UNIVERSITY OF SÃO PAULO
SCHOOL OF ECONOMICS, ADMINISTRATION, ACCOUNTING AND ACTUARIAL
DEPARTAMENT OF ADMINISTRATION
POSTGRADUATE PROGRAM IN ADMINISTRATION

Eduardo Ogawa Cardoso

Privacy-By-Design Attribution Model:
A Relative Weight And Spillover Effect Approach

São Paulo
2023

EDUARDO OGAWA CARDOSO

Privacy-By-Design Attribution Model:
A Relative Weight And Spillover Effect Approach

Original Version

Dissertation submitted to the Postgraduate Program in Administration of the Department of Administration of the Faculty of Economics, Administration, Accounting and Actuarial of the University of São Paulo, as a partial requirement for obtaining the title of Master of Science.

Concentration area: Marketing

Advisor: Prof. Phd. Otávio Bandeira De Lamônica Freire

São Paulo

2023

Prof. Dr. Carlos Gilberto Carlotti Júnior

Reitor da Universidade de São Paulo

Profa. Dra. Maria Dolores Montoya Diaz

Diretora da Faculdade de Economia, Administração, Contabilidade e Atuária

Prof. Dr. João Maurício Gama Boaventura

Chefe do Departamento de Administração

Prof. Dr. Felipe Mendes Borini

Coordenador do Programa de Pós-Graduação em Administração

Catalogação na Publicação (CIP)
Ficha Catalográfica com dados inseridos pelo autor

Cardoso, Eduardo Ogawa.
Privacy-By-Design Attribution Model: A Relative Weight And Spillover Effect Approach / Eduardo Ogawa Cardoso. - São Paulo, 2023.
73 p.

Dissertação (Mestrado) - Universidade de São Paulo, 2023.
Orientador: Otávio Bandeira De Lamônica Freire.

1. Administração. 2. Marketing. 3. Econometria. I. Universidade de São Paulo. Faculdade de Economia, Administração, Contabilidade e Atuária. II. Título.

Name: Cardoso, Eduardo Ogawa

Title: Privacy-By-Design Attribution Model: A Relative Weight And Spillover Effect Approach

Dissertation submitted to the Postgraduate Program in Administration of the Department of Administration of the Faculty of Economics, Administration, Accounting and Actuarial of the University of São Paulo for obtaining the title of Master of Science.

Approved in : __/__/____

Examination Board

Prof. Phd.: ____

Institution: ____

Evaluation: ____

Prof. Phd.: ____

Institution: ____

Evaluation: ____

Prof. Phd.: ____

Institution: ____

Evaluation: ____

ACKNOWLEDGEMENTS

My sincere appreciation goes out to the entire Department of Administration of the Faculty of Economics, Administration, Accounting and Actuarial of the University of São Paulo, whose collective wisdom and feedback have significantly enriched this work. Their commitment to fostering academic excellence and nurturing curious minds has been a constant source of inspiration.

I would like to express my profound gratitude to my advisor Otávio Bandeira De Lamônica Freire, whose insights, expertise, and unwavering support have been invaluable throughout this research journey. The depth and breadth of knowledge imparted by you have been instrumental in shaping my academic and professional trajectory.

I am deeply thankful to my parents, Eduardo and Sandra, and my sister Julia, for their enduring love, encouragement, and belief in my potential. Your sacrifices and resilience have paved the way for my achievements, and this dissertation stands as a testament to the values and virtues you instilled in me.

To Maria Carolina Cavalcante Dias, I extend a heartfelt thank you. Your unwavering support, both academically and personally, has been a beacon throughout this rigorous journey. It's been a privilege to work alongside such a dedicated and brilliant individual.

Lastly, I would like to thank all those who directly or indirectly contributed to the successful completion of this dissertation. Every conversation, critique, and word of encouragement has played a pivotal role in this scholarly endeavor.

Thank you for being a part of this journey.

ABSTRACT

Cardoso, E. O. (2023). Privacy-By-Design Attribution Model: A Relative Weight And Spillover Effect Approach (Master's Dissertation). Faculty of Economics, Administration, Accounting and Actuarial, Postgraduation Program in Administration, University of São Paulo, São Paulo.

This thesis presents the Privacy-by-design Attribution Model (PBDAM), a novel approach to multi-channel attribution modeling that addresses the challenges of capturing hidden interactions and ensuring equitable credit assignment while maintaining consumer privacy. PBDAM's performance is rigorously assessed through a comprehensive examination of both synthetic and actual empirical data, underscoring its superiority over conventional multi-channel attribution methods. Through analyzing a synthetic dataset with aggregated daily data and evaluating actual campaign data from major companies in Brazil, PBDAM demonstrates its ability to effectively capture hidden interaction information, generate fair attribution values, and provide valuable insights without compromising consumer privacy. The findings of this research establish PBDAM as a superior model in terms of performance, equitable credit assignment, and privacy-consciousness.

Keywords: Multi-Channel, Attribution, Privacy, Spillover Effect, Relative Weight

RESUMO

Cardoso, E. O. (2023). Modelo de Atribuição Privacy-By-Design: Uma abordagem com Relative Weight e Efeito Spillover. (Dissertação de Mestrado). Faculdade de Economia, Administração, Contabilidade e Atuária, Programa de Pós-Graduação em Administração, Universidade de São Paulo, São Paulo.

Esta tese introduz o Modelo de Atribuição Privacy-by-design (PBDAM), uma inovação na modelagem de atribuição multicanal que resolve os desafios de identificar interações ocultas e assegurar uma distribuição justa de créditos, tudo isso preservando a privacidade do consumidor. O PBDAM foi rigorosamente avaliado com base em conjuntos de dados sintéticos e empíricos, evidenciando sua superioridade em comparação com abordagens tradicionais. A análise de um conjunto de dados sintéticos e a avaliação de campanhas de grandes empresas brasileiras permitiram ao PBDAM demonstrar sua eficácia na captura de informações de interação ocultas, na geração de avaliações de crédito equitativas e na entrega de insights relevantes, sem abrir mão da privacidade. As conclusões alcançadas confirmam o PBDAM como um modelo avançado, com destaque em performance, equidade na atribuição de créditos e responsabilidade com a privacidade.

Palavras-chave: Atribuição, Multicanal, Privacidade, Efeito Spillover, Relative Weight

FIGURES LIST

TABLES LIST

FORMULAS LIST

SUMMARY

1 INTRODUCTION

In the current era, technological advancements harnessed the power to deliver tailored advertisements directly to consumers. As a result, this has led to a significant shift towards digital marketing, becoming one of the most widely adopted and profitable strategies (Leguina et al., 2020). However, as digital landscapes grow more complex, businesses must prioritize authenticity, honesty, and transparency to foster consumer trust (Kotler et al., 2017). Therefore, understanding this dynamic setup underlines the need for businesses to embed these values in their operations to thrive in the digital marketing era.

According to Li and Kannan (2017), consumers typically navigate across **multiple digital channels** before finalizing a purchase or subscription. These interactions called 'touchpoints,' signify encounters with channels that may not always lead to a website visit. Consumers interact with and are influenced by multiple touchpoints before reaching their end goal. This sequence of touchpoints is referred to as the 'consumer journey.' The unique combination of touchpoints for an individual consumer is considered a 'path.' Hence, these 'paths' highlight consumers' unique journeys, emphasizing the importance of understanding these interactions.

Regardless of industry-specific variations in consumer behavior patterns, businesses have realized that increasing the number of touchpoints or message frequency does not always lead to an improved experience or a noticeable impact on consumer purchases. Instead, precise and effective communication with the audience is essential. To achieve this, businesses have started mapping the entirety of the consumer journey and analyzed the data to understand how consumers engage across multiple marketing channels and determine the influence of each channel on the final consumer choice (Kotler et al., 2017).

Additionally, assessing the return on ad spend (ROAS) presents a significant challenge for advertisers using different media channels to reach clients. Market response models are valuable tools for advertisers to comprehend the sales outcomes resulting from different marketing or media mix strategies. This is achieved through observational data, encompassing various factors such as price, media expenditure, sales, and economic indicators. Importantly, these models can distinguish between exogenous variables, such as market expansion or demographic changes, and endogenous factors, such as promotional strategies or advertising expenditures. Therefore, utilizing of these models provides valuable perspectives on the efficacy of different marketing strategies in a range of market environments (Jin et al., 2017).

These models typically produce solid causal outcomes, such as return on advertising expenditure and budget allocation strategies, and aid in understanding historical media effectiveness. However, in order to accurately measure the influence of marketing or media mix, it is necessary first to comprehend the value of each channel (touchpoint) in the consumer journey. This process, called attribution modeling, employs advanced analytics to appropriately assign credit for desired consumer actions to each marketing touchpoint across all online and offline channels (Buhalis & Volchek, 2021).

According to the Marketing Science Institute (2022), attribution poses a significant research challenge in the marketing field, demanding further exploration. Hence, attribution modeling plays a crucial role in understanding the impact of each touchpoint in a multi-channel and cross-platform environment. This modeling approach can address various marketing issues, particularly those related to advertising effectiveness, revenue, and return on investment (Berman, 2018). It is especially valuable for media investment strategies and gaining insights into consumer behavior across different channels and platforms (Li & Kannan, 2014).

Besides attribution modeling, there are several methodologies for establishing causal relationships between channels and business outcomes, such as A/B testing. However, as Kohavi e Longbotham (2015) discussed, conducting randomized and controlled A/B tests in complex environments like online advertising can be challenging, often resulting in unreliable or time-consuming results.

Due to the uncertainty and imprecision of non-controlled tests, heuristic attribution models have become a relied-upon solution. Heuristic models involve a set of simplistic rules for distributing credit from a specific business outcome, such as sales or revenue, to a particular channel. Examples of heuristic models include last-touch, first-touch, time-decay, and linear models. The last-touch model assigns full credit to the last touchpoint, the first-touch model assigns full credit to the first touchpoint, the time-decay model employs an arbitrary rule based on time, where the most recent touchpoint receives the most credit which diminishes with time, and the linear model assigns equal credit to all touchpoints in the consumer journey.

The rules underlying these heuristic models often overlooks the specificity of the customer journey or disregard certain touchpoints. While the fairness of these models can be questioned, they are widely used in the digital advertising industry due to their simplicity and ease of understanding. The most prevalent heuristic model today is the last-touch attribution model, which attributes credit solely to the last touchpoint a consumer interacts with before taking the desired action, disregarding all prior interactions (Google Inc., 2021).

As mentioned before, consumers typically engage with multiple touchpoints, making it reasonable to incorporate all of these interactions in the model to represent actual consumer behavior accurately. This can help the marketers in the budget allocation and decision-making processes. Fortunately, we can track nearly every consumer interaction through digital channels. Numerous studies have shown that these tracking techniques provide ample data for constructing or utilizing algorithmic attribution models, which leverage historical data and complex algorithms to measure or infer attribution based on actual conversion performance rather than relying on human-defined rules and heuristics (Abhishek et al., 2012; Anderl et al., 2016; Berman, 2018; Leguina et al., 2020; Zhao et al., 2019)

Solutions incorporating economics and computer science inputs have become increasingly common in the current multidisciplinary landscape. This is facilitated by advancements in computer power over the past decade and the necessity of comprehensively understanding the consumer journey. Researchers have thus focused on developing more sophisticated models to allocate credit across all touchpoints, giving rise to multi-touch attribution models, also known as multi-channel attribution models (Anderl et al., 2016; Berman, 2018; Leguina et al., 2020; Li & Kannan, 2014).

Academics have proposed several multi-touch attribution models, including logistic regression models (Shao & Li, 2011), game theory-based approaches(Berman, 2018; Dalessandro et al., 2012), Bayesian models (Kannan & Li, 2017), mutually exciting point process models, VAR models (Kireyev et al., 2016), and hidden Markov models (Abhishek et al., 2012). In addition to these academic contributions, various internet-based industry firms, such as Google, Adobe, and Visual IQ, have developed heuristic and algorithmic attribution approaches based on these studies.

While multi-touch attribution models appear promising alternatives to traditional heuristic models, it is important to acknowledge the potential intrusiveness and disempowerment associated with data collection methods used in these models. Sophisticated techniques often rely on analyzing specific paths of the consumer journey, which typically involve consumer identification through personal data to accurately replicate interaction sequences (Berman, 2018). However, using personally identifiable information (PII) to target consumers can be highly intrusive, particularly if employed without their consent or knowledge. Additionally, consumers often lack control over the data collected, its use, and its sharing.

The General Data Protection Regulations (GDPR) were established to address these concerns. The GDPR encompasses principles aimed at giving consumers more control over their data and emphasizes the privacy-by-design principle, which advocates for incorporating privacy considerations into system design from the outset (Shovon et al., 2019).

Numerous studies have demonstrated that non-personally identifiable information (non-PII) or aggregated data can serve as viable sources for attribution modeling (Abhishek et al., 2012; Berman, 2018; Chapelle, Manavoglu & Rosales, 2014; Dalessandro et al., 2012; Liu et al., 2014; Shao & Li, 2011; Zhao et al., 2019). Fortunately, this type of information is readily accessible through media distribution platforms and does not necessitate consumer identification.

Furthermore, a significant body of research explores the relationship and synergy between different advertising channels regarding effectiveness. Data related to ad clicks, ad frequency, and ad impressions can unveil the consumer's indirect purchase intent and reveal hidden relationships between channels. Thus, aside from the sequence of engaged channels, it is reasonable to assume that other factors, such as the hidden synergy between channels, can influence the decision-making process in this context.

Unintuitively, there is a significant gap in the existing literature as no studies, to the best of our knowledge, have addressed or incorporated the relationships between touchpoints and their impact on actual business results within the attribution model. This omission highlights the need for further research in this area to enhance our understanding of the complex dynamics involved.

Additionally, we share a deep concern for consumers' data security and privacy. Consequently, relying on personally identifiable information (PII) or consumer-level data to support our attribution models is no longer viable. As a result, this issue raises an important question**. Is it possible to develop an attribution model that respects and protects consumer's privacy?**

This study aims to develop a privacy-by-design attribution model, addressing the lack of emphasis on consumer privacy and the hidden relationship between different interactions in most multi-touch attribution models. By presenting a privacy-by-design strategy, we intend to establish a mathematical framework based on aggregated daily data, including clicks, views and costs for each channel, to gain insights into business outcomes and desired consumer actions.

Unlike other techniques, this method does not rely on path sequence. Instead, we presume that the interaction between touchpoints is not processed in a serialized, linear manner. Our hypothesis is based on the assumption that interactions across different channels can mutually influence each other and have an impact on future interactions. We also propose a quantitative synergy between various channels and their associated metrics, recognizing that despite their high connectivity, all these metrics are essential to achieving ultimate business outcomes.

This interdependence can be measured using the spillover index, a mathematical function (Diebold & Yilmaz, 2012). Expanding on this idea, we consider each metric (clicks, views, costs, and conversions) from each channel as an individual entity, with distinct interactions and synergies that influence one another and future interactions.

When dealing with highly correlated data in analyzing online channels, such as cost, views, and clicks, it is not uncommon to encounter collinearity issues. Thus, in addition to the spillover index, we propose calculating each channel's contribution to the total predicted sales variance using a statistical function known as Relative Weight Analysis Johnson, 2000). This allows us to determine the relative importance of the channels when considered individually and in combination with other channels.

Finally, in order to address the challenge of integrating the spillover index and each channel's relative weight, we have opted for a simple yet elegant approach. We place the Spillover Index and the Relative Weight as orthogonal vectors in a cartesian plan and calculate PBDAM as the resultant using the Pythagorean Theorem - that way, we chose to prioritize the explicability and accessibility of the model.

The Privacy-by-design Attribution Model (PBDAM) is based on a combination of well-established approaches from marketing science, finance, and natural science. From a theoretical and scientific standpoint, our study provides a way to combine two different mathematical functions related to attribution modeling (spillover and relative weights), which could establish a new standard for understanding the advertising effectiveness landscape.

From a practical perspective, this new methodology may offer marketers and managers a reliable means to assess the performance of their company's marketing strategies. Adopting this newly developed attribution model can bring precise financial and strategic benefits while also taking steps toward restoring public trust in online advertising and ensuring compliance with GDPR regulations. Ultimately, we aim to assist managers in making better data-driven decisions while ensuring consistent performance and respecting the privacy of billions of consumers worldwide.

In Chapters 3 to 4, we present these contributions, describing our model specifications alongside several simulations and empirical results. In Chapter 2, our focus is primarily on reviewing prior scientific research related to attribution modeling and providing the relevant background on current privacy issues.

