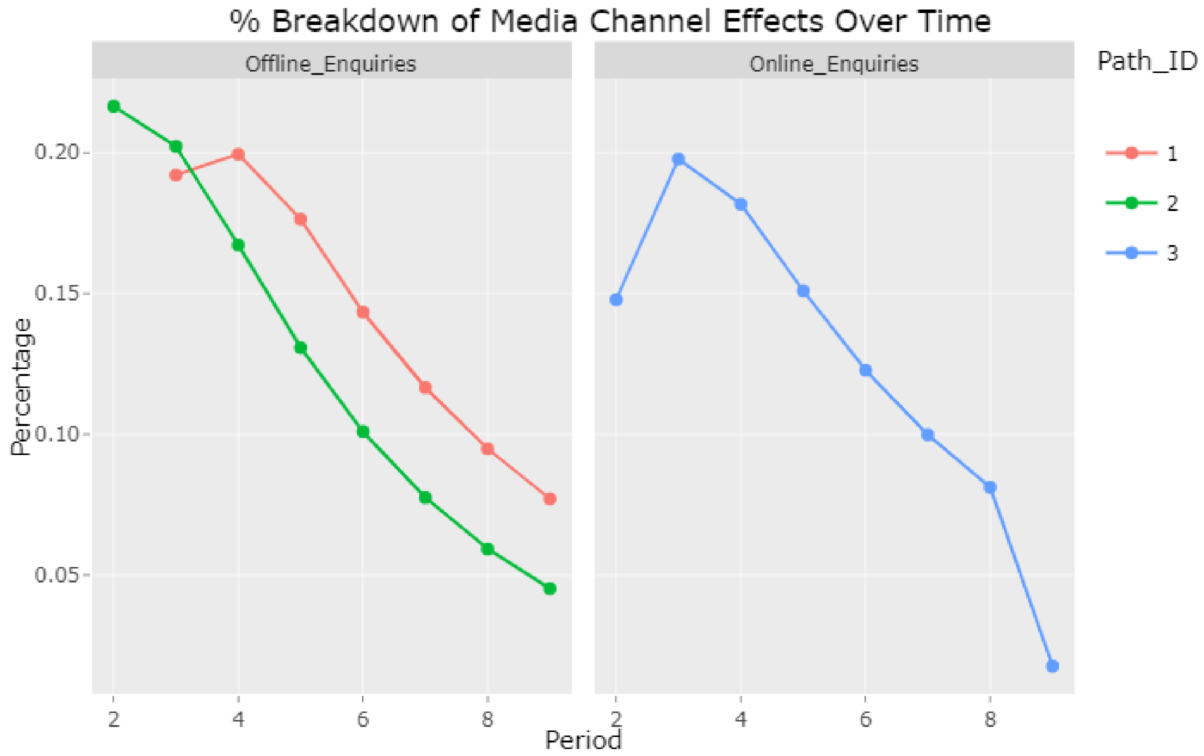

% Breakdown of Media Channel Effects Over Time
Offline_Enquiries
Online_Enquiries
Path_ID
1
2
3
4
Percentage
1.00
0.75
0.50
0.25
0.00
5
10
5
10
Period


% Breakdown of Total Media Channel Effects Over Time
Offline_Enquiries
Online_Enquiries
Percentage
1.00
0.75
0.50
0.25
0.00
5
10
5
10
Period

