

Samsami, Mahsa; Wagner, Ralf

**Article**

## Investment decisions with endogeneity: A Dirichlet tree analysis

Journal of Risk and Financial Management

**Provided in Cooperation with:**

MDPI – Multidisciplinary Digital Publishing Institute, Basel

*Suggested Citation:* Samsami, Mahsa; Wagner, Ralf (2021) : Investment decisions with endogeneity: A Dirichlet tree analysis, Journal of Risk and Financial Management, ISSN 1911-8074, MDPI, Basel, Vol. 14, Iss. 7, pp. 1-19, <https://doi.org/10.3390/jrfm14070299>

This Version is available at:

<https://hdl.handle.net/10419/258403>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*




<https://creativecommons.org/licenses/by/4.0/>



Article

# Investment Decisions with Endogeneity: A Dirichlet Tree Analysis

Mahsa Samsami<sup>1</sup> and Ralf Wagner<sup>2,\*</sup> 

<sup>1</sup> Departments of Business Management, University of Tehran, Kish 7941639982, Iran; mahsa.samsami@alumni.ut.ac.ir or samsami.mahsa@gmail.com

<sup>2</sup> Department of Sustainable Marketing, Faculty of Economics and Management, University of Kassel, 34125 Kassel, Germany

\* Correspondence: rwagner@wirtschaft.uni-kassel.de

**Abstract:** Ignoring endogeneity when assessing investors' decisions carries the risk of biased estimates for the influence of exogenous marketing variables. This study shows how to overcome this challenge by using Pólya trees in the quantification of impacts on investors' decisions. A total of 2255 investors recruited for this study received and opened a digital marketing newsletter about investing daily. Given the nature of investors' decisions characterized by heterogeneity and endogeneity, the response model is assessed with the Dirichlet process mixture and estimated with the Markov chain Monte Carlo method. Digital marketing substantially exceeds the impact of investor experience, but both have a significant positive impact on investors' trading volume. Findings obtained with the Dirichlet process mixture as a flexible model indicate that digital marketing even with latent endogenous factors makes an underlying contribution to the investors' actions in the stock market.



**Citation:** Samsami, Mahsa, and Ralf Wagner. 2021. Investment Decisions with Endogeneity: A Dirichlet Tree Analysis. *Journal of Risk and Financial Management* 14: 299. <https://doi.org/10.3390/jrfm14070299>

Academic Editors: Joanna Olbrys and Shigeyuki Hamori

Received: 2 April 2021

Accepted: 28 June 2021

Published: 1 July 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** digital marketing; Dirichlet process mixture; investor decision; lifetime value

## 1. Introduction

Despite the advancements in ethical investments and social business models (Chakrabarty et al. 2017; Lin et al. 2018; Martí-Ballester 2015) all over the world, a solid share of investors aim for profits and increase their capital regardless of occupation, education, and other background variables. Notably, a wide spectrum of information ranging from technical chart data to subjective feelings determines individuals' investment decisions. Frequently, heuristic-driven biases appear to impair the quality of short- and long-term investment decisions (Ahmad 2020). Investors are found to be sensitive to endogenous factors (e.g., price volatility), particularly towards daily news (Lillo et al. 2015). Investors aim to use this information to rationalize their decisions about investments and de-investments (Martin and Nisar 2007). They may also rely on the expertise of an advisor or invest their assets in mutual funds.

Investors who prefer to invest directly in the stock exchange are expected to consider both economic and political news along with stock exchange information and technical and fundamental analyses. With modern technology, the speed of information dissemination has changed to almost real-time. Investors easily access a host of information through various social networks, search engines, digital platforms, and other virtual environments but lack reliability filters, classifications, and contextualization. This is the value proposition of specialized advisors and various digital formats, such as online magazines, investment blogs, and newsletters.

For assessing information, investor expertise and experience are central (Dvořák 2005; Rzeszutek et al. 2015). Investor experience is captured by his or her lifetime value, which is understood through the timeline of their account opening. Brokerage firms and their

marketers can meet investors' needs via information syndication platforms or by using newsletters and other digital push communication formats.

In recent years, as can be seen, the financial industry has continuously thoroughbred evolution in service delivery owing to digitalization, which is specified by the enhanced speed of information processing in financial service processes (Gomber et al. 2017). Fintech has presented a great opportunity by providing customers with an extremely accessible way to fulfill investors' financial needs and criteria (Buckley and Webster 2016). Lee and You-Kyung in 2021 (Lee 2021) expressed the idea that fintech marketers need to develop media-based materials. Obviously, behavioral attributes are significantly influenced by both technological attributes and digital behavior (Singh et al. 2020).

Interestingly, the question is, to what extent is the behavior of investors related to these measures that challenge practitioners and researchers? The research question addressed in this paper is whether the experience (assessed by the investors lifetime value) or the information provided by digital push communication, or both, are driving investors' turnover in the stock market. Notably, investors' decision-making processes are not observed directly, but are latent and hidden. Various impacts—for instance, exchange rates, political affairs, sanctions, influencers, level of investing and financial knowledge, national culture, and investors' risk attitudes—affect investors' financial decision making (Petersen et al. 2015). A research stream has emphasized the importance of big data analytics in improving firm and investment performance (Goldstein et al. 2021; Ma et al. 2016; Maroufkhani et al. 2019; Olson and Chae 2012). However, the abundance of data does not compensate for analytic capabilities (Pauwels 2014) but instead calls for an improvement of analytical methodology for examining management decisions (Ranjan and Bhatnagar 2011; Saukkonen et al. 2018).

Notably, in order to measure not only market liquidity but also the dimensions of market liquidity drawn on intraday data, it is important to recognize the side that initiates a transaction and to distinguish between the so-called buyer- and seller-initiated trade (Olbrýs and Mursztyn 2018a). It is the validity of the market microstructure research that depends on the ability to classify accurately transactions being initiated as a buyer or seller (Olbrýs and Mursztyn 2018b). A precise quantification of the impact is needed to understand investors' decision-making processes (Abdallah and Hilu 2015) but is hampered by the following.

Firstly, the covariance of exogenous impacts, such as national economic development, changes in exchange rates, and individual companies' performance indicators. This leads to multicollinearity (Talay et al. 2019) and, therefore, likely result in biased estimates in econometric analysis (Lavery et al. 2019).

Secondly, there is heterogeneity of investors' decision processes (Dragotă and Șerbănescu 2010; Jalilvand et al. 2018). Only one component of this heterogeneity can be attributed to observable differences among investors, such as age, gender, or nationality. The core of human decision processes is unobservable and therefore calls for a model approximating heterogeneity using random distribution functions or mixture of distribution functions (Jedidi et al. 1997; Decker and Wagner 2000). Endogeneity results from learning of and adoption by companies and investors in the market. Therefore, the exogenous variables are not independent from error components in any response models (Chintagunta et al. 2006; Schröder and Hruschka 2012). Ignoring endogeneity implies the risk of biased estimates for any effects (Ebbes et al. 2009; Wansbeek and Wedel 1999).