2 LITERATURE REVIEW AND RELATED WORK

Today's Chief Marketing Officers (CMOs) face tremendous pressure to demonstrate the impact of their marketing budgets on their firm's business goals. With advertising and marketing budgets continuously increasing, advertisers and organizations need to measure the effectiveness of their media spending concerning key performance indicators (KPIs) and develop effective budget allocation strategies (Harmeling et al., 2017).

In addition to navigating non-traditional channels, such as digital marketing, CMOs must navigate an environment where consumers are exposed to multiple touchpoints and are becoming more knowledgeable. The interaction between these touchpoints varies across markets, industries, and enterprises, but the existence of synergy among them is undeniable, as shown in various studies (Abhishek et al., 2012; Berman, 2018; Flavián et al., 2020). Furthermore, data privacy regulations have made data collection and processing more challenging, adding to the situation’s complexity.

Traditionally, marketers have relied on market response models to understand the effectiveness of their marketing efforts. However, a crucial aspect of this process is the distribution of conversion credit to all channels involved, which is known as attribution. Therefore, this literature review aims to shed light on the development of attribution models as part of market response models. It will also explore the development and implications of new privacy legislation in this ever-changing environment, as privacy considerations play a crucial role in contemporary marketing practices.

2.1 MARKET RESPONSE MODELS

Every brand and product category has a sales-driving technique through which corporations communicate with the market, distribute products and services, and establish prices. The response of market consumers to these brand actions is reflected in sales, which in turn inform future marketing activities based on current and historical market knowledge (Hanssens et al., 2003).

A response model represents the relationship between a dependent variable and one or more independent variables. The dependent variable can be sales, market share, consumer awareness, or any other factor relevant to marketing managers. In marketing, response models primarily focus on the relationship between controllable marketing mix variables and performance indicators, such as sales, which serve as outcomes of marketing plans (Hanssens et al., 2003).

Explanatory variables are those believed to influence the dependent variable, and together with the dependent variable, they form the system of equations used to model market behavior. When these models incorporate competitive reaction functions, vertical market structures, cost functions, or other behavioral links, they are referred to as market mechanism models (Parsons, 1981). Empirical response models utilize cross-sectional or time series data (Parsons & Schultz, 1976).

The market response methodology heavily relies on econometric and time series analysis (ETS). The models employ various estimation techniques, including cross-sectional, time series, non-linear, "mixed model" with inequality sign constraints and shrinkage-based estimators. The model estimates the sales contributions of coupons, relative distribution, relative shelf price, feature ads, displays, temporary price reductions, pantry-loading, and advertising (Hanssens et al., 2003).

While sales are often used as a performance measure, market share can sometimes provide a more accurate reflection of a firm or brand's performance. Models with market share as the dependent variable effectively account for competition but may encounter estimation and testing challenges in markets with multiple brands. Additionally, response models can be developed for other significant dependent variables beyond sales and market share. In marketing, special attention is given to measuring the effects of advertising by analyzing its contributions to short-term incremental sales as well as long-term sustenance of base volume sales and sales momentum (Hanssens et al., 2003).

As brands improve their ability to track prospect inquiries and consumer interactions, they can convert this data into valuable proprietary information. This data, generated internally, primarily consists of individual-level records of prospect or consumer responses to marketing activities and time series data of consumer transactions.

It is important to note that while price, promotion, and distribution data are often obtained as by-products of sales-performance data collection techniques, advertising data requires additional efforts. Companies are aware of their advertising costs but need to estimate competitors' advertising by utilizing media counting, which involves counting advertising, determining its claimed market value, and assessing its relationship with sales (Hanssens, Parsons & Schultz, 2003).

In summary, response and sensitivity modeling in brand management are based on the notion that historical data on consumer and market response to the marketing mix contain valuable information. Market response models are a foundation for refining marketing mix variables such as pricing, sales promotions, advertising copy, media selection and timing, and other brand-specific marketing elements. Among the various market response models, Media Mix Modeling becomes relevant when focusing on understanding the relationship between media efforts and business outcomes.

2.1.1 Marketing & Media Mix Models

Marketing Mix Modeling (MMM) is a multiple regression technique that assists executives in making effective trade-off decisions regarding budget allocation and resource distribution across various marketing elements (Raj et al., 2012). These models assume that the marketing mix variables are set independently of the response function parameters (Dubé, Hitsch & Manchanda, 2005). Traditionally, the marketing mix is represented by the four Ps: product, price, promotion, and place (distribution).

Product relates to characteristics such as the firm's product portfolio, uniqueness, differentiation from competitors, or qualitative superiority. Promotion includes advertising, detailing, informational sales promotions, displays, and features. Price refers to the suggested retail price or incentive sales campaigns such as discounts, temporary price reductions, or offers. Place refers to product delivery, determined by factors like distribution, availability, and shelf space.

It's important to note that marketing mix modeling is not a quick fix or a one-time effort due to its data requirements and analytical capabilities. Its true value lies in its consistent and persistent application. As an integral part of a retailer's management information system, it provides comprehensive decision support by offering global and granular insights into marketing activities. This data also enables predictions of future consumer behavior, facilitating the efficient organization of marketing elements.

Furthermore, many advertisers utilize multiple media channels to enhance exposure to prospective clients, and evaluating return on ad spend (ROAS) is a crucial concern. Media mix modeling (MMM) is an analytical technique, typically using multivariate regression, proposed by

Borden (1964) and McCarthy (1978), which employs observational data (e.g., pricing, media spending, sales, economic factors) to estimate and anticipate the impact of different media mix strategies on sales. The output distinguishes between exogenous forces (e.g., market expansion, demographic trends) and controllable factors like promotional strategies or advertising expenditure (Jin et al., 2017).

Randomized trials, commonly regarded as the gold standard for addressing causal questions, can be conducted at various levels, such as consumer, retailer, or regional levels (Vaver & Koehler, 2011). These models generate reliable causal outcomes, providing metrics like return on advertising expenditure and budget allocation strategies. However, the precision of experiments is limited by the ability to target manipulations and track outcomes. Despite their gold standard status, obstacles often prevent randomized experiments from being employed frequently to produce answers that MMM can provide. Conducting multiple tests for different scenarios may be necessary to answer similar questions as MMMs.

In conclusion, MMM relies on time series data, which can be evaluated using historical data or experimentation, and its primary objective often revolves around optimizing budget allocation. MMMs offer an effective means to understand the effectiveness of media through historical data. However, it is important to note that accurately evaluating the impact of marketing or media mix requires appropriate attribution, which involves assigning value to each channel appropriately.

2.1.2 Attribution Models

Marketing and media mix models are often regarded as methods for budget optimization. However, the attribution problem primarily focuses on achieving a precise and consistent interpretation of the impact of each consumer encounter on the final decision. Attribution entails utilizing advanced analytics to allocate appropriate credit to each marketing touchpoint, both online and offline, for the desired consumer action (Harmeling et al., 2017). Consequently, attribution modeling can be defined as a mathematical representation of the heuristics or algorithmic procedures employed to assign credit.

In the early stages of research, academics approached the attribution problem as a prediction issue (Li & Kannan, 2014). They utilized historical interaction data to predict attribution for a specific consumer action. Shao and Li (2011) propose a bagged logistic regression model to identify the best predictive fit model, followed by applying basic probabilistic models (such as second or higher-order models) to assign credit for the outcome. Similarly, Dalessandro et al. (2012) suggest a data-driven causal approach that compares the outcome with and without a specific touchpoint, assuming all other factors remain constant. They estimate the attribution of a particular touchpoint using the concept of Shapley value (Shapley,1953)

Since the publication of these two works, significant progress has been made in the field. Considering that attribution is the central focus of this study and our primary objective is to better understand the research field, we conducted a bibliometric analysis of 156 articles published in the last 50 years related to attribution modeling. Scopus and Web of Science are two widely used bibliometric databases globally. Both platforms proved equally effective and facilitated the completion of the prescribed activities without difficulty. Nevertheless, Scopus was selected as the data source for this study due to its excellent coverage and extensive result overlap with Web of Science, which was statistically sufficient to be considered as the standalone database.

The first search phrase contained the term “marketing attribution models,” for this is the primary context in which our research was conducted. Additionally, we used Boolean operators to narrow our search and include synonyms and related terms, for instance, “multi-channel attribution” and “multi-touch attribution,” as well as a regex operator to include any possible variation on the word model or modeling. Finally, we limited our research to final-stage English-language articles in fields such as business, social science, and econometrics. To summarize, Figure 1 presents the query string used to conduct this bibliometric search:

Figure 1

Query string for advanced search on Scopus

```
((((multi OR digital) AND media) OR channel OR multichannel OR multitouch OR advertising OR
marketing ) AND ( attribution OR “attribution model*"))

AND ( LIMIT-TO ( DOCTYPE ,  "ar" ) )  AND  ( LIMIT-TO ( SUBJAREA ,  "BUSI" )  OR  LIMIT-TO (
SUBJAREA ,  "SOCI" )  OR  LIMIT-TO ( SUBJAREA ,  "ECON" ) )  AND  ( LIMIT-TO ( LANGUAGE ,
"English" ) )  AND  ( LIMIT-TO ( PUBSTAGE ,  "final" ) )
```

The initial search yielded a substantial number of results (21,670). Although the search queries were valid, narrowing down the search criteria was challenging without excluding highly relevant articles. A qualitative examination was conducted to ensure the inclusion of articles directly related to marketing attribution and to meet our research objectives. We scrutinized the titles, abstracts, and keywords of all publications in the search results, selecting 156 articles.

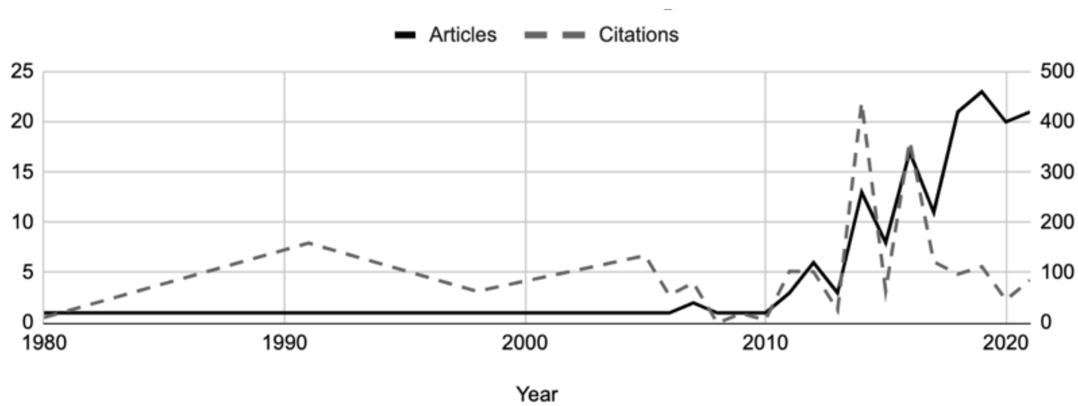
For bibliographic evaluation and the assessment of scientific stakeholders such as authors and sources, we employed Bibliometrix (Aria & Cuccurullo, 2017), an R-powered library. This tool facilitated the analysis of bibliometric data, allowing us to evaluate the performance of scientific agents and create the intellectual structure of the research area.

The data analysis process consisted of two stages. The first stage focused on scientific actors, including sources and authors relevant to the research question. Through bibliometric analysis of total citations and citation indexes, we assessed the performance of these scientific agents. The second stage involved the analysis of conceptual and intellectual networks, specifically co-citation analysis. By examining the co-citation networks of authors, we gained insights into the social organization and intellectual structure of the research field.

While conducting a year-by-year review, we found that the first publication on attribution modeling dates back to 1982. However, a longitudinal analysis reveals that attribution is a relatively new subject, with most publications and research conducted in the last decade. Prior to the surge in research on attribution modeling around 2010, many studies focused primarily on advertising effectiveness as a broader topic, with an emphasis on return on investment. This shift in focus reflects the significant transformation of the advertising industry, transitioning from a creativity-driven field to a data-driven activity that aims to demonstrate its relevance. This analysis also indicates that primary research procedures in attribution modeling closely resemble those employed in marketing mix modeling, often utilizing various forms of regression models.

Figure 2

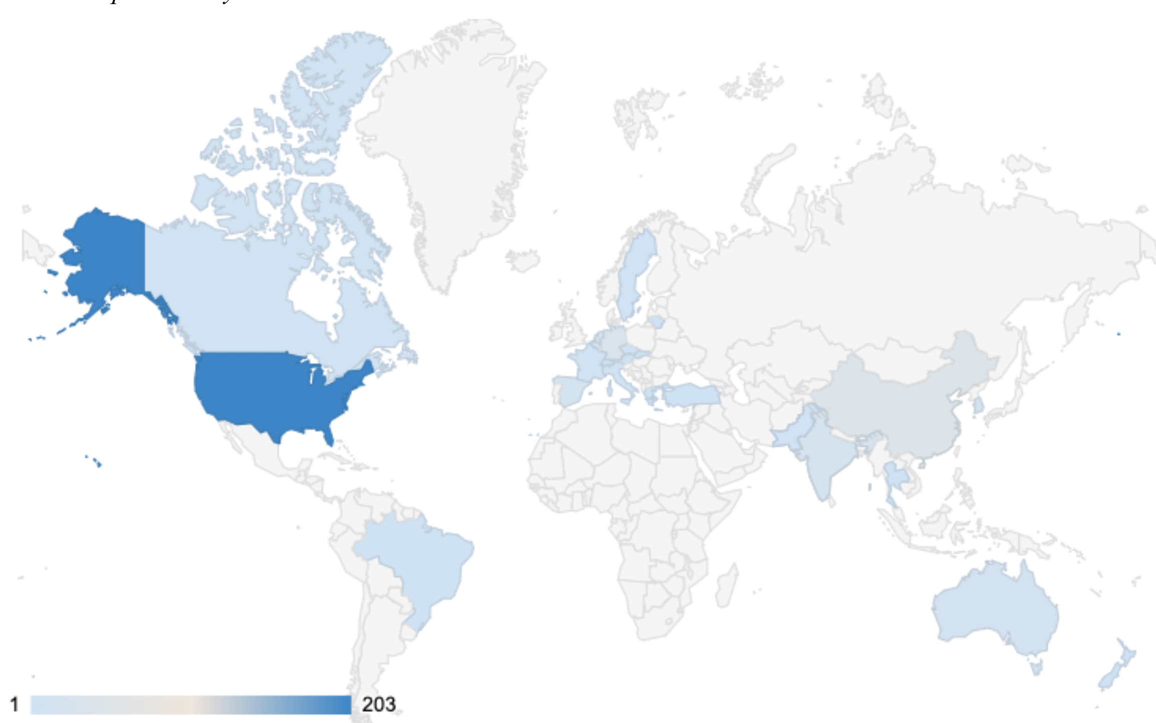
Number of publications and citations per year



Most studies occured in North America, with a few exceptions in India and China. This fact aligns with the broader Business & Management science field and the incredible amount of marketing investment done by companies in the U.S. (Research and Development Expenditure (% of GDP) | Data, 2019). **Figure 3** shows a detailed map of the publishing density in each country.

Figure 3

Publications per country



When analyzing the relevant sources in the field of attribution modeling, it was found that the most highly cited works have appeared in the Journal of Advertising, the International Journal of Research in Marketing, and the Journal of Marketing Research (see Table 1). The Journal of Advertising has long been regarded as a leading journal in the advertising field. In contrast, the Journal of Marketing Research, as an AMA journal, has primarily focused on econometric studies. It is worth noting that the Proceedings of the ACM SIGKDD also appeared in the ranking, indicating a trend towards including grey literature. This trend may stem from the field's need for rapid evolution, which requires a more agile form of publication, albeit with less emphasis on traditional scientific rigor.

Table 2 provides an overview of the ten most relevant sources based on the number of articles published. Once again, the Proceedings of the ACM SIGKDD appear in the top ranking, further supporting the observation that the field's rapid expansion necessitates a more agile publishing format. Furthermore, through qualitative examination of the published papers, it was observed that a significant portion of the work is focused on methodological advancements rather than producing theoretically robust publications. Additionally, a considerable number of empirical research studies and case studies have been published in the conference proceedings. The International Journal of Research in Marketing and the Journal of Marketing Research were identified as the most reputable sources based on the number of articles published.