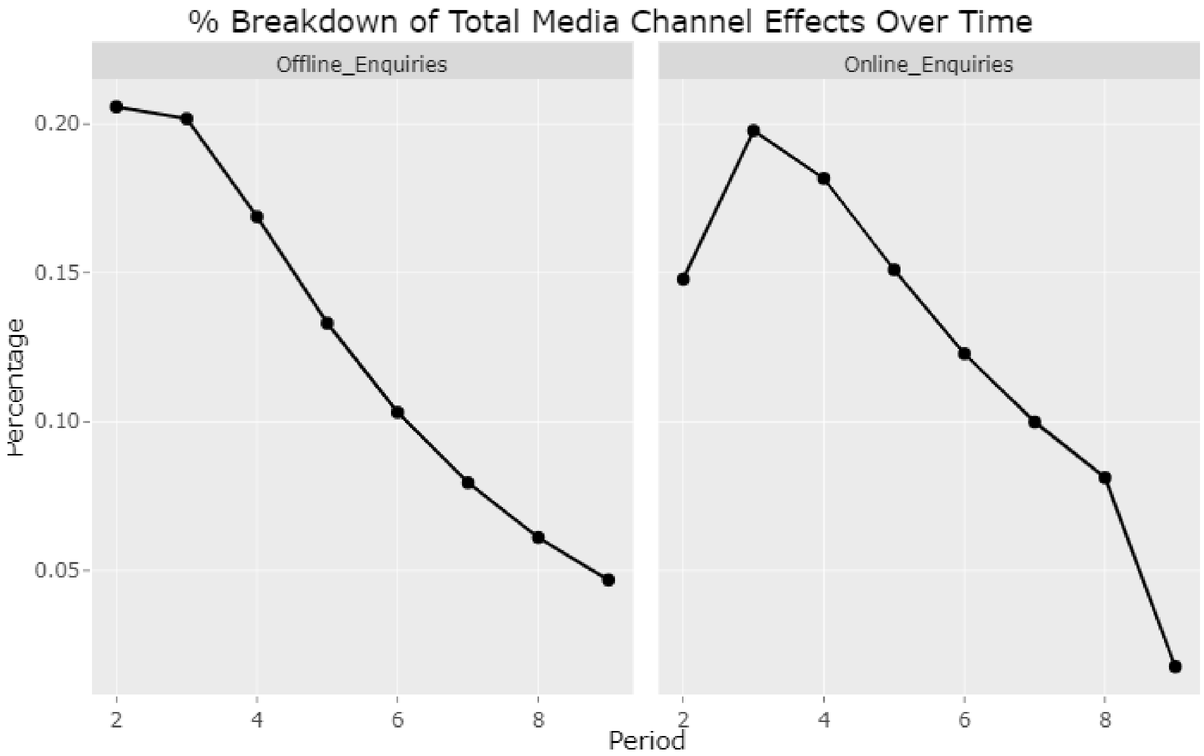
Paid Search Engagement Rate



Path	**Path Length**	**Effect Type**	**Total IRF**
2 - Paid Search Engagement Rate -> Offline Enquiries	2	Direct	0.2048
1 - Paid Search Engagement Rate -> Website Page Views -> Offline Enquiries	3	Indirect	0.0107
3 - Paid Search Engagement Rate -> Website Page Views -> Online Enquiries	3	Indirect	0.0341

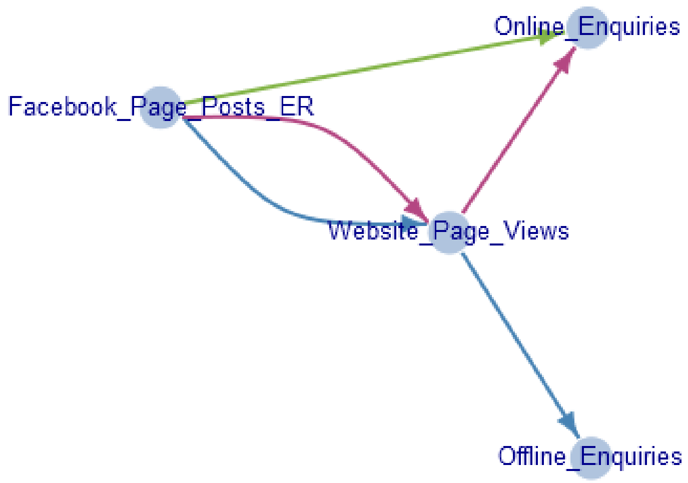


% Breakdown of Media Channel Effects Over Time
Offline_Enquiries
Online_Enquiries
Path_ID
1
2
3
0.20
0.15
0.10
0.05
Percentage
2
4
6
8
Period