According to rapid changes in the digital marketing environment, the importance of a practical understanding of digital marketing tools is observed (Nikunen et al. 2017; Yamin 2017). In a study of mail marketing effects, Schröder and Hruschka (2017) assessed three alternative approaches for modeling unobserved heterogeneity on account of finite mixtures, continuous mixtures, and a mixture of Dirichlet processes. The use of Dirichlet processes mixture (DPM) models has become computationally feasible with the development of Markov chain methods for sampling from the posterior distribution of the parameters of the component distributions and of the associations of mixture components

with observations (Neal 2000). The results of this study clarify that the number of segments and components that must be specified a priori to calibrate finite mixture models have a substantial impact on the results. However, besides fully controlled experimental studies this a priori knowledge is hardly given. This study contributes by introducing a DPM model for mastering the challenges of multicollinearity, heterogeneity, and endogeneity. It also demonstrates the model's calibration and provides the empirical evidence of the impact of digital marketing efforts and investor lifetime value.

Challenging the research objective (the influence of digital marketing in the internet area drawing on newsletter on investor behavior), it is assumed that investor lifetime (time of presence in the capital market since the opening of the account) need also to be considered. The empirical part of this study adopted a newsletter as one of the digital marketing tools that have a substantial impact in motivating investors to react in the stock market. This communication tool consolidates news, fundamental and technical analytical reports, etc., in new quality content that gives rise to a substantial impact potentially outnumbering the impact of investors' previous experience.

Guindani and Johnson (2018) and Cipolli and Hanson (2019) provide the example of recent application domain using PT for classification and response modeling purposes. Although Denison and Mallick (2007, p. 4) argued that "Polya trees can be used to define probability distributions on any space of Interest", only very few applications in the domain of modeling the response to marketing measures we identified (Schröder and Hruschka 2012; Schröder and Hruschka 2017). Moreover, this study demonstrates a novel approach of considering the heterogeneity of decisions in empirical research by taking advantage of Pólya trees in the model fitting procedure. For modeling investors' decision making, considering the impact of the effect of latent variables on the response variable is advantageous. In sending the newsletter as a digital marketing tool, modeling unobserved heterogeneity with the DPM model and finite mixture is emphasized (Schröder and Hruschka 2017). Thus, although some unobserved factors can affect investor behavior, digital marketing tool plays a fundamental and significant role in investor behavior.

The remainder of the paper is structured as follows. Next, related research on investors' decisions and fitting Dirichlet process mixture models is discussed. Then, the research design, methodology, and data are described. Thereafter model fitting results are reported, and conclusions, limitations, and future research opportunities are outlined.

## 2. Related Research

Digital mail as a means of marketing communication is differentiated from other communication instruments by the trustworthiness attributed by the recipients (Feld et al. 2013), which is very important (Nikunen et al. 2017)—just as it is in the finance industry (Rizzo 2018). Information conveyed by mail supports companies in persuading customers to purchase (Schröder and Hruschka 2012) or investors to buy stocks or index certificates (Lillo et al. 2015; Schröder and Hruschka 2017). However, catalog exposure has a more long-lasting effect (Venkatesan et al. 2007; Zantedeschi et al. 2016). Wu et al.'s (2018) empirical findings showed that email contributes to increasing customer purchase rates over the short- and long-term. Consequently, digital mail is critical for decreasing the operational cost of digital marketing, and even a reduced response rate from customers has a considerable impact on the success of a marketing mail campaign (Bult and Wansbeek 1995; Gönül et al. 2000; Gopal et al. 2001). In addition, social networks perform a key role in the formation and dissemination of service quality preference, firm performance, and innovation capacity (Imrie Brian 2013; Palacios-Marqués et al. 2015). Studies of media habits provide evidence that the use of social networks in marketing also affects the behavior of members significantly (Adrian and Nicole 2009; Palalic et al. 2020; Tikkanen et al. 2009).

Modern digital communication production technologies in combination with growing databases enable digital marketers to design and customize emails, newsletters, catalogs, and other communication instruments separately for each customer. Thus, increasing literature recognizes the importance of personalized marketing as a marketing tool because

it is considered an essential factor in predicting customers' preferences and advertising customized products or services (Tang et al. 2013). Personalized marketing has accentuated contextual information, through customer's current environment/situation or pricing (Liu et al. 2020; Tang et al. 2013; Yelkur and DaCosta 2001).

### 2.1. Digital Marketing for Financial Services

Digital marketing changed rapidly in recent years (Nikunen et al. 2017), being considered as a key factor owing to full integration with other marketing and communication activities (Schultz and Peltier 2013), and secured improvements in testing, database quality, and timing in financial services marketing (Thwaites and Lee 1994), particularly for banks to cope with future changes and competition (Alan et al. 2018; Miguéis et al. 2017; Moyo 2018; Parlar and Acaravci 2017). Relationship-based interactions with their customers are focused by the vendors in order to improve digital marketing engagement (Tiago and Veríssimo 2014). Furthermore, marketing methods and analysis of huge amounts of customer electronic data (Elsalamony 2014; Sharma et al. 2015) have received significant attention in the financial and insurance industries as the campaign cost is minimized, and the return rate tends to be maximized (Asare-Frempong and Jayabalan 2017; Miguéis et al. 2017; Mitik et al. 2017; Trivedi et al. 2020). Recording and combining customer data is needed to advance marketing practices (Shah and Murthi 2020), and recent developments in technology have led to improving the performance of bank marketing campaigns (Elsalamony 2014; Parlar and Acaravci 2017). Liao et al. (2004) argued that electronic catalog design and marketing are served as a facilitator in merging catalog and web-based marketing drawing on market segmentation.

Taiminen and Karjaluoto (2015) argued that SMEs failing the fundamental change of communication by digital tools will not fully exploit the potential benefits they provide. In the financial industry, digital marketing is attracting increasing attention because marketers can target a profitable set of consumers. Thus, customers also may find an existing product or a new product interesting (Barman et al. 2016). Asllani and Halstead (2015) used data analysis to provide crucial information regarding investor behavior in tandem with financial behavior in the stock market. Professional and individual insights are part of investment decisions. These methods include newspapers and other media, noise in the market as well as fundamental and technical stock price analysis.

Notably, two decades ago Lee (2002) identified a cluster of banking clients ready to forgo personal interactions and to receive investment services through digital marketing. Banks as financial institutions similar to brokerages retain a solid amount of information about their customers through which they can customize campaigns for customers through a multitude of channels, including mail, email, phone, and in person (Miguéis et al. 2017; Mitik et al. 2017; Parlar and Acaravci 2017).

Rapid changes in the availability of information on the web have considerable influence on information gathering, processing, and interaction among investors in the stock market. Lin et al. (2016) and Bank et al. (2019) provided evidence of digital (marketing) communications and social media's impacts on investor behavior. One of the most important point refer to the financial behavior of investors in buying and selling. Jalilvand et al. (2018) stated that Iranian investors are not homogenous. Investors in the stock market behave differently across demographic, psychological, and economic variables, as well as investors' concern such as sources of information, investment criteria, and attitude toward risk (Dragotăand Șerbănescu 2010). These sources of (unobservable) heterogeneity in the data need to be considered when fitting a model for analyzing investors' decisions.

### 2.2. Fitting Dirichlet Process Mixture Models

Given that the nature of the hierarchical extension is independent of the specific non-parametric model and the problem of combining inference in related nonparametric Bayesian models, Müller et al. (2004) introduced a Dirichlet process (DP) mixture model and developed a Markov chain Monte Carlo (MCMC) method. The DP generates distributions

that are discrete with a cumulative probability of one, making it awkward for continuous density estimation. This limitation can be fixed through convolving its trajectories generally by using DP random measures as the mixing measure in a mixture over some simple parametric forms (Müller et al. 2015). One of the critical properties is the a.s. (almost surely) discreteness of a random measure  $F \sim DP(M, F_0)$  (De Iorio et al. 2004). Müller and Mitra (2013) expressed that a DP random measure is a.s. The a.s. discreteness of the DP is awkward enabling a generalization of the proposed hierarchical extension to other nonparametric models beyond the DP (Müller and Mitra 2013; Müller et al. 2004). For each inference problem including regression and hierarchical models, Müller and Quintana (2004) provided an outline of nonparametric Bayesian models and approaches including DP models and variations, and Pólya tree (PT) extensions of parametric models.