Table 1

Most Relevant Sources by Number of Citations

Source	Articles	Citations	JCR	SJR	H-Index
Journal of Advertising	4	314	5.522	3.092	109
International Journal of Research in Marketing	10	279	4.513	3.725	102
Journal of Marketing Research	5	221	5.000	6.321	171
Proceedings of the ACM SIGKDD	10	189	-	1.004	182
Management Science	3	110	4.883	4.954	255
Journal of Current Issues and Research in Advertising	2	107	-	0.544	34
Quantitative Marketing and Economics	1	79	-	2.108	33
Journalism and Mass Communication Quarterly	1	63	-	2.020	80
Marketing Science	3	56	3.716	5.938	127
Journal of Retailing	1	49	5.245	3.184	136

Note. *The Journal Citation Reports – Clarivate Analytics (JCR) – https://jcr.clarivate.com presents the impact factor of a journal, based on the calculation of the number of citations in the current year divided by the original items published in that journal during the previous two years. The SCImago Journal & Country Rank (SJR and H-Index) - https://www.scimagojr.com/ presents the measure of scientific influence of academic journals calculated by the number of citations received by a journal and the importance or prestige of journals from where these quotes come from.*

Table 2

Most Relevant Sources by Number of Articles

Source	Articles	Citations	JCR	SJR	H-Index
Proceedings of the ACM SIGKDD	10	189	-	1.004	182
International Journal of Research in Marketing	10	279	4.513	3.725	102
Journal of Marketing Research	5	221	5.000	6.321	171
Journal of Advertising Research	4	17	3.154	0.831	86
Journal of Advertising	4	314	5.522	3.092	109
International Conference on Information and Knowledge Management, Proceedings	4	19	-	0.371	125
Marketing Science	3	56	3.716	5.938	127
Management Science	3	110	4.883	4.954	255
Journal of Marketing Analytics	3	2	-	0.305	11
Journal of Marketing	3	38	9.462	7.799	243

Based on our analysis, we were able to identify the key sources associated with media attribution research and publication. Bradford's Law states that documents on a specific topic are distributed according to a particular mathematical function, which means that increasing papers on a subject requires increasing the number of journals or information sources (Sangam, 2015). Consequently, if scientific journals are organized to minimize the proliferation of articles on a single issue, they can be categorized into a core group of periodicals that focus more narrowly on the topic. This implies that only a limited number of journals would be needed to provide the majority of publications on a specific specialized scientific area.

The data analysis indicates a notable overlap between the most cited and the most influential sources, validating the initial findings. This result supports the direction for future research and development to focus on the most prestigious journals in this particular subject area. Following Bradford's Law, the primary sources in media attribution research are identified as the Journal of Research in Marketing, the Proceedings of the ACM SIGKDD, and the Journal of Marketing Research. The remaining members of the core group are listed in Table 3.

Table 3

Core Sources according to Bradford's Law

Core Sources
International Journal Of Research In Marketing
Proceedings of the ACM SIGKDD
Journal Of Marketing Research
International Conference On Information And Knowledge Management, Proceedings
Journal Of Advertising
Journal Of Advertising Research
International Journal Of Information Management
Journal Of Digital And Social Media Marketing
Journal Of Marketing
Journal Of Marketing Analytics
Management Science

Table 4 lists the ten most influential authors regarding the number of articles published. With five papers published on attribution, Bajaras J. has established himself as the field's most active author. R. Akella and A. Flores also made the top ten, with four publications each. Although these writers were quite active in the field, their papers received an average of 3.8 citations per published article.

Table 4

Most Relevant Authors by Number of Published Articles

Author	**N. Pub.**	**Total Cit.**	**H-Index**	**G-Index**	**M-Index**	**Year Start**
Barajas J	5	17	2	4	0.182	2012
Akella R	4	16	2	4	0.182	2012
Flores A	4	16	2	4	0.182	2012
Holtan M	4	16	2	4	0.182	2012
Li Y	4	10	2	2	0.250	2015
Kannan Pk	3	255	3	3	0.333	2014
Lee S	3	8	2	2	0.222	2014
Lu Q	3	28	3	3	0.375	2015
Pauwels K	3	122	3	3	0.429	2016
Zhang Y	3	31	2	3	0.222	2014

Table 5 summarizes the ten authors who are frequently cited in the field. Kannan Pk. is the most referred author, indicating that he is the dominant figure in the subject. Even though his articles were published later than the rest on the list, his contributions to the field have been noteworthy. Additionally, the list includes highly referenced authors Dahlén M., Gotlieb Jb., and Sarel D. Unlike Kannan, these authors began their research early, even before the field reached its peak apogee in 2010, indicating that they may belong to the field's foundations.

Table 5

Most Relevant Authors by Number of Citations

Author	Total Cit.	N. Pub	H-Index	G-Index	M-Index	Year Start
Kannan Pk	255	3	3	3	0.333	2014
Dahlén M	186	2	2	2	0.111	2005
Li H	183	1	1	1	0.111	2014
Gotlieb Jb	159	1	1	1	0.031	1991
Sarel D	159	1	1	1	0.031	1991
Pauwels K	122	3	3	3	0.429	2016
Anderl E	111	2	2	2	0.286	2016
Schumann JH	111	2	2	2	0.286	2016
Xu L	99	2	2	2	0.222	2014
Dalessandro B	98	2	2	2	0.182	2012

According to the results of this bibliometric analysis, we can see in **Table 6** that the article "Attributing conversions in a multi-channel online marketing environment: An empirical model and a field experiment" by Li H., Kannan P.K. (2014), although relatively recent, can be considered a milestone in the development of the attribution field. It is worth noting that this article does not rest on a robust theoretical foundation; instead, it is based on empirical confirmation of past research.

The following two articles, published in the Journal of Advertising, provide an overview of the theoretical foundation underpinning all contemporary research. Both of these studies, by Gotlieb J.B., Sarel D. (1990) and Dahlén (2005), show the research field's fundamental nature, emphasizing advertising effectiveness rather than statistical methodology development.

Once more, it is worth noting that, despite being a proceeding, the ACM SIGKDD is represented here with two of the most cited publications. Given their recentness, both works primarily focused on statistical methodological improvement, emphasizing revenue optimization, which are features of the field's early works.

Table 6

Most Relevant Influential Articles

Year	Title	Source	Authors	Citations
2014	*Attributing conversions in a multi-channel online marketing environment: An empirical model and a field experiment*	Journal of Marketing Research	Li H., Kannan P.K.	183
1991	*Comparative advertising effectiveness: The role of involvement and source credibility*	Journal of Advertising	Gotlieb J.B., Sarel D.	159
2005	*The medium as a contextual cue: Effects of creative media choice*	Journal of Advertising	Dahlén M.	134
2014	*Path to purchase: A mutually exciting point process model for online advertising and conversion*	Management Science	Xu L., Duan J.A., Whinston A.	89
2012	*Bid optimizing and inventory scoring in targeted online advertising*	Proceedings of the ACM SIGKDD	Perlich C., Dalessandro B., Hook R., Stitelman O., Raeder T., Provost F.	86
2014	*Online ads and offline sales: Measuring the effect of retail advertising via a controlled experiment on Yahoo!*	Quantitative Marketing and Economics	Lewis R.A., Reiley D.H.	79
1998	*Does web advertising work? Memory for print vs. online media*	Journalism and Mass Communication Quarterly	Sundar S.S., Narayan S., Obregon R., Uppal C.	63
2016	*Mapping the consumer journey: Lessons learned from graph-based online attribution modeling*	International Journal of Research in Marketing	Anderl E., Becker I., von Wangenheim F., Schumann J.H.	62
2011	*Data-driven multi-touch attribution models*	Proceedings of the ACM SIGKDD	Shao X., Li L.	58
2016	*The effectiveness of different forms of online advertising for purchase conversion in a multiple-channel attribution framework*	International Journal of Research in Marketing	de Haan E., Wiesel T., Pauwels K.	57

In Figure 4, the intellectual structure of the research domain is depicted through co-citation analysis. Calculating the centrality index, the study domain is divided into three distinct clusters.

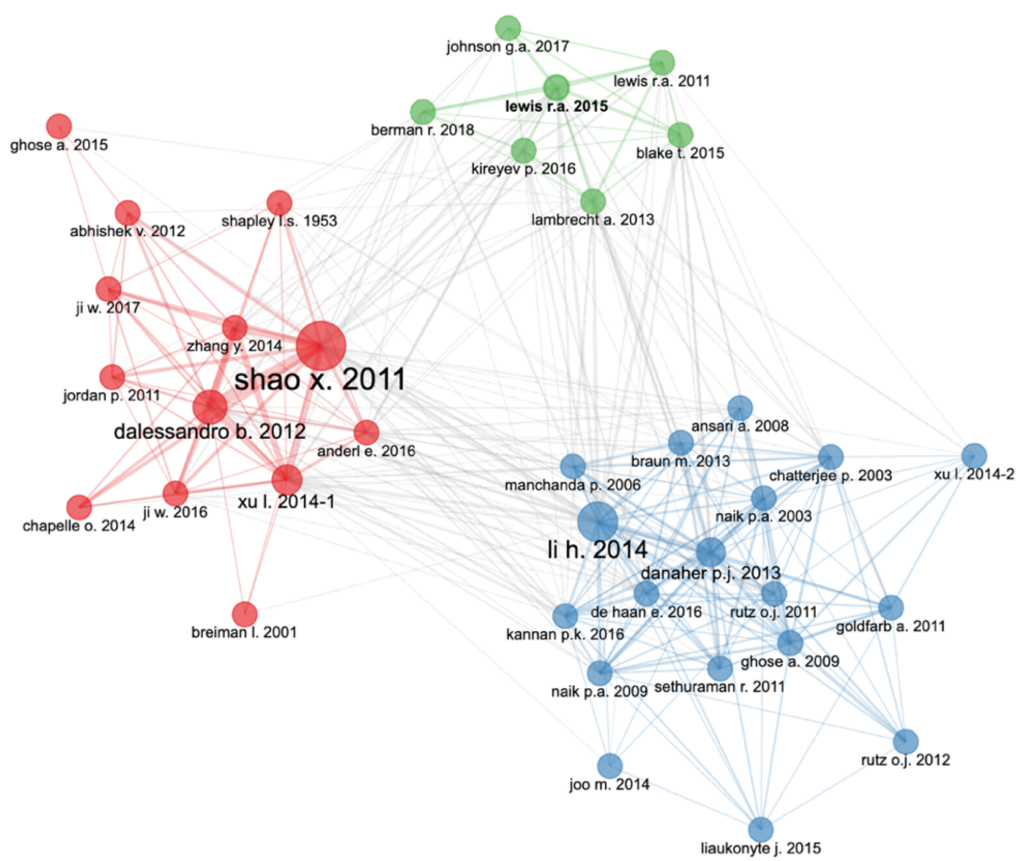
Cluster 1, represented in blue, consists of authors who primarily focus on analyzing the effectiveness of advertising in influencing consumer behavior. Several papers in this cluster also address the challenges of understanding the interactions between online and offline channels.

Cluster 2, highlighted in red, emphasizes the statistical and mathematical advancements in the field. Researchers in this cluster propose new approaches for assessing the impact of channels on desired consumer actions, often exploring the relationship between digital campaign metrics and financial data. This cluster frequently uses simulated data to validate its theoretical premises, relying on computational resources and algorithms.

The third and smallest cluster, depicted in green, is primarily concerned with empirical validation using established approaches such as Shapley value and Markov Chains. This group, which includes researchers from both academic and private affiliations, conducts controlled experiments to better understand the causal effects of budget optimization through more sophisticated attribution models.

Figure 4

Co-citation Analysis



By analyzing the arithmetic frequency of the author's keywords, we could see in **Table 7** the dataset's most frequent words. The analysis demonstrates that "online advertising" is the field's fundamental theme, while multi-channel attribution and multi-channel marketing are its most essential extensions.

The findings confirm that multi-channel attribution mainly focuses on the online environment, which may seem obvious from a practical aspect, but it helps us understand the studies' concentration area. Another intriguing conclusion is that, despite its seeming timidity, the privacy issue is beginning to emerge as a potential cause for concern. The whole set of words is listed below.

Table 7

Most Common Word

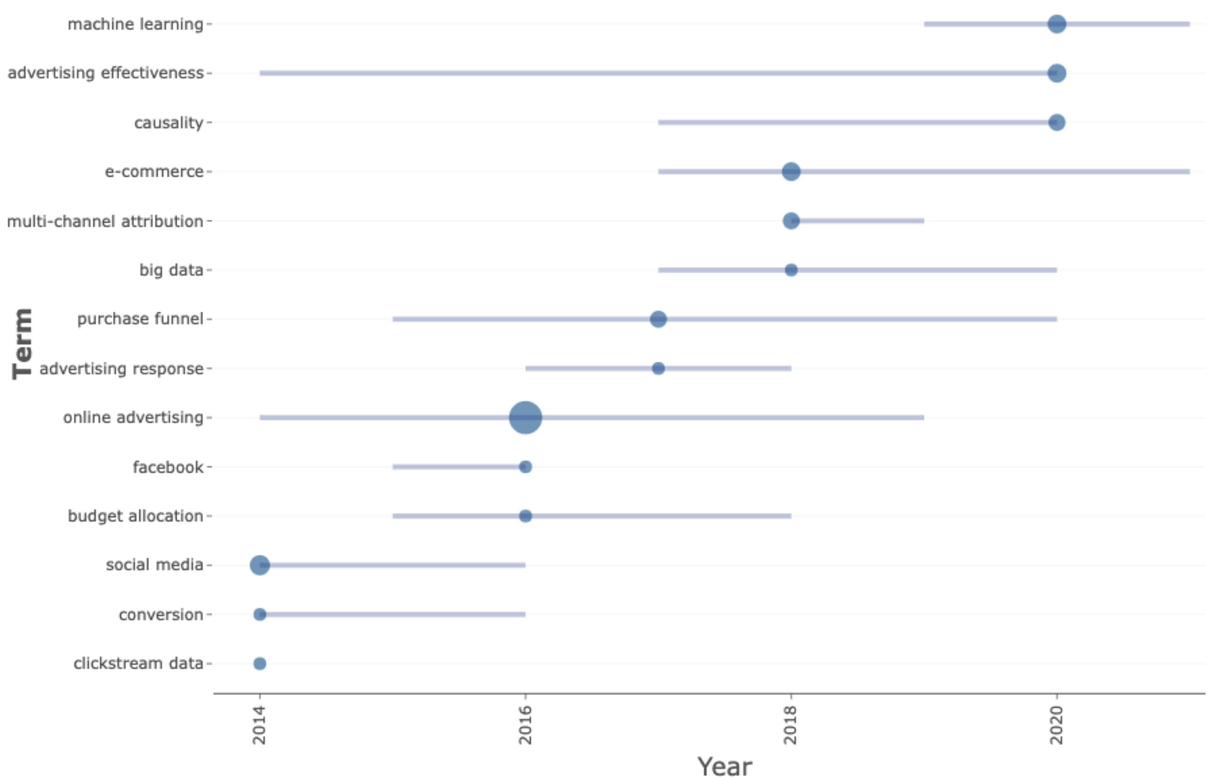
Words	Occurrences
online advertising	59
multi-channel attribution	40
multi-channel marketing	10
e-commerce	6
social media	6
advertising effectiveness	5
machine learning	5
causality	4
Privacy	4
purchase funnel	4

As depicted in **Figure 5**, *machine learning* has been a popular topic over the past several years. The findings align with previous observations regarding the evolution of techniques that utilize more powerful computing resources for credit allocation throughout the consumer journey. Advertising effectiveness has been a topic of interest across different eras, considering that one of the key objectives of attribution is to understand the impact of media efforts on a firm's marketing outcomes.

Initially, attribution models primarily focused on budget allocation, similar to approaches seen in media mix modeling (Shao & Li, 2011). However, the landscape has since progressed towards models based on path analysis (Li & Kannan, 2014; Xu et al., 2014). This shift is noteworthy because, as illustrated in Figure 5, budget allocation ceased to be the primary focus of the discipline after 2020.