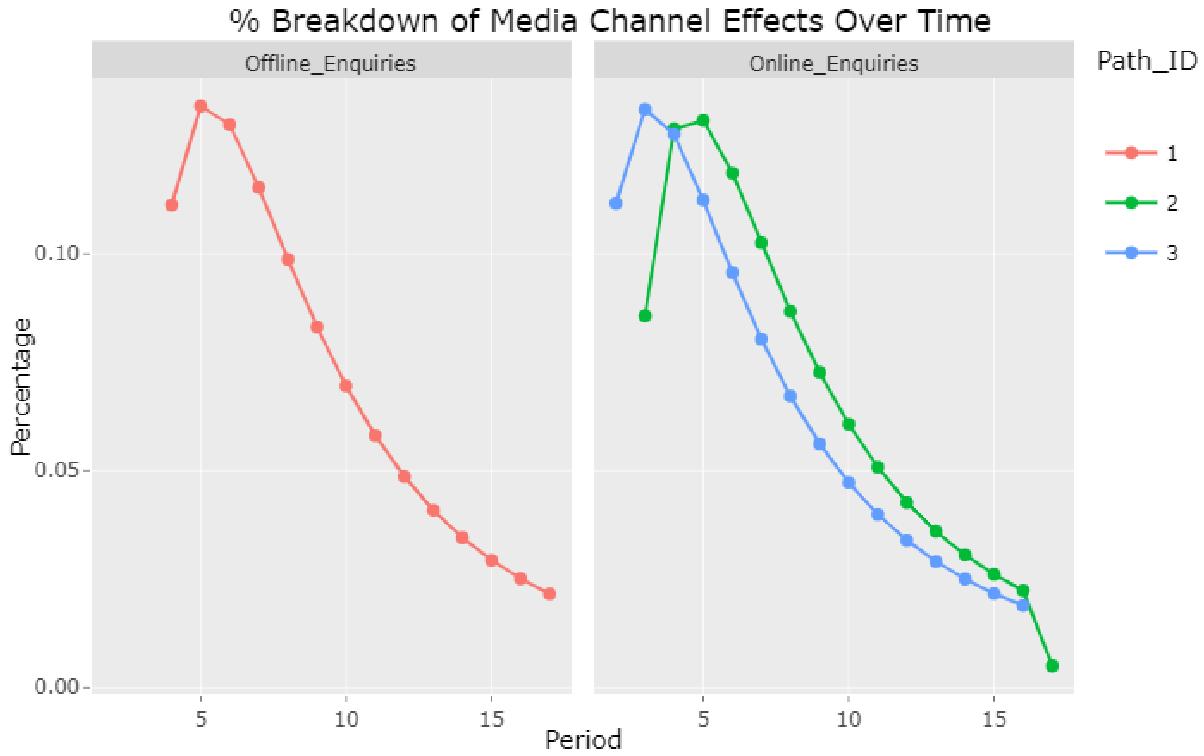


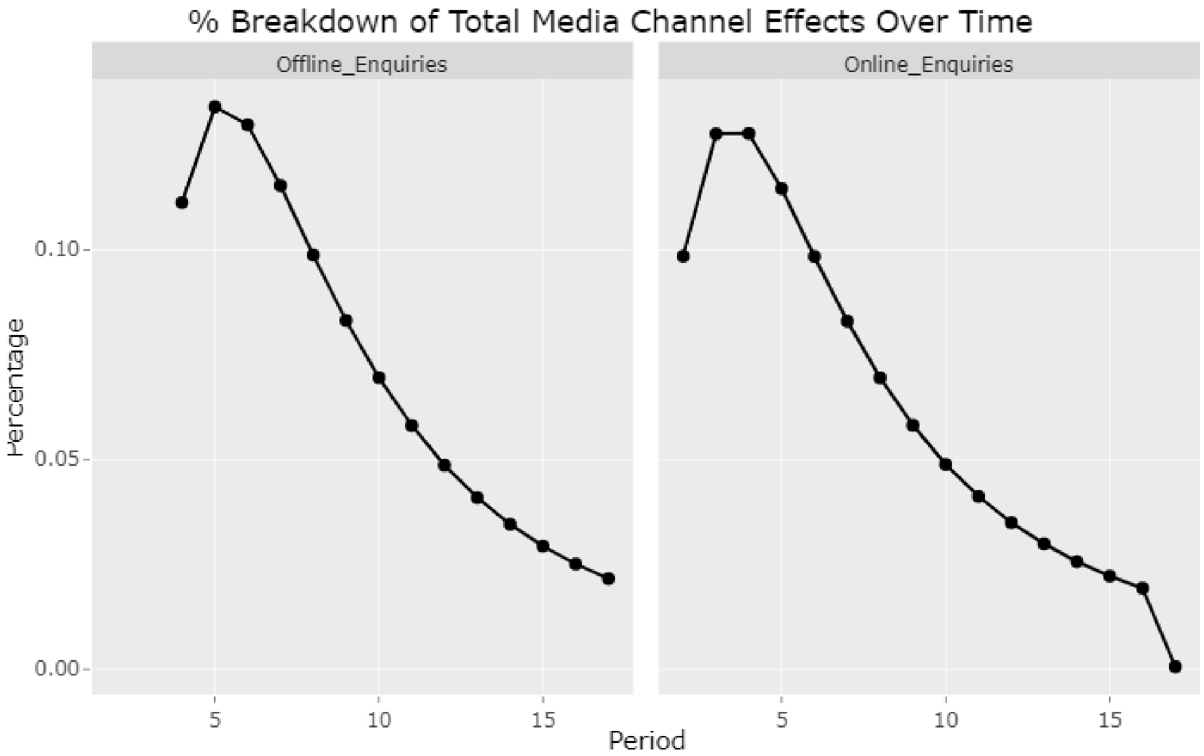
% Breakdown of Total Media Channel Effects Over Time
Offline_Enquiries
Online_Enquiries
0.20
0.15
0.10
0.05
Percentage
2
4
6
8
Period