The nonparametric Bayesian approach defines full support in the space of distributions for multiple unordered categorical variables. This support without the constraint of the dependence structure can be accomplished through a DPM of product multinomial distributions, which is also a convenient form for posterior computation (Dunson and Xing 2009). Hanson and Johnson (2002) stated that by marginalizing the PT, an exact inference is possible up to the MCMC error. Considering various model evaluations, it becomes evident that logit models frequently constitute a good choice. However, Bayesian neural nets and Tobit models turn out to be good alternatives (Schröder and Hruschka 2012). Given three alternative approaches for modeling unobserved heterogeneity, which are driven by finite mixtures, continuous mixtures, and DPM, which was preferred for this study model? Schröder and Hruschka (2012) concluded that according to some research, finite mixtures bring about a poorer performance than their continuous rivals; however, continuous mixtures fail to reproduce multimodal as well as skewed distributions. Moreover, a finite mixture is regarded as superior to the continuous mixture with respect to pseudo-Bayesian factors (Schröder and Hruschka 2017), and it does not provide a sound approximation or interpretation considering the infinity category. Consolidating these results, the DPM turns out to be a promising candidate for analyzing investors decisions.

### 3. Research Design and Data Contextualization

#### 3.1. Data Description and Response Component

##### 3.1.1. Sampling

The research data in this study is drawn from two main sources. One database originated from newsletter software which consists of daily fundamental and technical reports about stocks, economic and political news, and other related news. These newsletters are sent to customers and opened by most of the investors on a daily basis. The relevance of the newsletters' content is validated expert judgments in a procedure adopted from Scholz and Wagner (2005). Data for four years, beginning in May 2014 and ending in May 2019, were pooled: 7335 investors investing in the stock market. A total of 2255 investors recruited for this study received and opened the newsletter daily. More than three-quarters of the sample was male (78.84%).

The other data source, an investor information database from a stock exchange brokerage firm, contains investors' personal information (ID customer, email address, postal code, investor lifetime, and bank account) along with information about investors' financial transactions in the capital market and investing in the stock market either directly or indirectly, that is, their turnover (buying and selling in the stock market), buying and selling commissions, and account balance).

##### 3.1.2. Measurement Turnover

###### 1. Turnover

Jalilvand et al. (2018) stated that investors in Iran are not homogenous, and the financial behavior of investors in buying and selling in the capital market differs across demographic, psychological, and economic variables, as well as investors' concern such as sources of information, investment criteria, and attitude toward risk (Dragotäand

Șerbănescu 2010). The response variable ( $y$ ) is related to turnover during four years, which is included the amount of selling and buying of investors. It is a numerical continuous amount scale.

## 2. Newsletter

Investors react differently to exogenous (e.g., news) and endogenous (e.g., price volatility) factors (Lillo et al. 2015). Thus, the investor community is assumed to be a multi-category. Similar to other decision makers, who react differently to the same marketing due to differences (Guelman et al. 2015), capital market investors' behavior has been subject to considerable discussion.

The Mailerlite package has been used for the production of the newsletters and MailChimp for the monitoring of each newsletter's performance by means opening rate and time. This enabled an iterative improvement by adjusting titles and content to the recipients' interests.

Email activity and subscribers' engagement is tracked by the website admin using Mailerlite, who is experienced as a customer relationship manager in brokerage. Data records cover these activities and subscribers' engagement by means of unsubscribing counts, spam complaints, bounces, link activity, and other metrics that are relevant to assess the success such as tracking of sends and openings. Relevant to this study, the response measurement provides evidence who has received the newsletter and who has opened it. Subscribers are tracked on an individual level.

For fitting the model, the newsletter as a covariate variable ( $X_1$ ) is the arithmetic mean of four years of opening by the individual investor.

As soon as investor subscribe and submit his/her email for receiving newsletter, newsletter will be sent to his/her inbox from the same day. Although investors receive their newsletters everyday, they are unlikely to open all of them immediately, some remain unread at all. For instance, a typical investor may receive around 3000 newsletters in the time window of four years, but open just around 2000 out of this set.

Notably, the number of newsletters sent per day varies with respect to the events relevant to the stock market. Moreover, the number is not equal for the subscribers, because the subscription can be customized to the investors' interests (company or industry reports, fundamental and technical analysis, governmental decisions, etc.). So, for our analysis, we considered the arithmetic mean of openings by the individual newsletter recipient calculated of a four-year period.

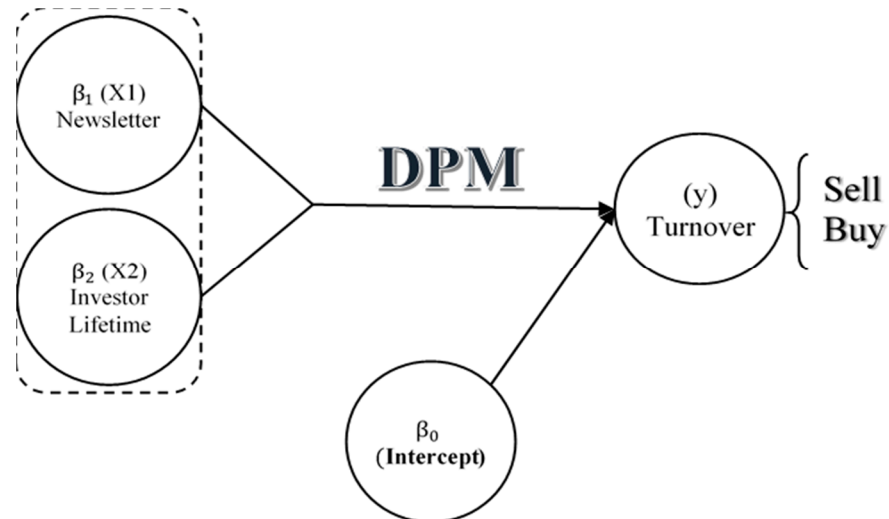
## 3. Investor lifetime value

Predicted customer lifetime value is of interest for companies because the issue of how and when customers might break their relationship with companies is still controversial (Chang and Zhang 2016). In addition, customer lifetime value could be a contributing factor in building and fostering a long term and strong relationship between customers and firms and eventually making more profit and growth in competitive markets (Ferrentino et al. 2016). Therefore, mapping risk-taking investors' preferences as informative or not, investor lifetime plays a fundamental and pivotal role in customers' investment in the stock market. Due to the point that estimation of value produced by customers over their lifetime is essential for many companies and serves as a facilitator for customers segmentation, it is essential to analyze their behavior and allocating resources or formulating strategies (Ferrentino et al. 2016). Importantly, in a financial services application, analyzing customer lifetime value provides more in-depth insights into the effects of the covariates (Benoit and Van den Poel 2009).

In this respect, investor lifetime along with responses to the newsletter were expected to provide a substantial explanation for investors' behavior. Thus, analyzing the effect of both covariate variables on the response variable  $y$  (turnover for four years = selling + buying), which is the continuous amount, can provide crucial information about investor behavior. Investor lifetime value is evaluated by the length of time investors who are active in the capital market from the date of the account is opened. It is assessed by the number of days

in which investors have been investing in the stock market through their account in the brokerage firm.