Figure 5

Trending Topics



Co-word analysis creates keyword clusters. They are regarded as themes, the density and centrality of which can be utilized for categorizing themes and mapping in a two-dimensional graphic (Aria & Cuccurullo, 2017). “Thematic map is a highly intuitive diagram that allows us to examine themes based on the quadrant in which they are located: (1) upper-right quadrant: motor themes; (2) lower-right quadrant: basic themes; (3) lower-left quadrant: emerging or disappearing themes; (4) upper-left quadrant: very specialized/niche themes” (Aria & Cuccurullo, 2017)

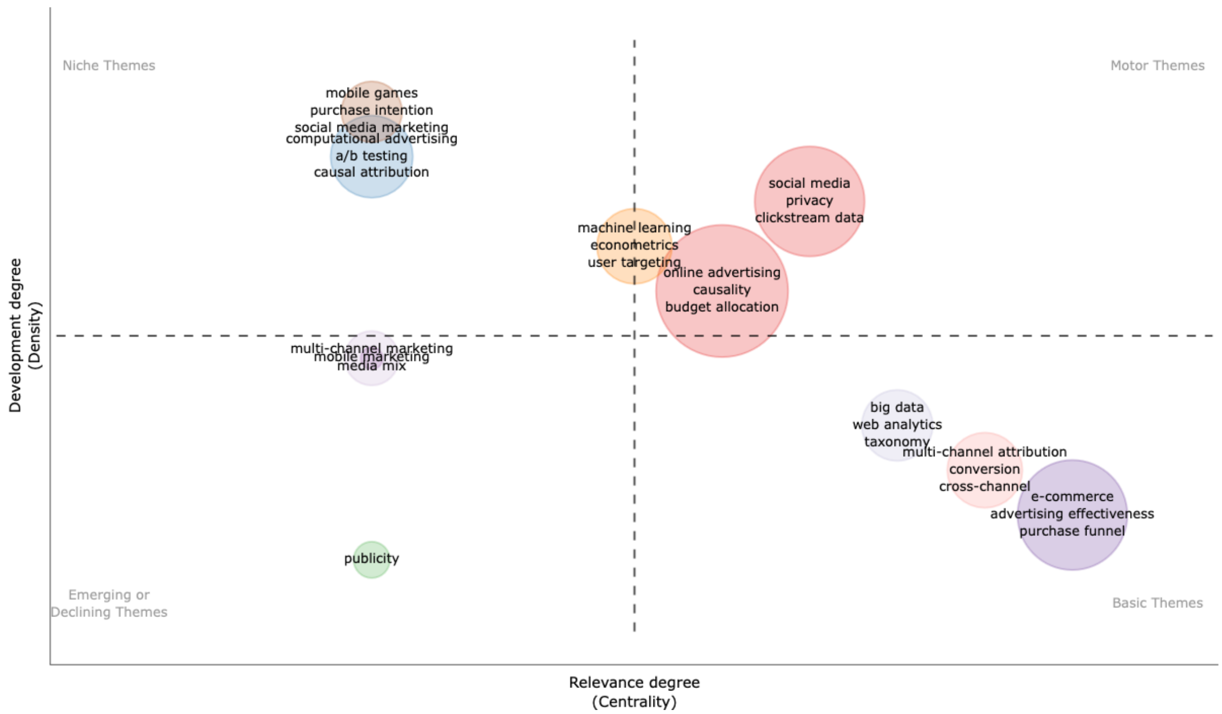
In the upper-right quadrant, we can find the subject field's driving themes, which consist primarily of papers on online advertising effectiveness with budget allocation and also a strong presence of social media and privacy-related concerns.

On the lower-right quadrant, we have our fundamental themes, which, according to our keyword selection for this bibliometric analysis, are multi-channel attribution, advertising effectiveness, and big data as essential topics.

In the lower-left quadrant, again concurring with the preceding states, we see media mix-related elements. Even though this is not our primary focus in this study, it is an important topic to examine, as the early attribution models significantly relied on the premises of media mix modeling.

Figure 6

Thematic Map



The research domain's intellectual structure was discovered via a co-occurrence network analysis utilizing authors' keywords. Using the calculation of the centrality index, we separated the study domain into 04 (four) clusters. As seen in **Figure 7**, each cluster was produced by inter-article solid linkages.

In cluster 1, which is blue and the largest, there are numerous online advertising-related words. This cluster appears to be the core topic in multi-channel attribution, which is consistent with our key findings during the refining results phase (2.3), where the most recent studies based on path analysis did not appear to include offline channels. The majority of the work that was able to incorporate offline channels relied on aggregate data and regression analysis (Shao & Li, 2011).

In addition, this cluster demonstrates the use of cost-related data and the budget allocation connection, which again appears to lead to a media mix modeling structure instead of a current multi-channel attribution modeling structure.

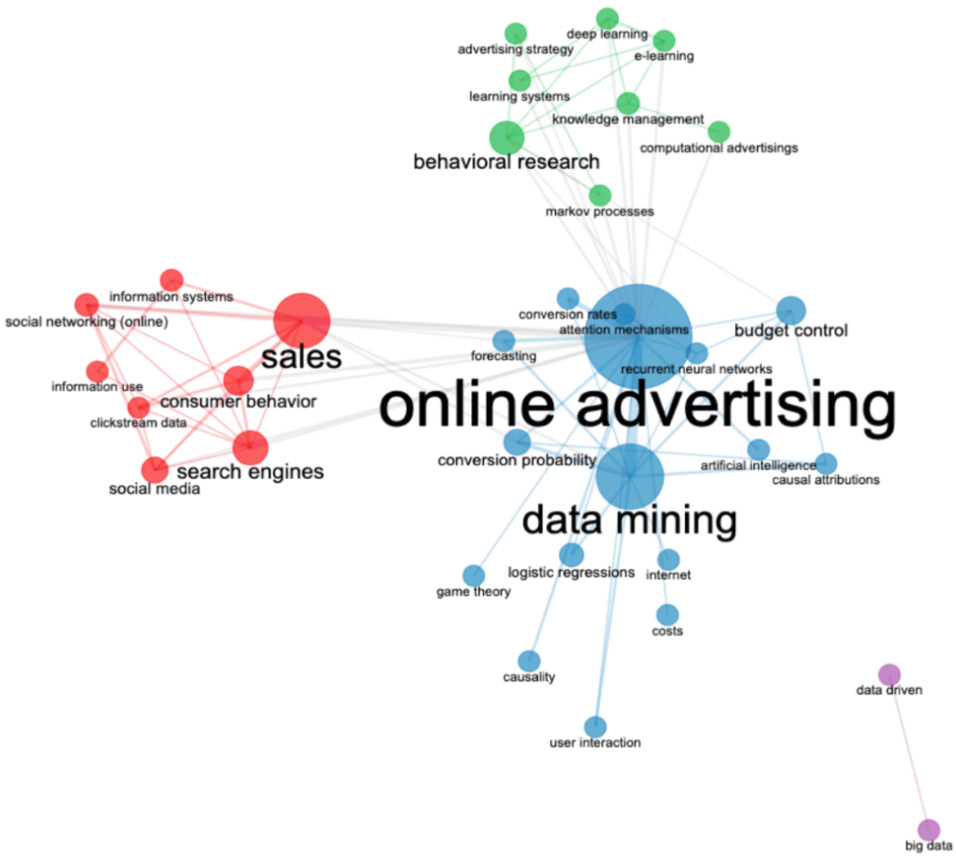
In cluster 2, highlighted in red, the word *sales* appears prominently alongside consumer behavior, indicating a relationship with the initial use of attribution as a prediction problem, in which researchers attempted to use a search engine and social media data to predict whether a particular behavior would result in a sale.

Cluster 3, shown in green, concentrates on words associated with behavioral studies and other "learning" systems. This cluster tended to group a more theoretical set of topics (advertising strategy and knowledge management) with more contemporary methodologies, such as deep learning and the Markov process. We feel this may be related to the word similarity rather than the subject similarity, and it may be necessary to do additional research to comprehend the relationship better.

The last and smallest is the purple cluster, which groups *data-driven* papers with *big data*. This may appear sensible because, although the term "data-driven" is not precisely new, its usage has increased since the rise of big data, making it more accessible and informative.

Figure 7

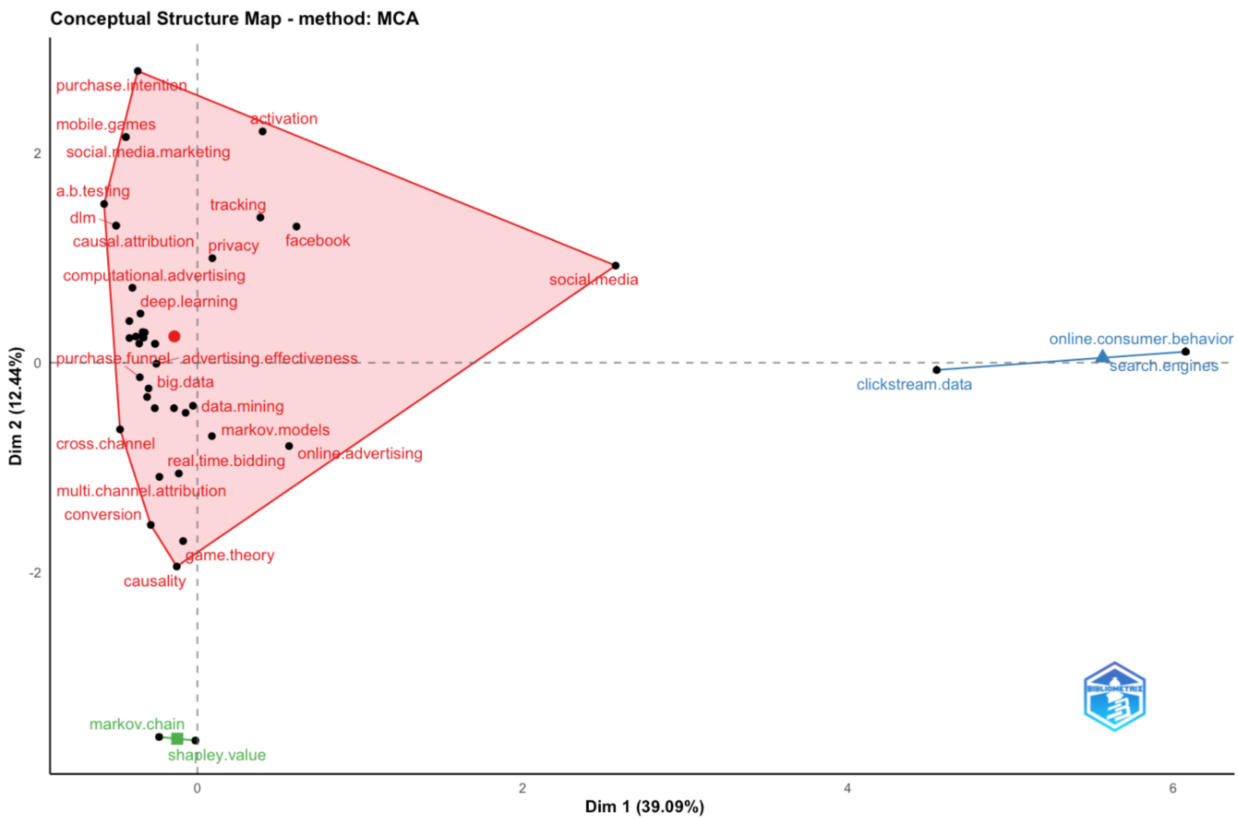
Keyword's Co-Occurrence Network



Multiple Correspondence Analysis (MCA) was performed on the keywords included in the dataset. **Figure 8** compresses voluminous data with several variables into a low-dimensional space to create an understandable two-dimensional graph that uses plane distance to indicate the similarity between the keywords. The closer a keyword is to the center, the more attention it has received in recent years. "*The results are interpreted based on the relative placements of the points and their distribution along the dimensions; the closer two words are shown on the map, the more similar their distribution*" (Aria & Cuccurullo, 2017).

Figure 8

Conceptual Structure Map



Cluster 1, in red, is the most representative, containing the majority of the terms and *advertising effectiveness*, as well as *purchase funnel* and *deep learning*, which appear to define our research subject best. The repeated appearance of online *consumer behavior* and *search engines* in the blue cluster indicates a significant relationship between these two topics. According to our earlier investigation, the green cluster containing solely *Markov Chains* and *Shapley Value* makes sense since it is the foundation for modern path-based multi-channel attribution models.

Currently, attribution is present in most digital analytics tools and software used by marketers, even in its basic form, such as heuristic rules. The most commonly used heuristic model is the last-click attribution model, which assigns full credit to the last marketing channel a consumer interacts with before taking the desired action, disregarding prior interactions (Google Inc., 2021). Despite its limitations in accurately representing the consumer journey, marketers and advertisers widely adopt this model as a market standard due to its interpretability and ease of implementation.

Multi-touch attribution is a pressing issue in digital advertising, particularly when multiple media channels are involved (Kannan & Li, 2017). Given that modern internet advertising campaigns are typically deployed across various platforms, consumers are exposed to the same advertising campaign multiple times before making a final decision (Ji, Wang & Zang, 2016).

Both theoretical and empirical evidence suggest that employing sophisticated marketing attribution methods helps maximize marketing return on investment (de Haan et al., 2016; Zhao, 2019). This has led to increased research demand and the development of new context-specific attribution methods (Ghose & Todri-Adamopoulos, 2016; Li & Kannan, 2014; Mukherjee & Jansen, 2017; Xu et al., 2014).

Academics have proposed various data-driven attribution systems, including logistic regression models (Shao & Li, 2011), game theory-based approaches (Berman, 2018; Dalessandro et al., 2012), Bayesian models (Jin et al., 2017), mutually exciting point process models (Xu et al., 2014), VAR models (Kireyev et al., 2016), and hidden Markov models (Abhishek et al., 2012). Additionally, internet-based industry firms like Google, Adobe, and VisualIQ have developed heuristic and data-driven attribution approaches based on these studies.

Further advancements in attribution modeling have been made possible through machine learning and neural network approaches. Ren et al. (2018) introduced the Dual-attention Recurrent Neural Network (DARNN) to address the multi-touch attribution problem, deriving attribution values directly from the conversion estimation objective. Shortly after, Li et al. (2018) presented the Deep Neural Net with Attention Multi-Touch Attribution Model (DNAMTA), which can be seen as a version of Ren's study. DNAMTA is a supervised learning technique that determines if a sequence of events leads to conversion, capturing the dynamic interaction effects between media channels. These studies serve as cornerstones for current attribution modeling and provide a roadmap for future development in the field.

However, in the context of multi-channel environments, the complete touchpoint path of consumers is often overlooked when evaluating channel effectiveness, despite the effectiveness and accessibility of data-driven attribution models to a wider audience. This is evidenced by the continued prevalence of comparatively simple heuristics, such as last-click attribution models (Li & Kannan, 2014).

Recognizing the importance of a comprehensive understanding of attribution, the academic community has started studying and synthesizing the literature (Gaur & Bharti, 2020). However, identifying patterns in contemporary literature can be challenging when a research topic is complex, in its early stages of investigation, and constantly evolving (di Stefano et al., 2010).

2.1.3 Attribution Summary

The accuracy of attribution models relies on their ability to assign value to touchpoints based on their actual influence on decision-making. Early academic research viewed attribution as a prediction problem. It proposed various data-driven attribution systems, including logistic regression models, game theory-based techniques, Bayesian models, mutually exciting point process models, VAR models, and hidden Markov models.

However, the complete consumer touchpoint journey is often overlooked when determining channel effectiveness in multi-channel scenarios, even though data-driven attribution methods have been proven more effective and accessible. Research has shifted towards multi-channel attribution models in recent years, considering the entire consumer journey.

With advancements in analytics technologies, it is now possible to track nearly every consumer interaction with digital marketing touchpoints, allowing researchers to create sequential interaction paths and develop statistical attribution models. It is important to note that while most models use consumer-level data, some models use aggregated data, achievable when technologies provide granular data at the consumer level.

Table 8 summarizes the most developed approaches, with probabilistic models being the most commonly used, particularly Markov Chains. Markov Chains are reliable and easy to create, but they have limitations when dealing with consumer pathways with more than four touchpoints, requiring significant computational resources.

Regression models are still employed, especially when aggregate time-series data is available, and income is a crucial variable. Linear models, despite their simplicity, can be effective in complex environments. Moreover, regression models can incorporate offline channels, which has been a challenge for consumer path-based methodologies in the past.

While machine learning-powered attribution models seem promising, it is essential to consider the intrusiveness and potential empowerment issues related to data collection methods. These advanced procedures often rely on path analysis, which involves identifying the consumer through personal data to replicate path sequences. Utilizing personally identifiable information (PII) without proper consent or knowledge of the consumer can be intrusive. It is crucial to seek more privacy-safe solutions for attribution models, even as we strive for accuracy and reliability in our approaches.

Table 8

Summary of Methodologies

Attribution Models	User Path Based	Heuristic Models	First Interaction
			Last Interaction
			Position Based
			Time-Decay
			Linear
		Probabilistic Models	Shapley
			Markov Chains
			LSTM
	Aggregation Based	Regression Models	Linear Regression
			Logistic Regression
			Vector Autogression

Note. Developed by the author (2023).

2.2 PRIVACY

2.2.1 GDPR

The complexity of determining whether a company's data processing activities are consistent with applicable laws, particularly in an international environment, is increasing as data protection standards rise. By their very nature, data may transcend borders effortlessly and play a crucial role in the global digital economy. According to Porter & Kramer (2018), data have become a valuable asset in recent years and have even been referred to as the future currency.