Facebook Page Posts Engagement

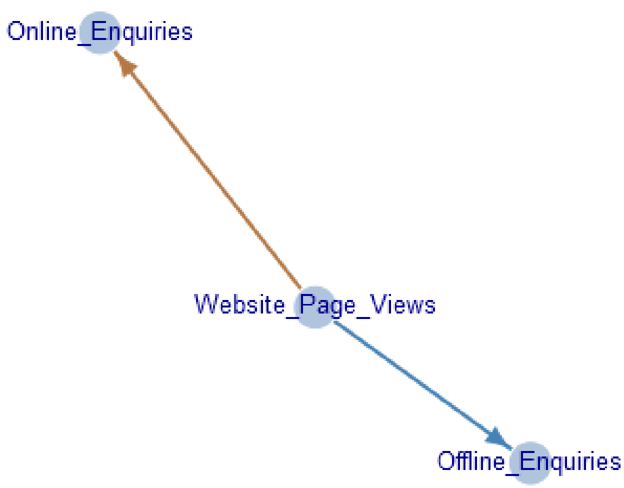


Path	Path Length	Effect Type	Total IRF
3 - Facebook Page Posts Engagement -> Online Enquiries	2	Direct	0.4648
1 - Facebook Page Posts Engagement -> Website Page Views -> Offline Enquiries	3	Indirect	0.0196
2 - Facebook Page Posts Engagement -> Website Page Views -> Online Enquiries	3	Indirect	0.0625

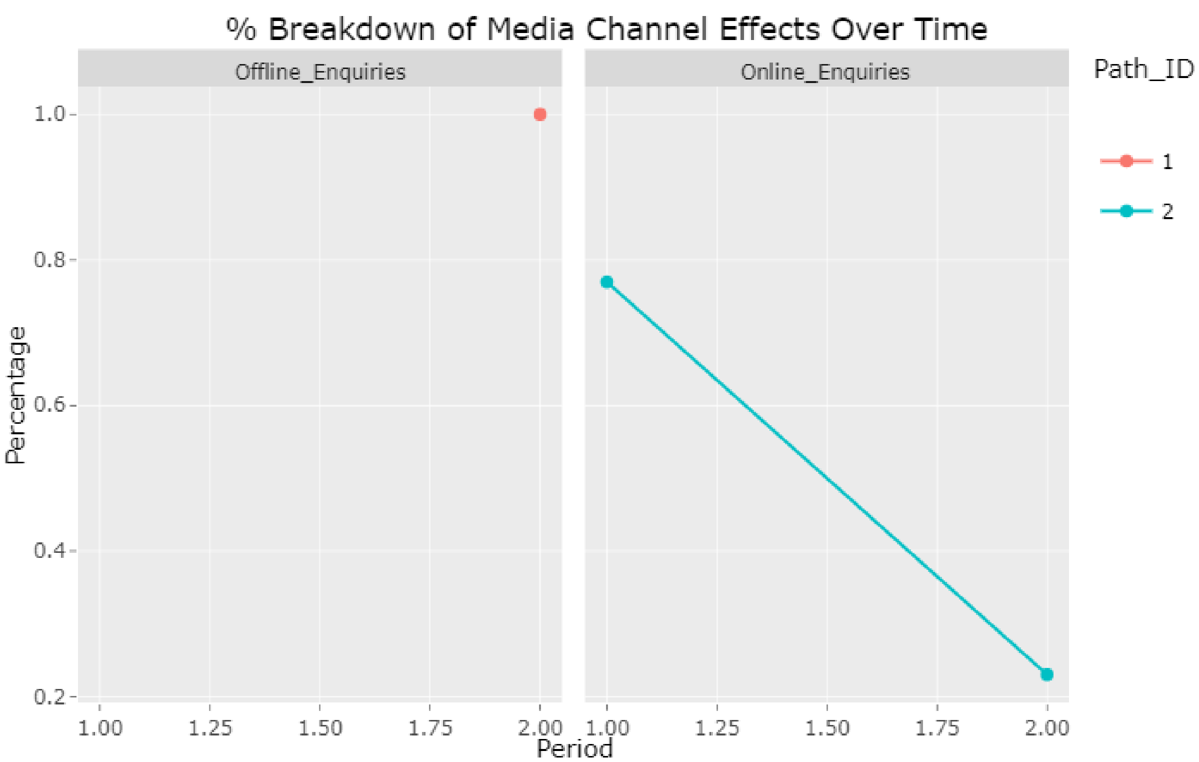

% Breakdown of Media Channel Effects Over Time
Offline_Enquiries
Online_Enquiries
Path_ID
1
2
3
Percentage
0.00
0.05
0.10
5
10
15
Period


% Breakdown of Total Media Channel Effects Over Time
Offline_Enquiries
Online_Enquiries
Percentage
0.00
0.05
0.10
5
10
15
Period

Website Page Views



Path	Path Length	Effect Type	Total IRF
2 - Website Page Views -> Online Enquiries	2	Direct	0.2093
1 - Website Page Views -> Offline Enquiries (all of the effect occurs in one period)	2	Direct	0.0656



Appendix 2 – Summary Table of Literature in Measuring Advertising Media Effectiveness (Chapter 3)

The table below places the research conducted in Chapter 3 within the context of the extant marketing literature.