Figure 1 illustrates that we consider  $X_1$  (the daily average calculated over four years newsletters opened by individual investors) and  $X_2$  (the length of time investor is active in the capital market from the date the account is opened) as exogenous and their influences on  $y$  as the response variable. This connection will be evaluated by DPM.



**Figure 1.** Research conceptual model.

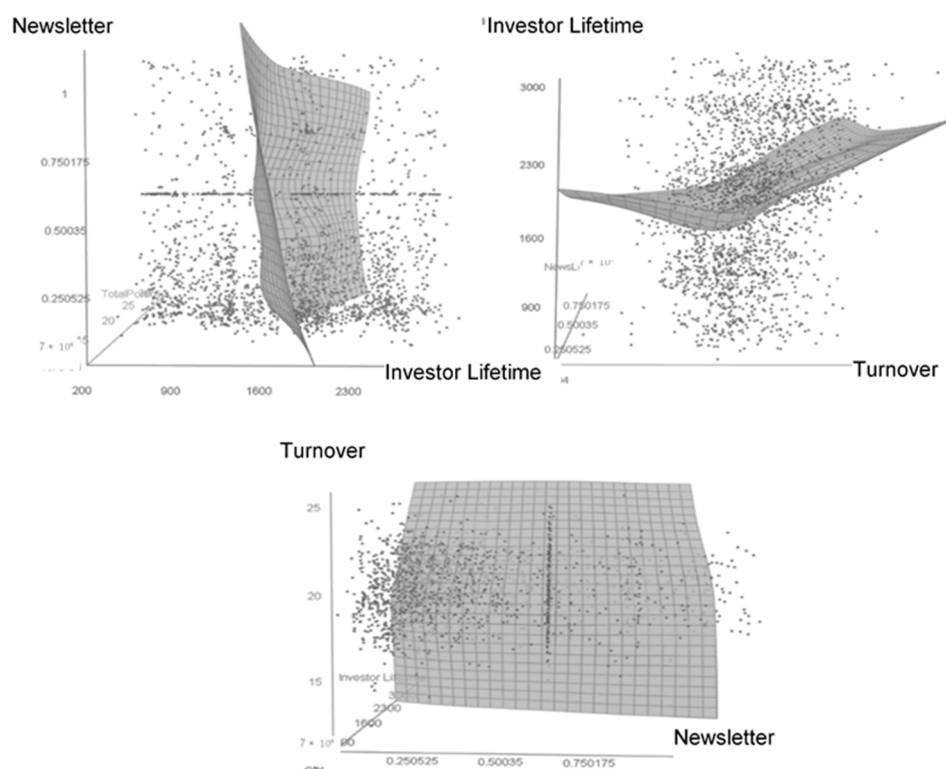
### 3.2. Endogeneity in the Model Specification and Dirichlet Process Mixture Model

An error ( $\epsilon$ ), or more accurately, measurement errors are caused by an instrument. For instance, when investors open an account, staff may record some of the information incorrectly, or investors make mistakes when they sign up for the newsletter. During the process of obtaining data from a database, all collected samples have errors: For the most part, they are unavoidable, and it is worth considering net errors like them. Exogenous variables are not independent of error components in any response models (Chintagunta et al. 2006; Schröder and Hruschka 2012).

Another component are unavoidable random effects. Sending a newsletter as an important digital marketing tool is considered an observed heterogeneity, an independent variable. Moreover, the auxiliary variable investor lifetime is discussed in the model. However, there may be other auxiliary variables that are unobserved components and cannot be considered in the model or are not visible, including education level, investor place of residence, level of knowledge about capital markets and investments, influencers, news and analysis content in the newsletter, and others.

Previous studies on digital marketing did not deal with latent variable descriptions. No a priori assumption on the exact amount of cardinality of segments or components that is necessary for a finite mixture approach need to be justified. For visual inspection, the two-covariate variables and one response variable of the model are plotted in Figure 2. As can be seen, it is the number of clusters that is controversial. Among some researchers, calculating the number of clusters appeared to be controversial, and it is questionable.





**Figure 2.** The 3D plot of data according to the conceptual decision model.

The nonparametric Bayesian approach without the constraint of the dependence structure is fitted via DPM (Dunson and Xing 2009). As Müller et al. (2004) introduced a DP mixture model and developed a Markov chain Monte Carlo (MCMC) method, the hierarchical extension is proven to be independent of the specific non-parametric model, which is the problem of combining inference in related nonparametric Bayesian models. Müller and Quintana (2004) provided an outline of nonparametric Bayesian models to inference problems, which is related to regression and hierarchical models. In addition, these approaches include DP models and variations, and PT models. Finite mixtures, continuous mixtures, and DPM are alternative approaches for modeling unobserved heterogeneity. As Schröder and Hruschka (2012) expressed, finite mixtures are likely to provide a poorer fit than their continuous alternatives. However, continuous mixtures cannot reproduce multimodal as well as skewed distributions. The finite mixture is compatible with continuous mixture due to pseudo-Bayesian factors (Schröder and Hruschka 2017). Random effects in generalized linear mixed models are mapped as latent states or cluster assignments. Most studies must assume latent variables are normally distributed (because of heteroscedasticity in regression) or follow an extreme value distribution in logistic responses. The nonparametric Bayesian approach enables to overcome this limitation by allowing that the distribution of random effects is unknown (Dey et al. 2012). An important application of nonparametric approaches brings about in modeling random effects distributions in hierarchical models (Müller and Quintana 2004). Additionally, in the nonparametric context, modeling is complicated by the point that the random quantities over which is defined the hierarchy are infinitely dimensional (Müller et al. 2004). This approach contains a regression at the level of the nonparametric model, which was in Müller et al.'s (2004) investigation of the special case of DPM.

The DPM is preferable to a finite mixture or continuous mixture because the selection of components is a constraint, and the number of segments and levels of the partition community can be estimated by a parameter of the model. Additionally, the selected distribution for errors or components is not necessarily identified as a fixed distribution, which is predefined in advance. Instead, it can be chosen from a variety of different

distributions with one probability. In fact, we no longer impose those random effects as normal or other distributions. Thus, this model is designed to be more flexible.

We consider a model with the following structure:

$$y_i = x_i\beta + v_i(i = 1, \dots, n) | v_i | G \sim G \quad G | \sigma^2 \sim DPM(.) \tag{1}$$

where the distribution  $G$  is unknown and random.  $\beta$  is the coefficient vector of the regression model,  $x_i$  is related to the covariate vector along with  $y_i$  which is the response variable, and  $v_i$  represents the latent variable that is supposed to come from a DPM.