The processing of personal data occurs in numerous fields of economic and social activity, and the development of information technology dramatically simplifies the processing and sharing of such data. In this context, the European Union (EU) approved the General Data Protection Regulation (GDPR) to harmonize further the standards for data protection across the EU Member States and to increase the level of privacy for those affected. The GDPR went into effect on May 25, 2018. Due to its expansive, international reach of application, it will also involve a large number of non-EU businesses. Entities should assess whether they fall under the GDPR's scope of application and work expeditiously to comply with its requirements.

Compliance with the GDPR may necessitate a time- and cost-intensive evaluation of an organization's current data protection practices. Consequently, businesses may need to modify their data processing architecture and procedures. This Regulation applies to the processing of personal data entirely or partially by automated means and to the non-automated processing of personal data that form part of or are intended to form part of a filing system. To summarize its material breadth, the GDPR applies to personal data processing.

The Regulation will become applicable to businesses as soon as data processing occurs. In order to assure a high level of protection, the (material) scope is read expansively. In other words, any data level processing that may occur having an EU citizen's data or on EU soil will be under the GDPR norm. This also is true for multi-location data storage servers.

2.2.2 Privacy-by-default vs Privacy-by-design

The GDPR prioritizes preventative data protection principles. Entities should address the ideas of Privacy by Design and Privacy by Default, as the need to design and apply them is directly enforceable.

Privacy by Design (Art. 25 Sec. 1 GDPR) is predicated on the understanding that the conditions for data processing are essentially determined by the software and hardware employed for the task. The minimally intrusive processing of data must be based on technological concepts for preventive protection. Developers and manufacturers must keep data minimization in mind when building new technologies. Examples include IT solutions geared toward data minimization and thorough and timely pseudonymization of personal data.

The principle of Privacy by Default (Article 25 Sec. 2 GDPR) would protect consumers against the prevalent trend among firms to collect as much personal data as possible. By default, only the personal data required for the intended purpose of the data processing will be collected. The notion concerns the quantity of collected personal data, the scope of their processing, the length of their storage, and their accessibility. The controller must take adequate technical and organizational measures for this objective. When a controller employs a processor, the processor must enable the controller to achieve Privacy by Default.

Consumers should not need to adjust the default settings of a service or product upon first usage or access to safeguard their privacy when privacy-friendly default settings are in place. If consumers wish to change these settings, for example, to allow continued use of their personal data or to share it with different parties, they should be required to opt in and modify the settings themselves. Privacy by Default will primarily protect persons who lack the technical expertise or time to apply privacy-friendly settings themselves.

Moreover, as the complexity and variety of online services and data use increase, it is difficult to assess the influence of technological settings on data protection. The primary use case for Privacy by Default should be privacy-friendly technological default settings when getting consent from data subjects.

Because the work we are creating is entirely focused on protecting consumer data privacy and because we apply an approach that reduces the amount of data collected, we may refer to our strategy as “privacy-by-design.”

2.2.3 Data Ethics & Online Trust

Each year, the generation and transmission of data via the internet continue to grow exponentially, and Cisco anticipated that annual data traffic would reach 4.8 Zettabytes by 2020.

While data collection enables numerous improvements in consumer service, the massive amount of data collected poses significant privacy risks, forcing businesses to prioritize data security and online trust.

In the digital environment, trust is a virtue developed by artificial agents and responsibility in the online environment. The term "online trust" refers to the trust that precisely defines the relationship of communication between individuals in a digital environment. There are two types of trust in the online world: (1) General trust, in which a person places their trust in the majority of websites, and (2) Specific trust, in which a person places their trust in a single website;

Data has always been a critical component of scientific research, including measurement, analysis, recording, representation, and validation, and this would be no different in marketing science. Initially, data generation increased in lockstep with computer generation. Still, the rapid growth of computer technologies has resulted in an exponential increase in the amount of data generated by consumers and collected by businesses (Harsh et al., 2018).

Currently, consent mechanisms for data collection in digital environments are either non-existent or non-compliant with the recently implemented GDPR. Respectively, data subjects face significant privacy concerns, as they lack proper control over their data (Joyee De & Imine, 2019). Therefore, consent compliance is a critical step in protecting consumer privacy.

Data ethics is a new branch of ethics that examines how data is generated, recorded, and shared, as well as how data-driven algorithms such as machine learning, artificial intelligence, and robotics are used and how they follow a set of guidelines to produce more desirable output, such as informational control (Harsh et al., 2018).

While this study is not intended to provide guidelines or a framework for ensuring information control or contentment compliance, it can address the data-driven attribution problem without jeopardizing consumer privacy.

2.2.4 Privacy Summary

The processing of personal data happens in a wide variety of spheres of economic and social activity. The advancement of information technology has considerably simplified the processing of such data as well as the sharing of it.

While data collection enables numerous improvements in consumer service, the massive amount of data collected poses significant privacy risks, forcing businesses to prioritize data security and online trust.

Currently, consent mechanisms for collecting data in digital settings are either non-existent or not compliant with the General Data Protection Regulation (GDPR), which was recently implemented. On the other hand, data subjects are vulnerable to considerable privacy risks since they do not have adequate control over their data.

Because our work is focused on protecting consumer data privacy and applying an approach that reduces the amount of data collected, we may refer to our strategy as "Privacy-by-design."

3. PRIVACY-BY-DESIGN ATTRIBUTION MODEL (PBDAM)

The privacy-by-design strategy aims to address the privacy concerns associated with modern attribution modeling. Instead of relying on path analysis or consumer-level data, the proposed approach focuses on aggregated daily data such as clicks, views, costs and conversions for each channel. This helps protect the privacy of consumer data while still allowing for a comprehensive understanding of business outcomes and desired consumer actions.

The attribution problem differs from media mix modeling in that it aims to provide a precise and consistent interpretation of the impact of each consumer interaction on the final decision rather than just optimizing the budget allocation. To assess each touchpoint's impact accurately, the model is structured into three phases, which are detailed in the following sections. These phases likely involve steps such as data preprocessing, model development, validation, and interpretation.

By following a privacy-by-design approach and leveraging aggregated data, the proposed model seeks to strike a balance between maintaining consumer privacy and providing valuable insights into attribution. Developing methodologies that respect privacy while delivering meaningful and actionable results for marketers and advertisers is essential.

3.1 METHOD AND MODEL SPECIFICATION

Modern media mix model influence the transition from a path-analysis-based approach to a time-series-based analysis in the proposed model. The following three phases are implemented to assess the impact of each channel on marketing performance:

1. Generalized Forecast Error Variance Decomposition (FEVD) Matrix: Inspired by Diebold and Yilmaz (2009), the model utilizes a Generalized FEVD matrix. This matrix allows for a comprehensive understanding of the effect of the entire sample, creating a spillover table. Each channel metric's predicted influence, known as the spillover index, on all other channel metrics is determined through this matrix. It quantifies how changes in one channel affect other channels in the marketing mix.
2. Relative Weight Analysis: In this phase, the model employs relative weight analysis to determine the relative importance of each channel in driving the final marketing performance in the form of. This analysis helps assign appropriate weights to each channel based on its contribution to the desired consumer action.

3. Privacy-by-Design Attribution Model (PBDAM): Finally, the model assigns the final credit to each channel using the spillover index and relative weight as a vector. These vectors are placed orthogonally in a Cartesian plane, and the resultant is calculated to determine the credit allocation for each channel. The PBDAM framework ensures that privacy concerns are addressed while providing insights into the impact of each channel on marketing performance.

By combining these three phases, the proposed model offers a privacy-conscious approach to attribution modeling, leveraging time-series analysis and drawing on the principles of media mix modeling. This allows for a comprehensive understanding of channel effectiveness and credit allocation in a privacy-safe manner.

3.1.1 Spillover Effect

Spillover Effect happens when an agent's actions or behaviors indirectly affect other agents' outcomes through peer effects, social interactions, or externalities. In other words, an interdependence is the influence between two separate variables that initially seem unrelated. Our premise is that, in a multi-channel marketing strategy, each interaction made by a consumer in one of the marketing channels may have a positive or negative influence on future interactions, as demonstrated in several pieces of research focusing on synergies in a multi-channel environment (Abhishek et al., 2012; de Haan et al., 2016; Flavián, Gurrea & Orús, 2020; Ghose & Todri-Adamopoulos, 2016; Kireyev et al., 2016; Mukherjee & Jansen, 2017).

Attribution models based on consumer-level data usually contemplate this premise using path analysis such as Markov Chains or Game Theory using Shapley's Value. However, our approach will be based on the Spillover Index proposed by Diebold and Yilmaz (2009) which the spillover index will be calculated for each channel metric (clicks, views, costs, and conversions).

3.1.1.1 Spillover Index

The Spillover Index is a measurement instrument for accurately assessing the spillover index from an econometric standpoint. In the tradition of Engle et al. (1990), this method gives a spillover measure based on vector autoregressive (VAR) models. The index is based on Forecast

Error Variance Decompositions (FEVD), which quantify the proportion of the movement in a variable's development over time attributable to its shocks and that attributable to shocks in other variables in the vector autoregression by determining how much of the total variance forecast is attributable to each. The spillover index is a measure of interdependence among variables, with a higher index value indicating that a more significant proportion of market shocks can be attributed to cross-variable shocks as opposed to own-variable shocks. In contrast to their earlier work, Diebold and Yilmaz (2012) employ the generalized VAR framework of Koop et al. (1996)and Pesaran et al. (2001).

The benefit of this kind of VAR is that variance decompositions are independent of the variable ordering. In addition, Diebold & Yilmaz (2012) proposed the Generalized Forecast Error Variance Decomposition (GFEVD) method, which allows us to divide each variable's forecast error variances into portions related to the various system shocks. In contrast to the initial model, which measured the total spillovers in a simple VAR framework (i.e., with potentially order-dependent results driven by Cholesky factor orthogonalization), GFEVD estimates the directional spillovers in a generalized VAR framework that eliminates the potential dependence of results on ordering.

3.1.1.2 Generalized Forecast Error Variance Decomposition

Let us start by considering a covariance stationary N-variable VAR(p), where $\varepsilon \sim (0, \Sigma)$ is a vector of independently and identically distributed disturbances.

$$1. \quad X_t = \sum_{i=1}^{p} \Phi_i X_{t-i} + \varepsilon_t$$

The Moving average representation is

$$2. \quad X_t = \sum_{i=0}^{\infty} A_i \varepsilon_{t-i}$$

Diebold and Yilmaz (2009) mentioned that "*instead of attempting to orthogonalize shocks, the generalized approach allows correlated shocks but accounts for them appropriately using the historically observed distribution of the errors. As the shocks to each variable are not* *orthogonalized, the sum of the contributions to the variance of the forecast error (the row sum of the elements of the variance decomposition table) is not necessarily equal to one*."

Now we define the variance shares as the fractions of the H-step-ahead error variances in forecasting X_i that is due to shocks to X_i, for i = 1, 2, . . . , N, and cross variance shares, **or spillovers**, as the fractions of the H-step-ahead error variances in forecasting xi that are due to shocks to X_i, for i, j = 1, 2, . . . , N, such that i is different of j.

Denoting the Generalized VAR framework H-step-ahead forecast error variance decompositions by Θ_{ij}^{g}(H), for H = 1,2…, we have:

$$3. \quad \Theta_{ij}^{g}(H) = \frac{\sigma_{jj}^{-1} \sum_{h=0}^{H-1} \left(e'_i A_h \Sigma e_j\right)^2}{\sum_{h=0}^{H-1} \left(e'_i A_h \Sigma A'_h e_i\right)}$$

where Σ is the variance matrix for the error vector ε, σ_{jj} is the standard deviation of the error term for the j^{th} equation, and e_i is the selection vector, with one as the i^{th} element and zeros otherwise. Before continuing, we applied the normalization as suggested by DY using:

$$4. \quad \tilde{\Theta}_{ij}^{g}(H) = \frac{\Theta_{ij}^{g}(H)}{\sum_{j=1}^{N} \Theta_{ij}^{g}(H)}$$

After calculating the contributions from the GVAR variation decomposition, it is now possible to construct the total spillover index as follows:

$$5. \quad S^{g}(H) = \frac{\sum_{\substack{i,j=1 \\ j \neq i}}^{N} \tilde{\Theta}_{ij}^{g}(H)}{\sum_{i,j=1}^{N} \tilde{\Theta}_{ij}^{g}(H)} .100 = \frac{\sum_{\substack{i,j=1 \\ j \neq i}}^{N} \tilde{\Theta}_{ij}^{g}(H)}{N} .100$$

The generalized VAR approach enables us to identify the direction of spillovers across major asset classes. However, although analyzing the total spillover index is sufficient for determining the extent of shocks, we chose to compute the directional spillovers using the normalized components of the GFEVD matrix since they are independent of the variable ordering.

Hence, channel metric *i* receives spillovers from all other channels metrics *j* following:

$$6. \quad S_{i.}^{g}(H) = \frac{\sum_{\substack{j=1 \\ j\neq i}}^{N} \widetilde{\Theta}_{ij}^{g}(H)}{\sum_{i,j=1}^{N} \widetilde{\Theta}_{ij}^{g}(H)}.100 = \frac{\sum_{\substack{j=1 \\ j\neq i}}^{N} \widetilde{\Theta}_{ij}^{g}(H)}{N}.100$$

Similarly, we measure the directional volatility spillovers transmitted by channel i to all other channel j as:

$$7. \quad S_{.i}^{g}(H) = \frac{\sum_{\substack{j=1 \\ j\neq i}}^{N} \widetilde{\Theta}_{ji}^{g}(H)}{\sum_{i,j=1}^{N} \widetilde{\Theta}_{ji}^{g}(H)}.100 = \frac{\sum_{\substack{j=1 \\ j\neq i}}^{N} \widetilde{\Theta}_{ji}^{g}(H)}{N}.100$$

Inputs to the PBDAM are derived from the net volatility spillover shown below, which summarizes how much each variable (channel metric) contributes to overall channel variance.

$$8. \quad S_{i}^{g}(H) = S_{.i}^{g}(H) - S_{i.}^{g}(H)$$

The individual channel spillover index will be simplified to (γ) to prevent further notational clutter. Following the rationale for constructing a real channel influence, we now use a simple arithmetic method to add the spillover index (γ) from all channel-related metrics (k).

$$9. \quad \gamma_k = \gamma_{cost_k} + \gamma_{click_k} + \gamma_{view_k} + \gamma_{adstock_k}$$

Since we have already calculated the influence of channels across all channels, we will compute the influence of each channel on the final sales volume in the following step.

3.1.2 Relative Weight Analysis (RWA)

When dealing with highly correlated data in analyzing online channel data such as cost, views, and clicks, it is not uncommon to encounter collinearity issues. Collinearity is when two or more predictor variables in a regression model are highly correlated. This can lead to unstable or misleading results when assessing the relative importance of variables.

In the presence of collinearity, traditional indices such as correlation coefficients, regression weights, and partial and semi-partial correlations may not accurately assess the relative significance of variables. These indices assume independence among predictors and can be influenced by their collinearity.

Alternative approaches can be employed to address the collinearity issue and accurately evaluate the relative importance of variables. One such approach is using relative weights, as introduced by Johnson in 2000. Relative weights aim to assign credit to marketing channels based on their contribution to the predicted criterion variance, considering the predictors individually and in combination with other variables.

While relative weights have demonstrated accuracy, reliability, and computational ease, it is important to note that the approach still relies on consumer-identifiable data and primarily focuses on income rather than the indirect influence resulting from the interaction of multiple channels. Privacy concerns associated with consumer-identifiable data should be carefully addressed when implementing attribution models.

Exploring techniques such as feature selection methods, dimensionality reduction approaches (e.g., principal component analysis), or regularization methods (e.g., ridge regression or lasso regression) is advisable to mitigate collinearity issues. These techniques can help identify the most influential variables and alleviate the impact of collinearity estimating relative importance.

Overall, when analyzing highly correlated online channel data, considering the presence of collinearity and employing appropriate methods to evaluate the relative importance of variables is crucial to ensure accurate and reliable attribution modeling results. Nevertheless, RWA is a statistical technique used to determine the relative importance of highly correlated predictors in a regression equation (LeBreton). To resolve the correlated predictors' problem, the RWA generates a new set of uncorrelated predictors (Z_{X_k}) with the highest degree of similarity to the initial correlated predictors (X_j) stage. For a more detailed formulation of the matrix orthogonalization, consult the work of Johnson, 2000.