Research	**Model Type**	**Data Frequency**	**Data Timeframe**	**Data Description**	**Number of Media Channels**	**Media Channels**
Dekimpe and Hanssens (1995)	VAR	Monthly	76 months	Large home-improvement chain	2	Print, TV/radio
Dekimpe and Hanssens (1999)	VECM	Monthly	5 years	Market performance from a pharmaceutical company. Second dataset from packaged food product	1	Overall marketing expenditure
Naik and Peters (2009)	Hierarchical Linear	Weekly	86 weeks	Visits to car dealer or its website (need to leave details), advertising expenditure	6	TV, radio, magazines, newspapers, display, direct mail
Srinivasan et al. (2010)	VAR	Every 4 weeks	7 years	French data from Promethee, a brand performance tracker developed by Kantar Worldpanel. Data on 74 brands from four product categories	1	Total advertising expenditure
Wiesel et al. (2011)	VAR	Daily	2.5 years	Inofec BV - sells furniture products	5	Direct mail, fax, catalogue, email, paid search
Danaher and Dagger (2013)	Type II Tobit	Daily	4 weeks	Sales data and self-reported media exposure of 3007 loyalty program customers of an upscale Australian department store	10	TV, newspapers, radio, magazines, catalogue, direct mail, display, paid search, social media (how often they visited the company's page), e-mail
Dinner et al. (2014)	System of Linear Equations	Weekly	103 weeks	Major US clothing retailer (25 different markets) – 85% offline sales	3	Traditional (aggregate spend on newspapers, magazines, radio, TV, billboard), display, paid search

Research	Model Type	Data Frequency	Data Timeframe	Data Description	Number of Media Channels	Media Channels
Frison et al. (2014)	ECM	Monthly	6.5 years	Monthly – 6.5 years - Sales and advertising expenditure data on 261 brands in the consumer goods (CPG) market	6	TV, radio, newspapers, magazines, billboard, cinema
de Haan et al. (2016)	SVAR	Daily	385 days	Major European online retailer (similar to Amazon), data across 5 product categories	9	Includes email, paid search, retargeting, referrals, portals, price-comparison websites, TV, radio
Kireyev et al. (2016)	VECM	Weekly	1 year	Large commercial bank operating mainly in the southern U.S.	2	Display, paid search
Pauwels et al. (2016b)	BVAR	Weekly	2-4 years (depends on brand)	4 brands – scholastic test preparation, family-run office furniture supplier, travel, retail	2-5 (depends on brand)	Includes website, paid search, TV, catalogues
Pauwels et al. (2016a)	VAR	Weekly	55 weeks	Major US apparel retail brand, eWOM data from Crimson Hexagon	6	Paid search, print, organic search, TV, radio, eWOM
Srinivasan et al. (2016)	VAR	Weekly	40 weeks	Large consumer packaged goods manufacturer	4	Paid search, social media (Facebook), TV, website
This Research	VAR	Daily	394 days	Luxury travel brand	9	Facebook ads, Facebook page posts, display, paid search, video, website, outdoor, magazine and TV

Appendix 3 – Empirical Sources in Quantifying Advertising Media Effectiveness (Chapter 2)

The table below provides details about the empirical sources in the systematic literature review conducted in Chapter 2. The sources are grouped by the four major modelling approaches identified in the literature.

Reference	**Modelling Approach**	**Model Estimation Method**	**Model Type**	**Data Frequency**	**Data Timeframe**	**Data Description**	**Number of Media Channels**	**Media Channels**
Frison et al. (2014)	Persistence	Frequentist	ECM	Monthly	6.5 years	Monthly – 6.5 years - Sales and advertising expenditure data on 261 brands in the consumer goods (CPG) market	6	TV, radio, newspapers, magazines, billboard, cinema
Wiesel et al. (2011)	Persistence	Frequentist	VAR	Daily	2.5 years	Inofec BV - sells furniture products	5	Direct mail, fax, catalogue, email, paid search
Srinivasan et al. (2016)	Persistence	Frequentist	VAR	Weekly	40 weeks	Large consumer packaged goods manufacturer	3	Paid search, social media (Facebook), television
Kireyev et al. (2016)	Persistence	Frequentist	VECM	Weekly	1 year	Large commercial bank operating mainly in the southern U.S.	2	Display, paid search
de Haan et al. (2016)	Persistence	Frequentist	SVAR	Daily	385 days	Major European online retailer (similar to Amazon), data across 5 product categories	9	Includes email, paid search, retargeting, referrals, portals, price-comparison websites, TV, radio