For fitting this model to the data at hand, PT priors can put the positive mass in every relative entropy neighborhood of every positive density with finite entropy. In addition, empirical Bayesian models using DPs are generalized to PT (Lavine 1992, 1994), and a flexible nonparametric approach for modeling time-to-event data is presented through mixtures of PT (Hanson and Johnson 2002). PT priors, as DP priors, put a distribution on a collection of distributions. For a distribution  $F$ , a PT is constructed by dividing the sample space into finer-and-finer sets using successive binary partitioning. By default, it fits a median regression model using a scale mixture of PTs before the distribution of the errors (Hanson and Johnson 2002; Lavine 1992, 1994):

$$G | \sigma^2 \sim PT(P_i^{\sigma^2}, \alpha) \tag{2}$$

where  $\alpha$  is a principal determining factor in the number of clusters (“precision parameter”); it can be assumed that  $\alpha$  either is a constant number or follows a distribution function. Additionally,  $\sigma^2$  is the variance of the base distribution in DPM. Completing the model specification, independent hyper priors are assumed:

$$\begin{aligned} \alpha | a_0, b_0 &\sim \gamma(a_0, b_0) \\ \sigma^{-2} | \tau_1, \tau_2 &\sim \gamma\left(\frac{\tau_1}{2}, \frac{\tau_2}{2}\right) \end{aligned} \tag{3}$$

The precision parameter  $\alpha$  of the PT prior can be considered as random, following a two-parameter  $\gamma$  distribution,  $\gamma(a_0, b_0)$ , or fixed at some particular value.  $a_0$  and  $b_0$  denote hyper-parameters for  $\alpha$ . One way to select hyper-parameters is with a manual search; in other words, trial and error. For assessing prior distribution, we need to propose initial values for  $a_0, b_0, \tau_1$ , and  $\tau_2$  in the burn-in phase.

In the Bayesian approach, to consider a prior, sigma is regarded as inverse  $\frac{1}{\sigma^2}$ . Choosing sigma small promises the estimate from the later distribution to be highly accurate. The  $\tau_1$  and  $\tau_2$  hyper-parameters are considered inverse as well. We supposed  $a_0$  and  $b_0$  to be 8 and 3.5, and set  $\tau_1$  to 1.70 and  $\tau_2$  to 2.10. For initializing,  $\mu$  is considered equal to zero and the variance equal to 10, providing an uninformative prior.

Posterior distributions are drawn, and the average of the data is considered the parameter estimate. It is of interest that they are considered as a random sample from the posterior distribution. The MCMC parameters are shown and explained in Appendix A and Table A1.

Frequently, linear model structures have an intercept  $\beta_0$  estimating the average turnover for each PT assuming no impact arising from newsletter opening or from investors’ lifetime value. In this status, the PT average will be null. In other words, the PT is centered on a  $N(0, \sigma^2)$  distribution, by considering each  $m$  level of the partition (mixture component),  $P_i^{\sigma^2}$  to coincide with the  $k/2^m, k = 0, \dots, 2^m$  quantile of the  $N(0, \sigma^2)$  distribution. However,  $\beta_0$  similar to  $\alpha$  can be considered a constant number or can be based on a distribution. In the same manner,  $\beta_0$  is drawn from a distribution. In this case, the following PT prior is considered:

$$G | \mu, \sigma^2 \sim PT(P_i^{\{\mu, \mu^2\}}, \alpha) \tag{4}$$

Although in the prior PT was centered on 0 with  $\sigma^2$ , in this case, its average can be considered an amount that is non-zero, but  $\mu$ . Consequently, a prior will be taken into

account for  $\mu$ , where the  $PT$  is centered on the  $N(\mu, \sigma^2)$  distribution. The intercept term is automatically excluded from the model, and the hyper-parameters for the normal prior for  $\mu$  must be identified. The normal prior is expressed by:

$$\mu | \mu_b, S_b \sim N(\mu_b, S_b) \tag{5}$$

In this sense, it can be conceived that  $PT$  is in  $\mu$  points and is not in central of zero. Thus, we say that the intercept does not exist in the linear regression. We can have one  $\mu$  for a different subject (different  $\beta$ ), and we put  $\mu$  instead of a constant number.  $S_b$  is the variance of  $PT$ .

We used a Bayesian analysis function from the DP package library in R language to generate samples from the posterior density of a semiparametric linear regression model to estimate the parameters using a mixture of  $PT$  prior to the distribution of the errors (Hanson and Johnson 2002; Lavine 1992, 1994). The model is estimated with the MCMC simulation method. Use of DPM models has become computationally feasible with the development of Markov chain methods for sampling from the posterior distribution of the parameters of the component distributions and/or of the associations of mixture components with observations (Neal 2000).

MCMC is a simulation technique and a family of techniques that allow the fitting of complex models in a Bayesian framework along with sampling probability distribution:

Given

$$X^{(t)} = x^{(t)}$$

1. Generate

$$Y_t \sim q(y | x^{(t)})$$

2. Take

$$x^{(t+1)} = \begin{cases} Y_t & \text{with probability } \rho(x^{(t)}, Y_t) \\ x^{(t)} & \text{with probability } 1 - \rho(x^{(t)}, Y_t) \end{cases}$$

where

$$\rho(x, y) = \min \left\{ \frac{\bar{\pi}(y)}{\bar{\pi}(x)}, \frac{q(x|y)}{q(y|x)}, 1 \right\}$$

In the computational implementation of the model, Metropolis–Hastings steps are used to sample the posterior distribution of the regression coefficients and hyper-parameters. Just like a conventional independent sample, the Markov chain sample is summarized. The mean, variance, as well as percentiles as sample features, are used for corresponding features of the objective distribution (Link and Eaton 2012).

#### 4. Results

Results are obtained by applying the Bayesian analysis for a semiparametric linear regression model of the DP package in the R library. This function generates a posterior density sample from a semiparametric linear regression model using a mixture of Poly Trees prior, in the order of the distribution of the errors  $y_i = x_i\beta + v_i (i = 1, \dots, n)$  (e.g., Lavine 1992, 1994; Hanson and Johnson 2002).

To mitigate the undesirable sensitivity of the  $PT$  prior to partition boundaries, which is an important limitation of the  $PT$  prior, an additional mixing with respect to the centering measure is introduced. Thus, Müller and Rodriguez in 2013 expressed the idea in which it is assumed that the  $PT$  prior is proven to be centered via defining the nested partition sequence to be determined through dyadic quantiles of a desired prior mean. Subsequently, it can be seen that posterior inference is improved under a  $PT$  mixture prior (Müller and Rodriguez 2013).

We evaluated the effect of digital marketing along with investor lifetime on investors' portfolios and investor behavior after receiving reports and news in a daily newsletter. We have a three-stage structure. The portfolio as  $y$  is the response and dependent variable, and  $x_1$  and  $x_2$ , the newsletter and investor lifetime, respectively, are the covariate variables. These variables along with  $G$  and  $\gamma$  inverse distributions can construct or complete a linear model structure. After modeling, the linear model was fitted, and ANOVA revealed that not only was there a significant effect between the two variables, but the newsletter (pseudo-contour probabilities  $< 0.01$ ) and investor lifetime (pseudo-contour probabilities  $< 0.01$ ) have a statistically significant effect on investor behavior.

$\beta$  is the coefficient vector of the regression model. Due to the special structure of DPM, their estimated coefficient (e.g.,  $\beta_0, \beta_1, \beta_2$ ) was proven to be significant, whenever included zero. This result was statistically significant at the 0.05 level and the 95% confidence interval.