Then, the dependent variable is used to regress the transformed predictors, yielding a series of standardized regression coefficients (β_k). The original regression coefficients (λ_{j_k}) are now used to rescale these new coefficients to the original variables (Johnson, 2001). Two equations can represent the relationship between predictors and the outcome variables.

The first one represents the relationship between the predictor X_j, the original coefficients (λ_{j_k}), and the new set of uncorrelated predictors (Z_k), where ω represents the disturbance:

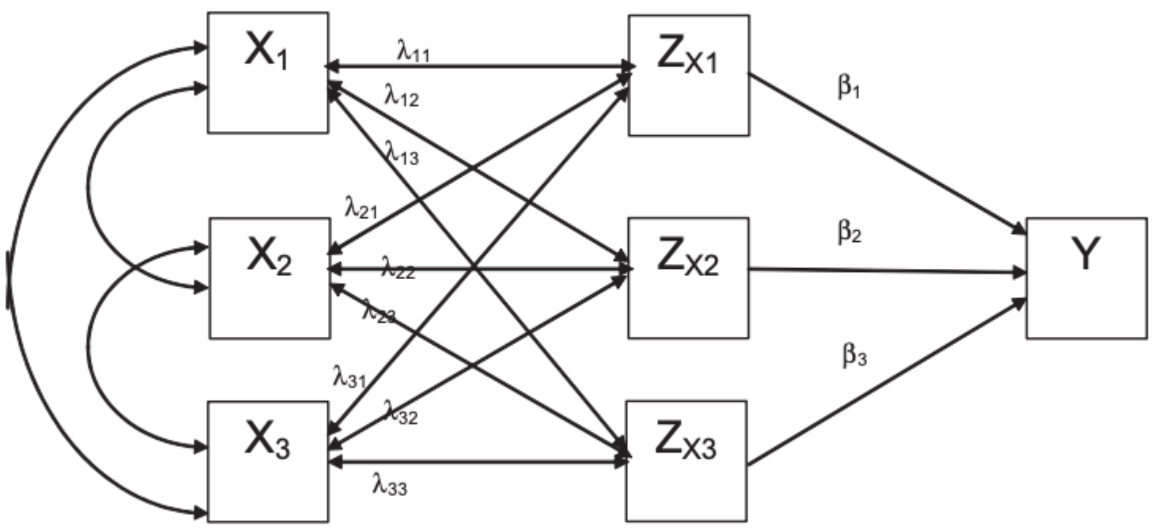
$$10. \quad X_1 = \lambda_{11}Z_1 + \lambda_{12}Z_2 + \lambda_{12}Z_3 + \omega$$

The second one describes the new set of uncorrelated predictors (Z_k) and the regression outcomes, where v represents the disturbance:

$$11. \quad Y_1 = \beta_{11}Y_1 + \beta_{12}Y_2 + \beta_{12}Y_3 + v$$

Figure 9

Graphical representation of univariate relative weights for three predictors



Graphical representation of univariate relative weights for three predictors. From "History and Use of Relative Importance Indices in Organizational Research" by J. W. Johnson and J. M. LeBreton, 2004, Organizational Research Methods, 7, p. 250. Copyright 2004 by Sage. Adapted with permission.

We can use the equation below to calculate the first variable's relative weight. The squared elements represent the relative contribution between uncorrelated and the original correlated predictors.

$$12. \quad \varepsilon_1 = \lambda_{11}^2 \beta_1^2 + \lambda_{12}^2 \beta_2^2 + \lambda_{13}^2 \beta_3^2$$

The sum of individual relative weights can comprise the squared multiple correlations.

$$13. \quad R^2 = \sum \varepsilon_j.$$

To calculate the relative weight (ε) of a channel (k) as a whole, you can aggregate the relative weights of each metric associated with that channel. The relative weight of each metric represents its contribution to the channel's overall performance. By summing the relative weights of all the metrics related to a particular channel, you can determine the relative weight of that channel. So:

$$14. \quad \varepsilon_k = \varepsilon_{cost_k} + \varepsilon_{click_k} + \varepsilon_{view_k} + \varepsilon_{adstock_k}$$

Where the sum of the relative weight (ε) of all the channels (k_i), added by the relative error (ω) is summed into unity.

$$15. \quad \sum_{i=1}^{n} \varepsilon_{k_i} \, \omega_i = 1$$

This is important since what we are trying to understand the sales credit earned by each channel, so the final sales volume is expected to stay the same in this formulation.

Because of its simplicity, relative weights are appealing. Since the Z_{X_k} is uncorrelated, the variance in the criterion allocated to each predictor may be assigned with certainty. The relative weights are also intuitively meaningful because they add to the entire model R^2. Simply put, the relative weight approach allows one to calculate the proportionate contribution of a single $X_j$to the prediction of Y by multiplying the proportion of variation in each Z_{X_k} accounted for by $X_j$by the proportion of variance in Y accounted for by $Z_{X_k}$and adding these products (Johnson, 2000).

3.1.3 Combining Spillover Index and Relative Weight

In order to combine the spillover index and each channel's relative weight, we choose a simple yet elegant approach. First, let us assume the Spillover Index (γ) and Relative Weight (ε) of each channel (k) as being both vectors ($\vec{\gamma}, \vec{\varepsilon}$). We will represent them **orthogonally in a cartesian plan,** placing both conjointly in the intersection (0,0). Now we calculate the PBDAM ($\vec{\psi}$) as the resultant of the vectors ($\vec{\gamma}, \vec{\varepsilon}$). So, by using the principles of analytical geometry and the Pythagorean Theorem, we describe the PBDAM as:

$$16. \quad \vec{\psi}_k = \sqrt{\vec{\gamma}_k^{\,2} + \vec{\varepsilon}_k^{\,2}}$$

Where the PBDAM ($\vec{\psi}_k$) is calculated as the square root of the sum of the squared spillover index (γ) and the squared relative weight (ε). So far, both spillover index and relative weight sum to unity, which logically makes sense since both represent the relative influence of each variable. However, using this approach, the final result for a mix of advertising channels does not sum to unity. To cover this issue, we simply normalize each component ψ_k as a fraction of the sum of all ψ. So,

$$17. \quad \psi_{k_i}(normalized) = \frac{\psi_{k_i}}{\sum_{i=1}^{n} \psi_{k_i}}$$

The sum of PBDAM of all channels, which can also be described as each channel's credit for the sales results, should be equal to 1.

$$18. \quad \sum_{i=1}^{n} \psi_{k_i}(final) = 1$$

3.1.4 Privacy-by-design Attribution Model - Summary

The Privacy-by-design Attribution Model (PBDAM) incorporates methodologies from various fields, including marketing science, finance, and natural sciences. The model is based on the assumption that each marketing channel conveys a message to consumers, and over time, this message accumulates a stock of consumer goodwill. However, this goodwill decays as time passes since the prior exposure to the message.

The PBDAM also considers the presence of synergy between different channels and their related metrics. Although these metrics may exhibit strong correlations, they are all deemed crucial in influencing the ultimate sales outcomes.

The process of the PBDAM can be summarized in the following steps:

1. Calculate the Spillover Index (γ) using GFEVD for all variables presented dataset:

$$S_{i.}^{g}(H) = \frac{\sum_{\substack{j=1 \\ j\neq i}}^{N} \widetilde{\Theta}_{ij}^{g}(H)}{\sum_{i,j=1}^{N} \widetilde{\Theta}_{ij}^{g}(H)} . 100 = \frac{\sum_{\substack{j=1 \\ j\neq i}}^{N} \widetilde{\Theta}_{ij}^{g}(H)}{N} . 100$$

2. For each channel (k) add the spillover index related to its metrics:

$$\gamma_k = \gamma_{cost_k} + \gamma_{click_k} + \gamma_{view_k} + \gamma_{adstock_k}$$

3. Apply the relative weight calculations to the same dataset using:

$$\varepsilon_1 = \lambda_{11}^2 \beta_1^2 + \lambda_{12}^2 \beta_2^2 + \lambda_{13}^2 \beta_3^2 \ldots$$

4. For each channel (k), add the relative weight related to its metrics:

$$\varepsilon_k = \varepsilon_{cost_k} + \varepsilon_{click_k} + \varepsilon_{view_k} \ldots.$$

5. Place the Spillover Index and the Relative Weight as orthogonal vectors in a cartesian plan and calculate the resultant using the Pythagorean Theorem:

$$\vec{\psi}_k = \sqrt{\vec{\gamma}_k^{\ 2} + \vec{\varepsilon}_k^{\ 2}}$$

6. Rescale the resultant vector to sum into unity.

$$\psi_{k_i}(normalized) = \frac{\psi_{k_i}}{\sum_{i=1}^{n} \psi_{k_i}}$$

3.2 MODEL EVALUATION

This section implements the Privacy-by-design Attribution Model (PBDAM) by utilizing both synthetic and real empirical data sets. This study aims to demonstrate the benefits and efficacy of the PBDAM compared to other commonly employed multi-channel attribution models.

In the initial instance, the PBDAM will be implemented on a synthetic dataset that includes aggregated daily data and subsequently compared to alternative multi-channel attribution models using consumer-level data. The purpose is to illustrate that the PBDAM can effectively capture hidden interaction information between channels and generate fair attribution values. The PBDAM demonstrates its capacity to allocate credit fairly by utilizing the combined variables and considering the fundamental interactions among channels.

The PBDAM methodology will also be implemented using real campaign data obtained from two prominent companies operating in distinct industries within Brazil. The aim is to showcase the efficacy in practical situations. Consequently, our work aims to elucidate the privacy-oriented characteristics and the appropriate utilization of aggregated non-personally identifiable information (non-PII) data for attribution modeling.

3.2.1 Data Description

The first dataset utilized in this study is a synthetic dataset derived from Google's internal data, ensuring a dependable and comprehensive source for analysis. The initial empirical dataset was acquired from the marketing department of a highly regarded private university in Brazil, referred to as Company A. The second donation was graciously provided by a local business proprietor in São Paulo, hereafter referred to as Company B. The three datasets comprise two distinct components that provide comprehensive insights into the performance of marketing channels and customer interactions.

The first part of the dataset comprises a daily data set spanning from October 1st, 2021 to December 31st, 2021. It aggregates each channel's metrics, including essential indicators such as clicks, views, conversions, and cost. These metrics serve as key performance indicators, offering a granular understanding of channel-specific activities and outcomes. The second part of the dataset captures customer-level data in the same time frame, which encompasses the media touchpoint path leading up to conversions. This data reveals the sequence and combination of interactions that consumers had with various channels before completing a conversion. This information is crucial for accurately attributing credit to each channel based on its influence in driving conversions. However, it is essential to highlight that for an interaction to be considered valid and recorded in the path, it must satisfy the criteria to be considered equivalent to a click. This premise will hold for all future studies.

By combining the daily dataset and customer-level data, the synthetic dataset provides a robust foundation for evaluating and comparing the Privacy-by-design Attribution Model (PBDAM) performance with other multi-channel attribution models. Additionally, it allows for a holistic analysis of the underlying interactions between channels and their impact on conversion outcomes.

3.2.2 Model Comparison and Results

The initial step involved implementing the most common heuristic models on the consumer-level data. The industry standard models that were utilized in our study include last-touch, first-touch, position-based, and time-decay.

While the first two models may seem straightforward, it is crucial to outline specific starting parameters for the latter two. For the position-based attribution model, as a starting point, we assigned a credit value of 0.3 to the initial position, 0.4 to the final position before conversion, and 0.3 to each intermediate position, divided by the total number of channels involved. Similarly, as an initial parameter for the time-decay attribution model, we adopted a 7-day half-life, a parameter frequently utilized as a benchmark in the majority of analytical tools (Google Inc, 2021). This choice was made to mirror real-world scenarios as closely as possible by employing commonly used parameters.

While widely used in the digital advertising industry due to their simplicity and ease of understanding, the rules underlying these heuristic models often overlook the specificity of the customer journey or disregard certain touchpoints, making them myopic or short-sighted. To account for this, we applied the Markov-Chain Model to the consumer-level data, using the removal effect as the metric for the attribution model (Kakalejčík et al., 2018).

Finally, the Privacy-by-Design Model (PBDAM) was implemented on the aggregated dataset, which does not contain personally identifiable information (PII). It is crucial to establish this distinction as the PBDAM, as indicated by its name, was not designed to apply to consumer-level data. Consequently, it can only be employed on aggregated data. Similarly, the heuristic models were not intended for use with aggregated data. In order to effectively compare both types of models, it is necessary to obtain consumer-level data and aggregate data for the same company within the same time frame. Hence, this approach was employed in our study.

3.2.2.1 Results – Synthetic Data

The synthetic dataset was provided by Google's internal analytical tool and included a sample size of over 69,484 fictional consumers and 412 distinct paths. The dataset under consideration consists of five distinct channels, which will be referred to as Alfa, Beta, Gama, Delta, and Epsilon, only for analytical convenience. Each channel may contain one or more metrics, depending on their characteristics.

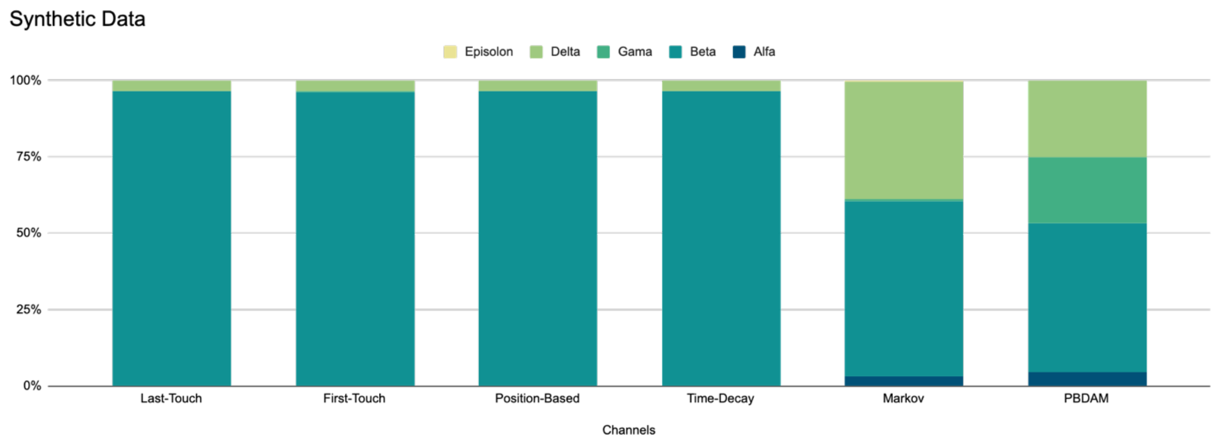
Table 9

Attribution Models Comparative – Synthetic Data

Synthetic Data	**Consumer Level Data - Consumer Journey (Path)**					**Aggregated Data**
Model Type	***Heuristics***				***Algorithmic***	***Algorithmic***
Channels	Last-Touch	First-Touch	Position-Based	Time-Decay	Markov	**PBDAM**
Alfa	0.0006	0.0005	0.0006	0.0006	0.0311	**0.0448**
Beta	0.9641	0.9623	0.9632	0.9637	0.5732	**0.4865**
Gama	0.0008	0.0010	0.0009	0.0008	0.0073	**0.2167**
Delta	0.0345	0.0362	0.0353	0.0350	0.3856	**0.2520**
Epsilon	0.0000	0.0000	0.0000	0.0000	0.0029	**0.0000**

Figure 10

Attribution Models Comparative – Synthetic Data



After conducting a comparative analysis, it is evident that PBDAM demonstrates more significant heterogeneity in credit allocation among the tested models. The insignificance of the Gama channel can be observed in all models except for PBDAM. Another notable observation pertains to the credit allocation of Delta, which exhibits a more pronounced effect, specifically in the PBDAM and Markov-Chain Model.

Beta significantly influences multiple attribution models, including last-touch, first-touch, position-based, and time-decay. Its attribution value remains consistently high at approximately 0.96 across all four models. This strongly indicates that Beta plays a significant role in the consumer's ultimate decision-making process.

In the Markov and PBDAM the Beta channel holds a considerable portion while sharing a notable contribution with the Delta and Gamma channels. In comparison with heuristics models, Beta's value decreases to 0.5732 in the Markov model context, whereas Delta experiences a surge to 0.3856, and Gama demonstrates an increase to 0.0073. The PBDAM model displays a more intricate portrayal of the landscape, whereas Beta and Delta demonstrate values of 0.4865 and 0.2520 correspondingly, while Gama experiences a notable increase to 0.2167.