Reference	Modelling Approach	Model Estimation Method	Model Type	Data Frequency	Data Timeframe	Data Description	Number of Media Channels	Media Channels
Pauwels et al. (2016b)	Persistence	Bayesian	BVAR	Weekly	2-4 years (depends on brand)	4 brands – scholastic test preparation, family-run office furniture supplier, travel, retail	2-5 (depends on brand)	Depends on brand - includes website, paid search, TV, catalogues
Pauwels et al. (2016a)	Persistence	Frequentist	VAR	Weekly	55 weeks	Major US apparel retail brand, eWOM data from Crimson Hexagon	6	Paid search, print, organic search, TV, radio, eWOM
Gallego et al. (2019)	Customized	Bayesian	Bayesian Structural Time Series	Weekly	4.5 years	Country-wide franchise of fast food restaurants	7	Out-of-home (billboards), radio, TV, online, search, press and cinema
Jin et al. (2017)	Customized	Bayesian	Bayesian Linear	Weekly	2.5 years	Advertising for shampoo advertiser – compiled by Neustar MarketShare	5	TV, magazines, display, YouTube and paid search
Wang et al. (2017)	Customized	Bayesian	Hierarchical Bayesian (one for each product category)	Weekly	2.5 years	Advertising for shampoo and soda product categories – compiled by Neustar MarketShare	7	TV, radio (soda category) magazines, display, YouTube, paid search (shampoo category) and other
Dinner et al. (2014)	Customized	Frequentist	System of Linear Equations	Weekly	103 weeks	Major US clothing retailer (25 different markets) – 85% offline sales	3	Traditional (aggregate spend on newspapers, magazines, radio, TV, billboard), display, paid search

Reference	Modelling Approach	Model Estimation Method	Model Type	Data Frequency	Data Timeframe	Data Description	Number of Media Channels	Media Channels
Joo et al. (2014)	Customized	Frequentist	ARMAX	Hourly	92 days	Number of Google searches containing certain keywords (set of chosen generic/branded words) related to financial firms/products, TV data from Kantar, records each advertisement and an estimated cost	2	TV, organic search
Liaukonyte et al. (2015)	Customized	Frequentist	Linear	TV advertisement insertion	1 year	TV advertising (Kantar Media), internet searches, survey giving information about advertising content (from survey)	3	TV, direct website, search engine referral
Hanssens et al. (2014)	Customized	Frequentist	Cross-Effects and Longitudinal Hierarchical Linear	Monthly	7.5 years	6 major brands, consumer attitude data on 8000 households in France	1	Overall marketing expenditure
Hanssens et al. (2016)	Customized	Frequentist	Linear	Weekly	2 years	6 leading US brands in the digital single-lens reflex (DSLR) camera market. Product reviews from Amazon	2	Product reviews, other advertising
Lovett et al. (2019)	Customized	Frequentist	Linear	Monthly	6.5 years	Information on 538 U.S. national brands from 16 product categories	3	TV, internet and other (print, radio and outdoor)
Kumar et al. (2017)	Customized	Frequentist	Time-Varying Effects (TVEM)	Weekly	3 years	Data across 5 distribution channels for top 6 brands of a CPG company that produces and distributes ice-cream across the US	2	Facebook impressions, TV

Reference	Modelling Approach	Model Estimation Method	Model Type	Data Frequency	Data Timeframe	Data Description	Number of Media Channels	Media Channels
Osinga et al. (2010)	Customized	Frequentist	Dynamic Linear Model (Kalman Filter)	Monthly	8 years	Sales and advertising data related to 89 prescription drugs (brands) introduced in the US between 1993 and 2000	3	Total advertising expenditure - direct to consumer advertising, direct to physician, other (e.g. journal advertising)
Kolsarici and Vakratsas (2018)	Customized	Frequentist	Time-Series Multivariate Adaptive Regression Splines (TS-MARS)	Monthly (hybrid cars), weekly (yoghurt and beer)	5.5 years (hybrid cars), 11 years (yoghurt and beer)	Sales data from durable (hybrid cars - J.D. Power and Associates) and packaged goods (yoghurt and beer - IRI) categories in the US market, advertising expenditure data from Kantar media	6	TV, magazine, internet, outdoor, newspaper, radio
Saboo et al. (2016)	Single Source	Frequentist	Time-Varying Effects (TVEM)	Monthly	3 years	Large Fortune 500 retailer selling a wide assortment of products related to home improvement, gardening needs, furniture and home appliances	2	Email, direct mail
Danaher and Dagger (2013)	Single Source	Frequentist	Type II Tobit (Probit model for purchase incidence, Tobit model for purchase outcome – log of dollar sales or profit)	Daily	4 weeks	Sales data and self-reported media exposure of 3007 loyalty program customers of an upscale Australian department store	10	TV, newspapers, radio, magazines, catalogue, direct mail, display, paid search, social media (how often they visited the company's page), e-mail