The output lists the estimated effect for newsletter  $\hat{\beta}_1 = 0.186$  and for investor lifetime,  $\hat{\beta}_2 = 0.000737$ . Consider that the intercept term in the posterior information for regression coefficients in the output (the "fixed" effects) corresponds to the mean of the frailty's distribution  $G$  in Equation (4). It is survival analysis that the hazard function is particularly useful in because it describes the way in which the instantaneous probability of failure for an individual change with time. For parametric inference, it needs to make assumptions when it comes to the distribution of failure times. Therefore, it can be common to consider nonparametric models (Wienke 2010). The simplest non-parametric estimate of a distribution function is proven to be the empirical distribution function, which means even in the case of continuous distributions, it is estimated by a discrete distribution. The posterior median (95% credible interval) for  $\sigma^2 = 5$ . The acceptance rates for the Metropolis-Hastings steps associated with the regression coefficients, frailties, centering variance, and precision parameter 0.256069, 0, 0.4333262, 0.4516119, respectively (Jara et al. 2011). The main differences are that the 95 per cent HPD regions for newsletter and investor lifetime are much wider than the 95 per cent confidence intervals (Kleinman and Ibrahim 1998).

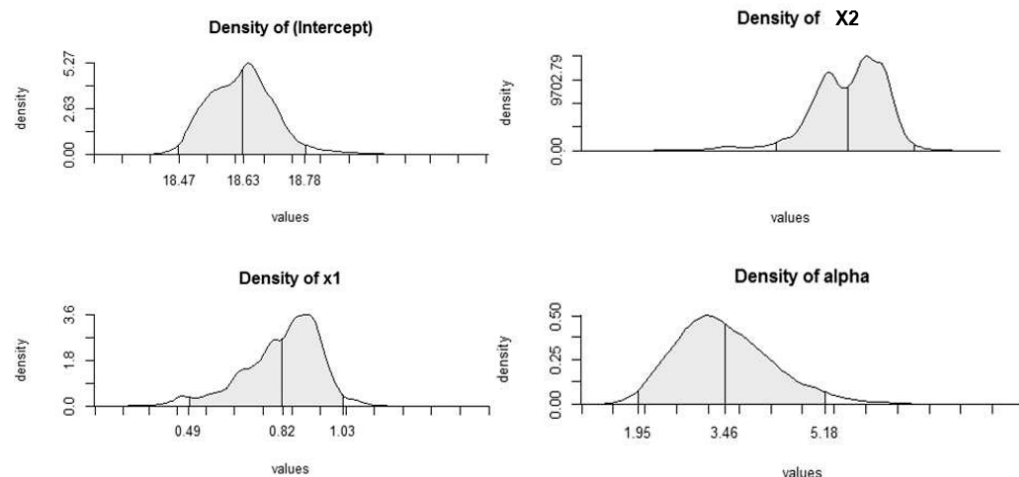
The posterior mean (95% HPD interval) for  $\beta_1$  (newsletter) under the normal assumption for the random effects is 0.8169000 (0.49310000; 1.03000000), and the posterior mean and 95% HPD associated with  $\beta_2$  (investor lifetime) is 0.0007375 (0.00066980; 0.00080010). In the Bayesian estimation, the highest posterior density (HPD) is of importance. Notably, the lower and upper bounds of the HPD do not include zero, as illustrated in Table 1; therefore, the estimated parameters for  $\beta_1$  (newsletter) and  $\beta_2$  (investor lifetime) differ significantly from zero. The boundaries of 95% HPD regions cover the estimated parameters. Notably, this result implies the inference on fixed effects to be robust to the non-normality of random effects in linear mixed-effects models (Vallejos 2008).

**Table 1.** Bayesian semiparametric regression model and the highest posterior density.

	Mean	Median	Standard Deviation	Naïve Std. Error	95% HPD-Low	95% HPD-Upp
$\beta_0$ (Intercept)	18.630000	18.6300000	0.0850100	0.0006011	18.4700000	18.7800000
$\beta_1$ (Newsletter)	0.8169000	0.8423000	0.1339000	0.0009468	0.49310000	1.03000000
$\beta_2$ (Investor Lifetime)	0.0007375	0.0007433	0.0000362	0.0000002	0.00066980	0.00080010
<b>Precision parameter:</b>						
sigma2	5.953570	5.957104	0.321260	0.002272	0.000000	0.000000

Acceptance rate for metropolis step = 0.256069, 0, 0.4333262, 0.4516119.

The amount of minimum, first quarter, median, third quarter and maximum are illustrated in Table A2. Log posterior predictive distribution varies between  $-8$  and  $-1$ . Not only does this amount approach zero, but also the range of change is acceptable since the probability ( $\rho$ ) range moves to one. Figure 3 depicts the density of the parameter estimates.



**Figure 3.** Histories and plots of metropolis steps.

Obviously, the densities of all four estimates are non-normal, although 20,000 observations are captured. This result indicates the suitability of the PT for fitting the response structure outlined in Figure 1. An additional assessment of the model fitting process (Appendix C and Figure 3) provides evidence of the histories and histograms of the metropolis steps, which is showing target distribution generated samples at different iteration points.

The histogram in Appendix C shows whether there is correlation. If there is a particular trend in the diagram, that is, dependence. However, if a trend is not observed, there's a random that is desirable. As elaborated in Appendix C, histograms are heterogeneous (Wiggly), being acceptable. On the other hand, plots in Figure 3 are almost normal, which shows a mixture model with the normal core and  $G \sim DP(\alpha G_0)$ .

## 5. Discussion, Limitations, and Future Research

This study proposed a model addressing heterogeneity in latent variables resulting from the fact that investors behave and react differently. The Bayesian estimation of the effects of the explanatory variables enables to capture unobservable heterogeneity and essentially reveals probability distributions for unknown quantities of interest given in the data. Therefore, in this study Bayesian approaches were considered instead of other models, such as regression.

In the application domain of investors' decisions, this study challenged the questions of the relevance of investors' lifetime value that accumulated over years in contrast to the short-term effect of digital marketing communications with a newsletter. In answer to our research question, experience (assessed by the investors lifetime value) or the information provided by digital push communication or both are driving investors' turnover in the stock market. Particularly, investors' decision-making processes are not observed directly but are latent and hidden. Following previous suggestions, the challenge of unobserved heterogeneity was addressed with the Dirichlet mixture model (Hruschka 2010; Schröder and Hruschka 2012, 2017). Given that the data were multi-categorical, the mixture model is suitable. However, endogeneity and latent variables should be addressed especially in digital marketing research, which is consistent with previous studies Hruschka (2010) Schröder and Hruschka (2017).

One of the most significant findings to emerge from this study is that digital marketing along with other hidden or unobserved factors plays a significant role in financial behavior and their reaction for buying and selling in the stock exchange, which is in line with findings by [Jalilvand et al. \(2018\)](#). Interestingly, as [Lillo et al. \(2015\)](#) concluded, the present results showed that investors react to news and information with other endogenous factors, such as price returns and volatility.

Regarding [Hruschka \(2010\)](#) and [Schröder and Hruschka \(2017\)](#), the second major finding was that investor experience, which is related to the date investors open accounts to invest in the stock market, has a positive significant effect on financial behavior, even if there is a host of factors that cannot be identified and observed.

Our starting point was theorizing about the role of digital marketing with news content, analysis, and information in financial behavior. One of the most important points in analytics is related to understanding and considering the factors affecting the use of analytics in marketing, namely, considering latent factors by using a flexible model.

DP models as one of the most popular nonparametric Bayesian models are a flexible model because considering components as a constraint is not necessary for this model, and the number of segments and the level of the partition community can be estimated with a model parameter. Additionally, the proposed model does not consider parametric distributions for random effects. Thus, digital marketing even with unobserved endogenous factors makes an underlying contribution to the investor's actions in the stock market.