The Alfa and Epsilon channels exhibit minimal impact across all models. On the other hand, Alfa demonstrates a notable increase in attribution value within the Markov model, registering a value of 0.0311. This increase is even more pronounced in the PBDAM model, where Alfa achieves a value of 0.0448. Although Alfa has limited impact in the last-touch, first-touch, position-based, and time-decay models, it seems to have a more significant role when examining the transitional dynamics between channels.

In brief, under less intricate, single-touchpoint frameworks, Beta emerges as the prevailing channel. However, a more thorough analysis reveals that the significance of Delta and Gama increases when considering more complex multi-channel interactions. The channel Alfa, which has been observed to have limited influence in simpler models, exhibits noteworthy contributions in the context of Markov and PBDAM models. This suggests that the potential impact of Alfa may have been previously underestimated or overlooked.

Therefore, the PBDAM and Markov models are similar in their ability to reveal the varied distribution of attribution across channels. This underscores the importance of the PBDAM model in providing a more comprehensive understanding of channel attributions and transitional dynamics that might be overlooked in simpler heuristic models.

3.2.2.2 Results – Company A

The dataset that is being considered has nine channels. For analytical simplicity, we will refer to these channels as Alpha, Beta, Gamma, Delta, Epsilon, Zeta, Eta, Theta, and Iota. Each of these channels can contain one or more metrics, depending on the qualities that are unique to that channel.

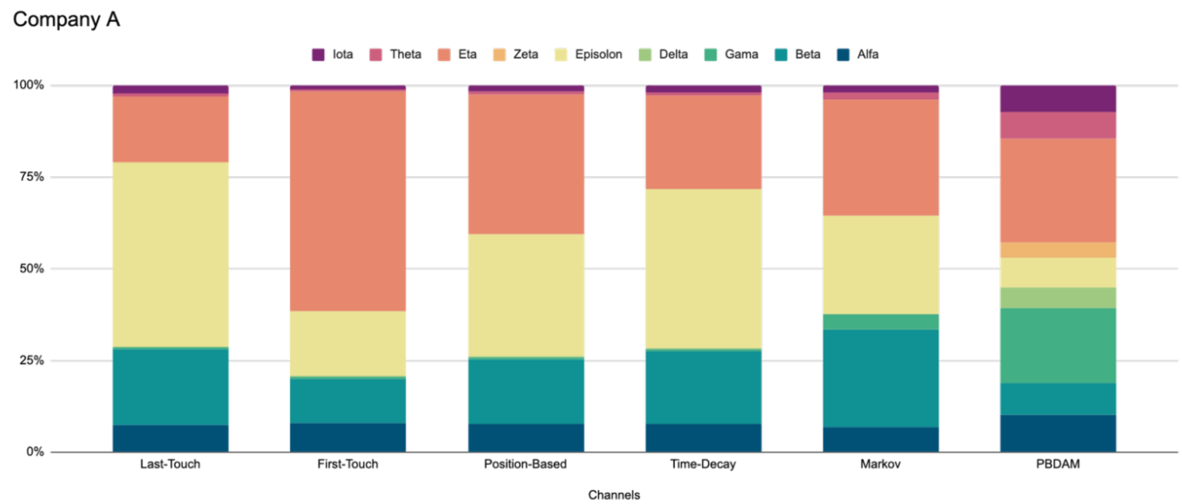
Table 10

Attribution Models Comparative – Empirical Data Set – Company A

Company A	**Consumer Level Data - Consumer Journey (Path)**					**Aggregated Data**
Model Type	***Heuristics***				***Algorithmic***	***Algorithmic***
Channels	Last-Touch	First-Touch	Position-Based	Time-Decay	Markov	**PBDAM**
Alfa	0.0740	0.0803	0.0771	0.0754	0.0694	**0.1006**
Beta	0.2073	0.1189	0.1748	0.1999	0.2652	**0.0893**
Gama	0.0048	0.0070	0.0064	0.0056	0.0404	**0.2028**
Delta	0.0016	0.0029	0.0022	0.0018	0.0010	**0.0561**
Epsilon	0.5020	0.1763	0.3332	0.4356	0.2674	**0.0818**
Zeta	0.0004	0.0004	0.0004	0.0004	0.0001	**0.0423**
Eta	0.1781	0.5979	0.3817	0.2525	0.3183	**0.2826**
Theta	0.0099	0.0048	0.0074	0.0089	0.0194	**0.0719**
Iota	0.0220	0.0115	0.0170	0.0200	0.0188	**0.0727**

Figure 11

Attribution Models Comparative – Empirical Data Set – Company A



Initially, Epsilon is the most influential channel in three models: last-touch, position-based, and time-decay. Its values are 0.5020, 0.3332, and 0.4356, respectively. This indicates that Epsilon has a substantial impact on the customer's decision-making process, whether that influence is at the final interaction (Last-Touch), distributed evenly across the customer journey (Position-Based), or favoring more recent interactions (Time-Decay).

However, the First-Touch model shows a different scenario. Eta has the highest value (0.5979), suggesting it's the most influential channel when considering the first interaction with the customers. This highlights the role of Eta in initiating the customer's journey.

According to the Markov model, the Eta channel exhibits the highest level of influence, as indicated by a value of 0.3183. Subsequently, the channels Epsilon and Beta demonstrate relatively lower levels of influence, with values of 0.2674 and 0.2652, respectively. The PBDAM model enhances the dispersion of influence, wherein Eta assumes the highest position (0.2826), closely trailed by Gama (0.2028) and Alfa (0.1006).

Specific channels demonstrate a marked rise in their influence in the Markov and PBDAM models compared to the single-touchpoint models. Gama's values increase to 0.0404 and 0.2028 in the Markov and PBDAM models, respectively. Similarly, despite having minimal influence in the single-touchpoint models, Zeta shows a significant rise to 0.0423 in the PBDAM model.

In summary, Epsilon and Eta emerge as dominant channels under the single-touchpoint models, with Epsilon significantly influencing on the latter part of the customer's journey and Eta having a pronounced impact at the beginning. However, when considering more complex interactions between channels, as shown by the Markov and PBDAM models, a more nuanced picture emerges, with Beta, Alfa, Gama, and even Zeta gaining significant attribution.

Once again, the PBDAM and Markov models share similarities in their ability to highlight the significant roles of seemingly unimportant channels in multi-touchpoint interactions, which are often overlooked in single-touchpoint models. Hence, the PBDAM model is essential for understanding the dispersion of influence across different channels in the customer's decision-making process.

3.2.2.3 Results – Company B

The dataset for the company B has eight channels. For analytical simplicity, we will refer to these channels as Alpha, Beta, Gamma, Delta, Epsilon, Zeta, Eta, and Theta. However, these may not correspond to the same channels as above, we chose the same nomenclature only to facilitate the analysis. Each of these channels can contain one or more metrics, depending on the qualities that are unique to that channel.

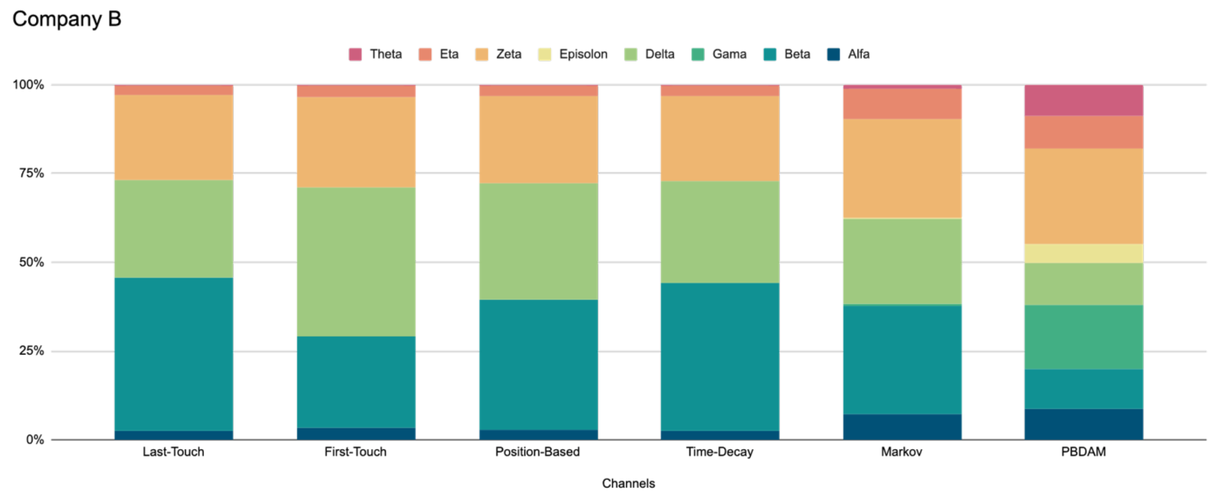
Table 11

Attribution Model Comparative – Empirical Data Set – Company B

Company B	**Consumer Level Data - Consumer Journey (Path)**					**Aggregated Data**
Model Type	***Heuristics***				***Algorithmic***	***Algorithmic***
Channels	Last-Touch	First-Touch	Position-Based	Time-Decay	Markov	**PBDAM**
Alfa	0.0252	0.0332	0.0283	0.0259	0.0727	**0.0821**
Beta	0.4317	0.2575	0.3678	0.4157	0.3040	**0.2234**
Gama	0.0000	0.0001	0.0001	0.0000	0.0062	**0.1623**
Delta	0.2742	0.4190	0.3277	0.2875	0.2401	**0.1212**
Epsilon	0.0002	0.0002	0.0002	0.0002	0.0035	**0.0358**
Zeta	0.2388	0.2555	0.2440	0.2402	0.2760	**0.2494**
Eta	0.0282	0.0326	0.0301	0.0287	0.0877	**0.0952**
Theta	0.0017	0.0018	0.0018	0.0017	0.0098	**0.0307**

Figure 12

Attribution Model Comparative – Empirical Data Set – Company B



It is worth noting that the last-touch, position-based, and time-decay models reveal Beta as the predominant channel, with values of 0.4317, 0.3678, and 0.4157, respectively. This suggests Beta's instrumental role in the latter stages of the customer interaction. In contrast, the First-Touch model positions Delta as a prime initiator of customer engagement, evidenced by a leading value of 0.4190. However, when analyzing Markov, despite Beta maintaining a significant presence (0.3040), Zeta demonstrates a robust influence, with an attribution value of 0.2760.

Furthermore, Alfa and Eta, despite their moderate presence in the single-touchpoint models, register a surge in their influence within the Markov and PBDAM models. Alfa exhibits an increase to values of 0.0727 and 0.0821, while Eta exhibits an elevation to 0.0877 and 0.0952 in the Markov and PBDAM models, respectively. This ascension in values insinuates their latent yet significant role in multi-touchpoint interactions.

However, the continuous low scoring of Epsilon and Theta across all models suggests a limited role in influencing customer conversion. Epsilon's values hover around 0.0002 in the single-touchpoint models, marginally improving to 0.0358 in the PBDAM model, reinforcing its minor role.

In conclusion, the data presents a dynamic and complex landscape of channel influence. While Beta, Delta, and Zeta emerge as dominantly influential in single-touchpoint models, the multi-touchpoint models foreground the importance of seemingly peripheral channels like Gama, Alfa, and Eta.

Such findings undervalue the indispensability of adopting a holistic and nuanced analytical approach in accurately interpreting the impact of different channels. Therefore, the PBDAM model is indispensable in accurately interpreting the impact of different channels.

4. CONCLUSION

The Privacy-by-Design Attribution Model (PBDAM) demonstrates an appreciation for the intricate interactions within multi-touch attribution frameworks. This model distinguishes itself by elucidating the diverse attribution patterns and recognizes the interplay and transition probabilities between varying marketing channels. Consistent with Li & Kannan (2014), the model seems to take into account the spillover and carryover effects across channels, being able to capture subtle effects throughout the entirety of consumer journey. Unlike heuristic models such as the Last-Touch or First-Touch approach, which are limited by their linear perspective, PBDAM incorporates a more inclusive analysis by crediting a broader range of channels.

Moreover, PBDAM exhibits an exceptional capability to expose the latent potential of previously underestimated channels. For instance, the Gama channel, which exhibited limited influence in other models, experienced a substantial increase in attribution under PBDAM in both synthetic and empirical data. This observation aligns with the studies of Abhishek et al. (2012), demonstrating that PBDAM effectively recognizes channels contributing significantly to the customer journey, notwithstanding their position as neither the initial nor final point of interaction.

Conforming to the findings of Kireyev et al. (2016), PBDAM also underscores that customer journeys are rarely linear or simplistic. The model elevates the significance of channels that may be overlooked in attribution models predicated on more straightforward rules, such as the first or last interaction. For example, when analyzing the results of company B, channels like Alfa and Eta, relatively insignificant in single-touchpoint models, registered increased attribution under the PBDAM lens.

In comparison with other models, PBDAM presents a unique perspective. For instance, when analyzing the Beta channel from synthetic data, predominant in last-touch, first-touch, position-based, and time-decay models, experiences a dilution of influence within PBDAM. This observation indicates a more equitable distribution of attributions within PBDAM, affirming its lack of bias towards the first or last interaction and a more comprehensive appraisal of the entire customer journey.

Interestingly, the attribution distribution of PBDAM closely mirrors that of the Markov model, a complex multi-touch model, as opposed to simpler models. This similarity, congruent across synthetic and empirical data, substantiates PBDAM's efficacy in capturing the subtleties of the customer journey. This consistency between PBDAM and Markov models underscores the robustness and reliability of the PBDAM model as an analytical tool, akin to the well-established Markov model. It implies that PBDAM can capture the nuances and intricacies of customer-channel interactions with as much precision and depth as the Markov model.

This is especially crucial for capturing transitional dynamics that might be overlooked in simpler models, thereby providing a more comprehensive understanding of the multi-faceted nature of channel interactions. Such a consistent performance across experiments not only validates the PBDAM model's efficacy but also emphasizes its utility in accurately interpreting the impact of different channels on consumer decision-making processes.

In conclusion, the PBDAM, through its comprehensive analysis of channels' effectiveness, emerges as a robust instrument in the attribution modeling landscape. It embraces the complexity of customer journeys, venturing beyond linear interpretations and providing a more comprehensive understanding of the varied paths leading to conversions.

4.1 WORK CONTRIBUTION

The Privacy-by-Design Attribution Model (PBDAM) emerges as an innovative contribution to multi-channel attribution, with numerous advantages and substantial contributions to academic understanding and managerial practice.

The core of PBDAM's advantages is the sophisticated quantification of the mutual influence and interactions between different marketing channels (Diebold & Yilmaz, 2012). This embodies a marked progression from traditional linear attribution models, offering a more nuanced understanding of interdependencies in multi-channel environments.

A prevalent issue in multi-channel attribution analysis is the high collinearity amongst various online channels, such as cost, views, and clicks. Incorporating of Relative Weight Analysis (RWA) in the PBDAM signifies an innovative approach to tackling this predicament. Through the

RWA, the model can accurately discern the relative importance of individual channels, both standalone and in synergy with other channels. This addresses the collinearity issues that frequently confound interpretation in multi-channel attribution studies.

The PBDAM extends beyond the confines of singular channel analysis and illuminates the synergies between various channels and their associated metrics. This allows the model to capture the complex dynamics often overlooked by traditional models, offering more accurate reflections of marketing strategies and their implications.

A key advantage of PBDAM resides in its emphasis on accessibility and explicability. The model facilitates understanding and utilization across diverse users, regardless of their statistical prowess. Integrating the spillover index and relative weight as orthogonal vectors in a Cartesian plane presents an intuitive visualization of the complex interactions between channels, enhancing interpretability.

The scientific contributions of the PBDAM are manifold. Its novel integration of the spillover index and Relative Weight Analysis into a comprehensive multi-channel attribution model fosters a more thorough understanding of the complex dynamics in multi-channel marketing. This contribution advances academic discourse in marketing analytics and equips practitioners with a robust tool for more effective resource allocation, potentially enhancing marketing ROI.

Furthermore, the PBDAM represents a significant stride forward in the multi-channel attribution modeling, with promising implications academia and business practice. The Privacy-by-Design Attribution Model (PBDAM) introduces an expanded perspective in attribution modeling by showcasing the intricate and multifaceted nature of customer journeys. This challenges the traditional thinking that customer journeys are simple, linear processes and encourages further exploration and research into complex, non-linear customer journey models. It pushes the boundaries of current attribution modeling techniques, encouraging the academic community to reassess and update traditional models.

PBDAM also bridges marketing, data science, and privacy studies. Its unique approach, which involves applying sophisticated data analysis techniques to understand multi-touch attributions while considering privacy issues, can contribute to academic insights in these disciplines and facilitate their integration. This integration could lead to breakthroughs in understanding customer behavior, optimizing marketing strategies, and preserving data privacy simultaneously.