Reference	Modelling Approach	Model Estimation Method	Model Type	Data Frequency	Data Timeframe	Data Description	Number of Media Channels	Media Channels
Kumar et al. (2016)	Single Source	Bayesian	Hierarchical Bayesian	Weekly	170 weeks	Survey answers, social media participation and offline sales of loyalty card customers of a speciality wine retailer	3	TV, email, social media (owned media, platform not specified)
Bollinger et al. (2013)	Single Source	Bayesian	Hierarchical Bayesian (Multivariate Logit)	Daily	2 years	Advertising exposure, demographic and purchase information for over 3000 households recruited by a Fast Moving Consumer Goods (FMCG) company	3	TV, display, social media (Facebook – owned and earned media)
Zantedeschi et al. (2016)	Single Source	Bayesian	Hierarchical Bayesian (Type I Tobit)	Daily	26 months	Advertising exposure, purchases of 300 customers of a multichannel speciality retailer	2	Email, catalogue
Feit et al. (2013)	Single Source	Bayesian	Hierarchical Bayesian (Multivariate Logit)	Daily	1 month	Digital media usage for a random sample of 2,000 ESPN users based primarily in North America for the 2010 FIFA World Cup	4	ESPN website, ESPN3 streaming video, ESPN Mobile, television
Song et al. (2016)	Single Source	Frequentist	Clustered Generalized Multivariate Autoregressive (CGMAR)	Online web session	2 years	9,805 customers of a North American specialty retailer selling women's clothing, accessories, and home items	2	Email, catalogue
Godfrey et al. (2011)	Single Source	Frequentist	System of linear equations	Quarterly	13 quarters	Customer survey for 1,162 customers from a car dealership, combined with customer repurchase behaviour	3	Phone, email, direct mail

Reference	Modelling Approach	Model Estimation Method	Model Type	Data Frequency	Data Timeframe	Data Description	Number of Media Channels	Media Channels
Danaher and Rossiter (2011)	Single Source	Frequentist	Linear (with random effects)	N/A	N/A	Survey with hypothetical scenarios - completed by 1,550 people	11	TV, radio, magazines, newspapers, catalogues, direct mail, email, SMS, door-to-door visits, telemarketing
Joo et al. (2016)	Single Source	Frequentist	Binomial	Hourly	3 months	180,000 randomly selected users of AOL.com	2	TV, organic search
Lin et al. (2013)	Single Source	Frequentist	Consumer-level demand model	Half-hourly	7 days	Universal McCann's Media in Minds Diary - 1,775 people in US who reported their media exposure	4	TV, newspapers or magazines, radio, internet
Fischer (2019)	Single Source	Frequentist	Ordered Logistic	Weekly	10 weeks	Survey data from TNS representative online panel (collected on behalf of Mercedes-Benz). More than 800 target customers	3	TV, print, online
Shao and Li (2011)	Attribution	Frequentist	Binomial Response - Logistic, Probabilistic	Individual touchpoints	4 weeks	72.5 million anonymous users randomly sampled from an advertising campaign of a consumer software and services company	5	Paid search, display, social media, email and video
Dalessandro et al. (2012)	Attribution	Frequentist	Game Theory/Probabilistic (implemented with a binomial response model)	Individual touchpoints	Dataset 1 - 2 weeks, Dataset 2 – not stated	Dataset 1 - an actual advertiser's online advertisement campaign. Dataset 2 – several advertisement campaigns internal to M6D3 (digital advertising company)	7 for Dataset 1, 1 for Dataset 2	Not stated for Dataset 1, retargeting for Dataset 2

Reference	Modelling Approach	Model Estimation Method	Model Type	Data Frequency	Data Timeframe	Data Description	Number of Media Channels	Media Channels
Li and Kannan (2014)	Attribution	Bayesian	Bayesian	Individual touchpoints	68 days	68 days – customer path data of 1997 visitors to the website of a firm in the hospitality industry	6	Organic search, paid search, referral, website, email, display
Anderl et al. (2016)	Attribution	Frequentist	Markov	Individual touchpoints	Not stated	4 clickstream datasets provided by online advertisers. Each dataset includes a minimum of 405,000 journeys per advertiser	7-8 (depends on dataset)	Depends on dataset - website, paid/organic search, price comparison website, affiliate, referrer, social media (Facebook display ads), retargeting, email, other
Xu et al. (2014)	Attribution	Frequentist	Mutually Exciting Point Process	Individual touchpoints	4 months	Random sample of 12,000 customers of major manufacturer and vendor of consumer electronics	3	Paid search, display and other (e.g. affiliate)
Abhishek et al. (2012)	Attribution	Frequentist	Hidden Markov	Individual touchpoints	11 weeks	6,432 users who participated in an online campaign for a car manufacturer. Data provide by digital advertising agency	3	Display, paid search, website
Zhang et al. (2014), Ji et al. (2016), Ji and Wang (2017), Hou et al. (2016)	Attribution	Frequentist	Survival	Individual touchpoints	2 months	Dataset openly published by Miaozhen Systems. It includes about 380 million cookies collected from PCs and mobile devices	46	Not stated
Yin et al. (2016)	Attribution	Frequentist	Genetic Data Mining Algorithm	Individual touchpoints	6 months	6,760 customer journeys	Not stated	Not stated