A company leading in the capital market should use this low-cost and effective tool. The interesting result for theory building is that the parameter of the newsletter substantially exceeds the lifetime value parameter. We learn that the relevance of the newsletter design and contents exceeds the relevance of targeting the long-lasting investors by far. This result supports the claim of [Smith \(2012\)](#) and [Tolani et al. \(2020\)](#) on the relevance of expected benefits driving the response to digital marketing efforts. Across contexts of respondents are more generalizable pattern seems to emerge.

Practitioners going beyond establishing and maintaining a long-lasting relationship with their customers can take advantage of the records on the newsletter software server describing the clients' behavior. By creating effective and useful content for investors, companies can turn potential customers into actual customers because each customer will be willing to share useful and practical information with others, and this digital marketing tool is very easy to share. Heterogeneity of situation in receiving the newsletter as digital direct marketing is higher because with the advent of smartphones along with laptops and personal computers, accessing in mailboxes is rather accessible for most of the people in comparison with many years ago. A complementary exogenous determinant was the influence of investor lifetime. Long investor lifetime is substantially significant even with unobserved endogenous factors.

As FinTech product and service developers in advanced economies often understand and find customers' journey in order to satisfy their own needs ([Buckley and Webster 2016](#)). This finding has important implications for developing digital marketing. The financial sector is increasingly from technology-driven firms in order to win customers and gain market share, which is referred to as "FinTech" in the online and offline media discourse ([Buckley and Webster 2016](#); [Dapp et al. 2014](#)). Given that there is empirical evidence of and direction of endogeneity in mail marketing ([Hruschka 2010](#)), it is important to consider one of the most important factors related to the role of communication marketing in social networks. DPM enable considering the challenges of multicollinearity, heterogeneity, and endogeneity. A much-debated question is whether marketing communication tools such as electronic word of mouth have more impact on investor behavior compared to digital marketing; thus, determining that the era of social network leadership has arrived. While this manuscript was being written, an incredible event happened: Users on the Reddit forum WallStreetBets decided to buy GameStop shares ([Chapman 2021](#); [Krantz 2021](#)). Initially, users thought the stock was undervalued, and then, they wanted to send a message to short-sellers. Acting in unison, users pushed up the share price sharply to

astronomical levels, leading to massive losses for short sellers. As can be seen, decision makers managed to change the fate of a company's stock. Future research could consider peer-to-peer communication tools as another auxiliary variable in the model. As can be seen, there is the ubiquity of fintech services. Clearly, behavioral attributes are significantly affected by technological attributes and digital behavior (Singh et al. 2020). Considering the advancement of FinTech vendors in improving and automating the delivery and use of financial services, it is recommended that in further research more sophisticated instances of digital marketing (going beyond a newsletter) are evaluated with respect to heterogeneity in adopting and usage among investors behavior. Empirical evidence clarifying the extent to which financial behavior and turnover in the stock exchange are triggered by digital marketing, partly through newsletter, establishes the practical relevance of the present study. Results suggest brokerage firms on the stock exchange to adopt various newsletter types and related contents to complement their modern digital marketing mix that persuades investors to buy and sell in the market, in other words, to increase their turnover.

A limitation of the present study is that some investor data were saved incompletely in the database of the brokerage server and newsletter software, because of advertising-blocking email clients and national data privacy regulations. Moreover, the parsimonious model used in this pioneering study might be enriched with additional exogenous determinants capturing observed heterogeneity, such as demographic factors, investor responsiveness, etc.

**Author Contributions:** Conceptualization, M.S. and R.W.; methodology, M.S. and R.W.; software, M.S.; validation, M.S.; data curation, M.S.; writing—original draft preparation, M.S. and R.W.; writing—review and editing, M.S. and R.W.; visualization, M.S.; supervision, R.W.; project administration, M.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Authors followed the APA Ethical Principles of Psychologists and Code of Conduct 2010 and 2016 Amendments.

**Informed Consent Statement:** Given that investors during opening account in brokerage firm have signed SLA, so informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Data cannot be available due to privacy reasons.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

### MCMC Parameters

Basically, MCMC samples are correlated but as the procedure goes on magnitude the correlation tends to zero. So, if we want to eliminate the correlation between the data, we need to burn some initial simulated values. After modeling and estimating, since the Bayesian model has to be estimated, it maximized the probability. With consideration of this probability, we run MCMC simulation with different burn-in and saving data, which took about four to ten hours on an Intel(R) Core(TM) i7-6500U processor and 8 GB of RAM as illustrated in Table A1.

**Table A1.** MCMC parameters.

	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6
nBurn	5.000	10.000	15.000	25.000	50.000	100.000
nSave	10.000	5.000	10.000	10.000	10.000	20.000
nSkip	10	15	15	15	15	15

### Appendix B

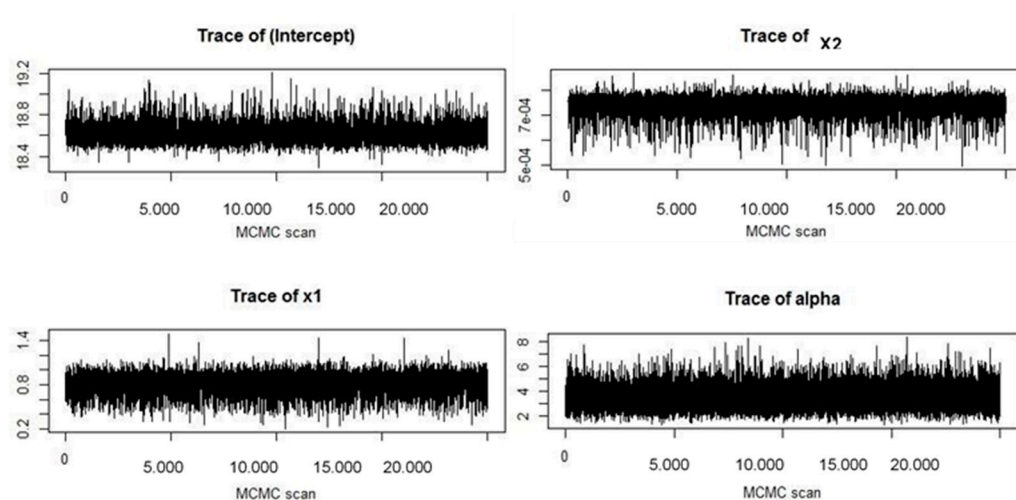
#### Posterior Predictive Distribution

**Table A2.** Posterior predictive distribution (log).

Minimum	−8.912
First Quarter	−2.467
Median	−2.054
Mean	−2.305
Third Quarter	−1.886
Maximum	−1.520

### Appendix C

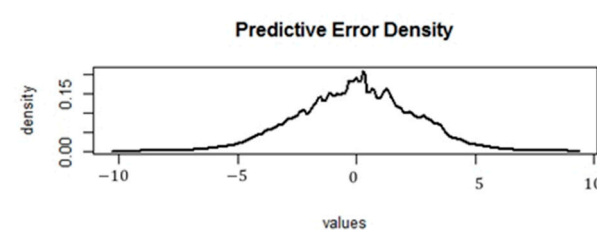
#### Histories and Histograms of Metropolis Steps



**Figure A1.** Histories and histograms of metropolis steps.

### Appendix D

#### Predictive Error Density



**Figure A2.** Predictive error density.

### References

- Abdallah, Salam, and Khalil Hilu. 2015. Exploring determinants to explain aspects of individual investors’ financial behavior. *Australasian Accounting, Business and Finance Journal* 9: 4–22. [CrossRef]
- Adrian, Palmer, and Koenig-Lewis Nicole. 2009. An experiential, social network-based approach to direct marketing. *Direct Marketing: An International Journal* 3: 162–76. [CrossRef]
- Ahmad, Maqsood. 2020. Does underconfidence matter in short-term and long-term investment decisions? Evidence from an emerging market. *Management Decision*. [CrossRef]
- Alan, Sule, Cemalcilar Mehmet, Karlan Dean, and Zinman Jonathan. 2018. Unshrouding: Evidence from bank overdrafts in Turkey. *Journal of Finance* 73: 481–522. [CrossRef]
- Asare-Frempong, Justice, and Manoj Jayabalan. 2017. Predicting customer response to bank direct telemarketing campaign. Paper presented at International Conference on Engineering Technology and Technopreneurship (ICE2T), Kuala Lumpur, Malaysia, September 18–20.