As digital privacy becomes a growing concern, PBDAM's privacy-by-design feature aligns with the increasing academic focus on privacy-preserving data analysis techniques. This approach can lead to a better understanding of how privacy and data analysis coexist, paving the way for future studies.

Lastly, PBDAM’s detailed attribution modeling explains how various marketing channels contribute to the customer journey. Managers can leverage this insight to allocate resources more efficiently, thereby maximizing marketing ROI. Managers can optimize their marketing spend and design more impactful marketing strategies by understanding which channels influence the customer's journey most effectively.

4.2 LIMITATIONS AND FUTURE RESEARCH

Despite the impressive results demonstrated by the Privacy-by-Design Attribution Model (PBDAM) in uncovering the complexity of customer journeys, this research acknowledges some limitations. A notable deficiency arises from the inability of PBDAM to integrate Adstock - a measure of the carry-over or lagged effect of marketing efforts (Broadbent, 1984).

Firstly, the lack of Adstock may result in inaccurate attribution of conversions, with the model assigning undue credit to recent marketing activities while overlooking the latent effect of prior exposures (Naik & Raman, 2003). Misattribution can lead to misguided strategic decisions regarding resource allocation across marketing channels.

Secondly, the omission of Adstock may distort the perceived efficacy of different marketing channels. Specifically, channels yielding immediate results might be overvalued, whereas the more latent channels could be undervalued. This could lead to an inefficient allocation of marketing investments, favoring channels with apparent short-term efficiency over those that deliver value over extended periods (Rutz & Bucklin, 2011).

Lastly, excluding Adstock limits PBDAM's ability to provide a thorough understanding of the temporal dynamics inherent in customer journeys. This could restrict the depth of insight into consumer behavior, thereby hindering the development of marketing strategies that align with actual consumer response patterns over time (Shuba, Marc & Koen., 2010).

To address these limitations, future iterations of PBDAM should aim to incorporate the Adstock factor. This addition will enhance the model's ability to account for the lingering effects of marketing activities, thus facilitating a more comprehensive and accurate representation of customer journeys. Further research could also focus on developing techniques to determine optimal Adstock rates for different marketing channels and exploring how these rates might vary across various industries or market contexts.

An additional limitation of PBDAM concerns the degrees of freedom within the model. The degrees of freedom in a statistical model refer to the number of values that can vary while maintaining a fixed value of some statistic. In regression models, degrees of freedom often have an inverse relationship with the complexity of the model; a more complex model typically affords fewer degrees of freedom (Greene, 2003).

In the context of PBDAM, the complexity of the model, while allowing it to capture intricate customer journeys, can limit the degrees of freedom. This constraint may limit the model's flexibility and ability to account for random variations or unexplained behaviors within the data.

For future research, exploring methods of expanding the degrees of freedom within PBDAM would be advantageous. Techniques such as regularization, which introduces a penalty on the complexity of the model, could be a fruitful avenue for investigation. By maintaining an optimal balance between complexity and degrees of freedom, it might be possible to enhance the generalizability of PBDAM, thereby improving its robustness across diverse datasets and marketing scenarios.

REFERENCES

Abhishek, V., Fader, P., & Hosanagar, K. (2012). *Media Exposure through the Funnel: A Model of Multi-Stage Attribution.* Available at SSRN: http://dx.doi.org/10.2139/ssrn.2158421

Anderl, E., Becker, I., Von Wangenheim, F., & Schumann, J. H. (2016). Mapping the customer journey: Lessons learned from graph-based online attribution modeling. *International Journal of Research in Marketing*, *33*(3), 457-474. https://doi.org/10.1016/j.ijresmar.2016.03.001

Anderl, E., Schumann, J. H., & Kunz, W. (2016). Helping Firms Reduce Complexity in Multichannel Online Data: A New Taxonomy-Based Approach for Customer Journeys. *Journal of Retailing*, *92*(2), 185–203. https://doi.org/10.1016/j.jretai.2015.10.001

Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Berman, R. (2018). Beyond the last touch: Attribution in online advertising. *Marketing Science*, *37*(5), 771-792. https://doi.org/10.1287/mksc.2018.1104

Borden, N. H. (1964). The concept of the marketing mix. *Journal of advertising research*, *4*(2), 2-7.

Broadbent, S. (1984). 'Modeling with Adstock'. Journal of the Market Research Society. 16, 4, 295-312.

Buhalis, D., & Volchek, K. (2021). Bridging marketing theory and big data analytics: The taxonomy of marketing attribution. *International Journal of Information Management*, *56*. https://doi.org/10.1016/j.ijinfomgt.2020.102253

Chapelle, O., Manavoglu, E., & Rosales, R. (2014). Simple and scalable response prediction for display advertising. *ACM Transactions on Intelligent Systems and Technology*, *5*(4). https://doi.org/10.1145/2532128

Dahlén, M. (2005). The medium as a contextual cue: Effects of creative media choice. *Journal of advertising*, 34(3), 89-98. https://doi.org/10.1080/00913367.2005.10639197

Dalessandro, B., Stitelman, O., Perlich, C., & Provost, F. (2012). Causally motivated attribution for online advertising. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/2351356.2351363

de Haan, E., Wiesel, T., & Pauwels, K. (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. https://doi.org/10.1016/j.ijresmar.2015.12.001

De Haan, E., Wiesel, T., & Pauwels, K. (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. https://doi.org/10.1016/j.ijresmar.2015.12.001

di Stefano, G., Peteraf, M., & Veronay, G. (2010). Dynamic capabilities deconstructed: A bibliographic investigation into the origins, development, and future directions of the research domain. *Industrial and Corporate Change*, *19*(4), 1187–1204. https://doi.org/10.1093/icc/dtq027

Diebold, F. X., & Yilmaz, K. (2012). Better to give than to receive: Predictive directional measurement of volatility spillovers. *International Journal of Forecasting*, *28*(1), 57–66. https://doi.org/10.1016/j.ijforecast.2011.02.006

Dubé, J. P., Hitsch, G. J., & Manchanda, P. (2005). An empirical model of advertising dynamics. *Quantitative Marketing and Economics*, *3*, 107-144. https://doi.org/10.1007/s11129-005-0334-2

Engle, R. F., Ito, T., & Lin, W-L. (1990). Meteor Showers or Heat Waves? Heteroskedastic Intra-daily Volatility in the Foreign Exchange Market. *Econometrica,* Econometric Society, 58(3), 525-542.

Flavián, C., Gurrea, R., & Orús, C. (2020). Combining channels to make smart purchases: The role of webrooming and showrooming. *Journal of Retailing and Consumer Services*, *52*. https://doi.org/10.1016/j.jretconser.2019.101923

Gaur, J., & Bharti, K. (2020). Attribution Modelling in Marketing: Literature Review and Research Agenda. *Academy of Marketing Studies Journal*, *24*(4), 1-21. https://www.abacademies.org/articles/attribution-modelling-in-marketing-literature-review-and-research-agenda-9492.html

Ghose, A., & Todri-Adamopoulos, V. (2016). Toward a Digital Attribution Model: Measuring the Impact of Display Advertising on Online Consumer Behavior. *MIS Quarterly*, *40*(4), 889–910. https://www.jstor.org/stable/26629681

Google Inc. (2021). About attribution models. https://support.google.com/google-ads/answer/6259715?hl=en

Gotlieb, J. B., & Sarel, D. (1991). Comparative advertising effectiveness: The role of involvement and source credibility. *Journal of advertising*, 20(1), 38-45. https://doi.org/10.1080/00913367.1991.10673205

Greene, W. H. (2003). *Econometric analysis*. Pearson Education India.

Hanssens, D. M., Parsons, L. J., & Schultz, R. L. (1990). Response models in marketing. *Market Response Models: Econometric and Time Series Analysis*, 3-26. https://doi.org/10.1007/978-94-009-1073-7_1

Hanssens, D. M., Parsons, L. J., & Schultz, R. L. (2003). *Market response models: Econometric and time series analysis* (Vol. 2). Springer Science & Business Media.

Harmeling, C. M., Moffett, J. W., Arnold, M. J., & Carlson, B. D. (2017). Toward a theory of customer engagement marketing. *Journal of the Academy of Marketing Science*, *45*(3), 312–335. https://doi.org/10.1007/s11747-016-0509-2

Harsh, R., Acharya, G., & Chaudhary, S. (2018). Epistemological View: Data Ethics, Privacy & Trust on Digital Platform. In *2018 IEEE International Conference on System, Computation, Automation and Networking (ICSCA)* (1-6). IEEE.

Ji, W., Wang, X., & Zhang, D. (2016). A probabilistic multi-touch attribution model for online advertising. *International Conference on Information and Knowledge Management, Proceedings*, *24-28-October-2016*, 1373–1382. https://doi.org/10.1145/2983323.2983787

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

Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, *35*(1), 1–19. https://doi.org/10.1207/S15327906MBR3501_1

Johnson, J.W. (2001). Determining the relative importance of predictors in multiple regression: Practical applications of relative weights. In F. Columbus (Ed.), *Advances in psychology research*, 5, 231-251). Huntington, NY: Nova Science.

Joyee De, S., & Imine, A. (2019). On consent in online social networks: Privacy impacts and research directions (short paper). In: Zemmari, A., Mosbah, M., Cuppens-Boulahia, N., Cuppens, F. (eds) Risks and Security of Internet and Systems. CRiSIS 2018. Lecture Notes in Computer Science(), vol 11391. Springer, Cham. https://doi.org/10.1007/978-3-030-12143-3_11

Kakalejčík, L., Bucko, J., Resende, P.A.A. & Ferencova, M. (2018). Multichannel Marketing Attribution Using Markov Chains. *Journal of Applied Management and Investments*, 7 (1), 49-60. https://econpapers.repec.org/article/odsjournl/

Kannan, P. K., & Li, H. "Alice." (2017). Digital marketing: A framework, review and research agenda. *International Journal of Research in Marketing*, *34*(1), 22–45. https://doi.org/10.1016/j.ijresmar.2016.11.006

Kireyev, P., Pauwels, K., & Gupta, S. (2016). Do display ads influence search? Attribution and dynamics in online advertising. *International Journal of Research in Marketing*, *33*(3), 475–490. https://doi.org/10.1016/j.ijresmar.2015.09.007

Kohavi, R., & Longbotham, R. (2015). Online controlled experiments and A/B tests. *Encyclopedia of machine learning and data mining*, 1-11. https://doi.org/10.1007/978-1-4899-7502-7_891-2

Koop, G., Pesaran, M. H., & Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of econometrics*, *74*(1), 119-147. https://doi.org/10.1016/0304-4076(95)01753-4

Kotler, P., Kartajaya, H., & Setiawan, I. (2017). *Marketing 4.0: moving from Traditional to Digital*. John Wiley & Sons.

Leguina, J. R., Rumín, Á. C., & Rumín, R. C. (2020). Digital marketing attribution: Understanding the user path. *Electronics (Switzerland)*, *9*(11), 1–25. https://doi.org/10.3390/electronics9111822

Lewis, R. A., & Reiley, D. H. (2014). Online ads and offline sales: measuring the effect of retail advertising via a controlled experiment on Yahoo!. *Quantitative Marketing and Economics*, *12*, 235-266. https://doi.org/10.1007/s11129-014-9146-6

Li, H., & Kannan, P. K. (2014). Attributing conversions in a multichannel online marketing environment: An empirical model and a field experiment. *Journal of Marketing Research*, *51*(1), 40–56. https://doi.org/10.1509/jmr.13.0050

Li, N., Arava, S. K., Dong, C., Yan, Z., & Pani, A. (2018). Deep neural net with attention for multi-channel multi-touch attribution. https://doi.org/10.48550/arXiv.1809.02230

Liu, Y., Laguna, J., Wright, M., & He, H. (2014). Media mix modeling – A Monte Carlo simulation study. *Journal of Marketing Analytics*, *2*(3), 173–186. https://doi.org/10.1057/jma.2014.3

McCarthy, E. J. (1978). Basic marketing: a managerial approach. (No Title).

MSI 2022-24 Research Priorities. (2022). Marketing Science Institute. https://www.msi.org/wp-content/uploads/2022/10/MSI-2022-24-Research-Priorities-Final.pdf

Mukherjee, P., & Jansen, B. J. (2017). Conversing and searching: the causal relationship between social media and web search. *Internet Research*, *27*(5), 1209–1226. https://doi.org/10.1108/IntR-07-2016-0228

Naik, P. A., & Raman, K. (2003). Understanding the Impact of Synergy in Multimedia Communications. *Journal of Marketing Research*, *40*(4), 375–388. https://doi.org/10.1509/jmkr.40.4.375.19385

Parsons, L. J. (1981). Models of Market Mechanisms. In Marketing Decision Models, Randall L. Schultz and Andris A. Zoltners, eds. New York: North-Holland, 77-98.

Parsons, L. J., & Schultz, R. L. (1976). Marketing models and econometric research. North Holland, Amsterdam.

Perlich, C., Dalessandro, B., Hook, R., Stitelman, O., Raeder, T., & Provost, F. (2012). Bid optimizing and inventory scoring in targeted online advertising. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 804-812). https://doi.org/10.1145/2339530.2339655

Pesaran, M. H., Shin, Y., & Smith, R. J. (2001). Bounds testing approaches to the analysis of level relationships. *Journal of Applied Econometrics*, *16*(3), 289–326. https://doi.org/10.1002/jae.616

Porter, M. E., & Kramer, M. R. (2018). Creating shared value: How to reinvent capitalism—And unleash a wave of innovation and growth. In *Managing sustainable business: An executive education case and textbook* (pp. 323–346). Springer.

Raj, K. A. A. D., Kanagasabapathi, B., Shrivastava, S., Krishnan, K., & Shah, M. (2012). Performance evaluation of Adstock models using market drivers in the consumer packaged goods (CPG) industry. *International Journal of Electronic Marketing and Retailing*, *5*(2), 173-186. https://doi.org/10.1504/IJEMR.2012.051031

Ren, K., Fang, Y., Zhang, W., Liu, S., Li, J., Zhang, Y., Yu, Y., & Wang, J. (2018). Learning multi-touch conversion attribution with dual-attention mechanisms for online advertising. *International Conference on Information and Knowledge Management, Proceedings*, 1433–1442. https://doi.org/10.1145/3269206.3271677

Rutz, O. J., & Bucklin, R. E. (2011). From generic to branded: A model of spillover in paid search advertising. *Journal of Marketing Research*, *48*(1), 87-102. https://doi.org/10.1509/jmkr.48.1.87

Sangam, S. L. (2015). Bradford's empirical law. *Journal of Library Development*, *1*(1), 1-13.

Shao, X., & Li, L. (2011). Data-driven multi-touch attribution models. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 258-264).

Shapley, L. (1953). *A Value for n-Person Games*. In: Kuhn, H. and Tucker, A., Eds., Contributions to the Theory of Games II, Princeton University Press, Princeton, 307-317. https://doi.org/10.1515/9781400881970-018

Shovon, A. R., Roy, S., Shil, A. K., & Atik, T. (2019). GDPR compliance: implementation use cases for user data privacy in news media industry. In *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)* (pp. 1-6). IEEE. https://doi.org/10.1109/ICASERT.2019.8934660

Shuba, S., Marc, V. A., & Koen, P. (2010). Mind-set metrics in market response models: An integrative approach. *Journal of Marketing Research*, *47*(4), 672–684. https://doi.org/10.1509/jmkr.47.4.672

Sundar, S. S., Narayan, S., Obregon, R., & Uppal, C. (1998). Does web advertising work? Memory for print vs. online media. *Journalism & Mass Communication Quarterly*, *75*(4), 822-835. https://doi.org/10.1177/107769909807500414

Susmel, R., & Engle, R. F. (1994). Hourly volatility spillovers between international equity markets. *Journal of international Money and Finance*, 13(1), 3-25. https://doi.org/10.1016/0261-5606(94)90021-3

Vaver, J., & Koehler, J. (2011). *Measuring Ad Effectiveness Using Geo Experiments*. https://research.google/pubs/pub38355/

Xu, L., Duan, J. A., & Whinston, A. (2014). Path to purchase: A mutually exciting point process model for online advertising and conversion. *Management Science*, *60*(6), 1392–1412. https://doi.org/10.1287/mnsc.2014.1952

Zhao, K., Mahboobi, S. H., & Bagheri, S. R. (2019). Revenue-based attribution modeling for online advertising. *International Journal of Market Research*, *61*(2), 195–209. https://doi.org/10.1177/1470785318774447