Reference	Modelling Approach	Model Estimation Method	Model Type	Data Frequency	Data Timeframe	Data Description	Number of Media Channels	Media Channels
Yadagiri et al. (2015)	Attribution	Frequentist	Game Theory/Probabilistic	Individual touchpoints	2 months for travel company, 100 days for retailer	Travel and Experience organisation - 1.5m users, ecommerce retailer – 0.4m users	9-10 (depends on dataset)	Depends on dataset - include website, display, email, referrals, social media, organic search, paid search
Sinha et al. (2014)	Attribution	Frequentist	Binomial Response - Logistic	Individual touchpoints	See right	Same as Yadagiri et al. (2015)	9	Includes website, display, email, referrals, social media, organic search, paid search
Wooff and Anderson (2013)	Attribution	Frequentist	Time Weightage (based on Beta distribution)	Individual touchpoints	Not stated	58,867 customer journeys from major UK online retailer	9	Includes affiliates, display, price comparison sites, referrals, organic search, paid search
Geyik et al. (2014)	Attribution	Frequentist	Probabilistic	Individual touchpoints	12 days	Two campaigns	4	Not stated
Nottorf (2014)	Attribution	Bayesian	Bayesian Mixture of Normals (Logistic)	Individual touchpoints	1 month	Several thousand customer journeys from an online shop	5	Search (paid search and organic), price-comparison site, retargeting, website, other (e.g. social media)

Reference	Modelling Approach	Model Estimation Method	Model Type	Data Frequency	Data Timeframe	Data Description	Number of Media Channels	Media Channels
Danaher and Van Heerde (2018)	Attribution	Frequentist	Binomial Response - Probit (generic model presented, Probit just used for their data)	Daily	1 year	3 brands from North American specialty retailer – several thousand customer journeys for each brand	6	Referrals, organic search, paid search, social media (mainly Facebook), email, catalogue
Brodersen et al. (2015)	Attribution	Bayesian	Bayesian Structural Time Series	Individual touchpoints	6 weeks	Advertising campaign run by one of Google's advertisers in the US	Not stated	Not stated
Kaatz et al. (2019)	Attribution	Frequentist	Markov	Individual touchpoints	1 month	German fashion retailer selling leather products	7	Direct, paid search, organic search, referral, display, social media (Facebook), email
Nisar and Yeung (2018)	Attribution	Frequentist	Probabilistic	Individual touchpoints	2 years	Large online sales platform	8	Display, organic search, paid search, price-comparison (referral), retargeting, social media, direct, email
Kakalejcik et al. (2019)	Attribution	Frequentist	Markov	Individual touchpoints	8 months	Data from Google Merchandise Store	7	Direct, referral, organic search, paid search, display, social media (Facebook, Twitter, LinkedIn, etc), other

Reference	Modelling Approach	Model Estimation Method	Model Type	Data Frequency	Data Timeframe	Data Description	Number of Media Channels	Media Channels
Yuvaraj et al. (2018)	Attribution	Frequentist	Logistic, Probabilistic	Individual touchpoints	Not stated	Anonymised dataset from Adobe	6	Organic search, paid search, display, social media, email, others
Ren et al. (2018)	Attribution	Frequentist	Dual-Attention Recurrent Neural Network	Individual touchpoints	Dataset 1 – 2 months, Dataset 2 – 1 month	Dataset 1 – Miaozhen, Dataset 2 - Criteo	Dataset 1 – 9, Dataset 2 - 10	Dataset 1 - social, paid search, portal, video, vertical community, unknown, union and music, Dataset 2 – not stated
Li et al. (2019)	Attribution	Frequentist	Two-Stage Choice Model	Individual touchpoints	1 year	Dataset from a leading U.S. media measurement and analytics company, who manage a large-scale consumer panel	4	Paid/organic search, email, display/referral, direct
Tinyakova et al. (2018)	Attribution	Frequentist	Markov	Individual touchpoints	Not stated	Not stated	Not stated	Not stated
Zhao et al. (2019)	Attribution	Frequentist	GAM	Individual touchpoints	3 months	Simulated dataset and real world dataset – source not stated	18	Not stated