- Asllani, Arben, and Diane Halstead. 2015. A multi-objective optimization approach using the RFM model in direct marketing. *Academy of Marketing Studies Journal* 19: 65–79.
- Bank, Semra, Evrim Erdogan Yazar, and Ugur Sivri. 2019. Can social media marketing lead to abnormal portfolio returns? *European Research on Management and Business Economics* 25: 54–62. [CrossRef]
- Barman, Debaditya, Kamal Kumar Shaw, Anil Tudu, and Nirmalya Chowdhury. 2016. Classification of bank direct marketing data using subsets of training data. In *Information Systems Design and Intelligent Applications*. New Delhi: Springer, pp. 143–51.
- Benoit, Dries F., and Dirk Van den Poel. 2009. Benefits of quantile regression for the analysis of customer lifetime value in a contractual setting: An application in financial services. *Expert Systems with Applications* 36: 10475–84. [CrossRef]
- Buckley, Ross, and Sarah Webster. 2016. Fintech in Developing Countries: Charting New Customer Journeys. *Journal of Financial Transformation* 44: 1–19.
- Bult, Jan Roelf, and Tom Wansbeek. 1995. Optimal selection for direct mail. *Marketing Science* 14: 378–94. [CrossRef]
- Chakrabarty, Bidisha, Sang Bong Lee, and Nitish Singh. 2017. Doing good while making money: Individual investor participation in socially responsible corporations. *Management Decision* 55: 1645–59. [CrossRef]
- Chang, Chun-Wei, and Jonathan (Zaozao) Zhang. 2016. The Effects of Channel Experiences and Direct Marketing on Customer Retention in Multichannel Settings. *Journal of Interactive Marketing* 36: 77–90. [CrossRef]
- Chapman, Ben. 2021. Hedge Funds and Short-Selling: How and Why Gamestop Shook the Stock Market. Available online: <https://www.gqindia.com/get-smart/content/hedge-funds-gamestop-short-selling-stocks-everything-you-need-to-know-about-how-a-sub-reddit-shook-the-stock-markets> (accessed on 29 January 2021).
- Chintagunta, Pradeep, Tülin Erdem, Peter Eric Rossi, and Michel Wedel. 2006. Structural modeling in marketing: Review and assessment. *Marketing Science* 25: 604–16. [CrossRef]
- Cipolli, William, and Timothy Hanson. 2019. Supervised learning via smoothed Polya trees. *Advances in Data Analysis and Classification* 13: 877–904. [CrossRef]
- Dapp, Thomas, Lars Slomka, Deutsche Bank AG, and Ralf Hoffmann. 2014. Fintech—The digital (r)evolution in the financial sector. *Deutsche Bank Research* 11: 1–39.
- De Iorio, Maria, Peter Müller, Gary Rosner, and Steven MacEachern. 2004. An ANOVA model for dependent random measures. *Journal of the American Statistical Association* 99: 205–15. [CrossRef]
- Decker, Reinhold, and Ralf Wagner. 2000. Remarks on the behavioristic analysis of competitive reactions. *OR-Spectrum* 22: 97–116.
- Denison, David GT, and Bani Mallick. 2007. Analyzing financial data using Polya trees. In *Bayesian Statistics and Its Application*. New Delhi: Anamaya.
- Dey, Dipak, Peter Müller, and Debajyoti Sinha, eds. 2012. *Practical Nonparametric and Semiparametric Bayesian Statistics*. Berlin: Springer Science and Business Media, vol. 133.
- Dragotă, Victor, and Vlad Șerbănescu. 2010. Some Issues concerning romanian investors' behaviour. Results of a survey. *Theoretical and Applied Economics* 17: 5–16.
- Dunson, David, and Chuanhua Xing. 2009. Nonparametric Bayes modeling of multivariate categorical data. *Journal of the American Statistical Association* 104: 1042–51. [CrossRef] [PubMed]
- Dvořák, Tomáš. 2005. Do domestic investors have an information advantage? Evidence from Indonesia. *The Journal of Finance* 60: 817–39. [CrossRef]
- Ebbes, Peter, Michel Wedel, and Ulf Böckenholt. 2009. Frugal IV alternatives to identify the parameter for an endogenous regressor. *Journal of Applied Econometrics* 24: 446–68. [CrossRef]
- Elsalamony, Hany. 2014. Bank direct marketing analysis of data mining techniques. *International Journal of Computer Applications* 85: 12–22. [CrossRef]
- Feld, Sebastian, Heiko Frenzen, Manfred Krafft, Kay Peters, and Peter Verhoef. 2013. The effects of mailing design characteristics on direct mail campaign performance. *International Journal of Research in Marketing* 30: 143–59. [CrossRef]
- Ferrentino, Rosa, Maria Teresa Cuomo, and Carmine Boniello. 2016. On the customer lifetime value: A mathematical perspective. *Computational Management Science* 13: 521–39. [CrossRef]
- Goldstein, Itay, Chester Spatt, and Mao Ye. 2021. Big Data in Finance. Available online: <https://ssrn.com/abstract=3809447> (accessed on 22 March 2021).
- Gomber, Peter, Jascha-Alexander Koch, and Michael Siering. 2017. Digital Finance and FinTech: Current research and future research directions. *Journal of Business Economics* 87: 537–80. [CrossRef]
- Gönül, Füsün, Byung-Do Kim, and Mengze Shi. 2000. Mailing smarter to catalog customers. *Journal of Interactive Marketing* 14: 2–16. [CrossRef]
- Gopal, Ram D., Zhiping Walter, and Arvind Tripathi. 2001. Admediation: New horizons in effective email advertising. *Communications of the ACM* 44: 91–96. [CrossRef]
- Guelman, Leo, Montserrat Guillén, and Ana María Pérez-Marín. 2015. A decision support framework to implement optimal personalized marketing interventions. *Decision Support Systems* 72: 24–32. [CrossRef]
- Guindani, Michele, and Wesley Johnson. 2018. More nonparametric Bayesian inference in applications. *Statistical Methods & Applications* 27: 239–51.
- Hanson, Timothy, and Wesley Johnson. 2002. Modeling regression error with a mixture of Polya trees. *Journal of the American Statistical Association* 97: 1020–33. [CrossRef]

- Hruschka, Harald. 2010. Considering endogeneity for optimal catalog allocation in direct marketing. *European Journal of Operational Research* 206: 239–47. [CrossRef]
- Imrie Brian. 2013. The influence of social capital on service quality evaluation. *Management Decision* 51: 871–89. [CrossRef]
- Jalilvand, Abolhassan, Mojtaba Rostami Noroozabad, and Jeannette Switzer. 2018. Informed and uninformed investors in Iran: Evidence from the Tehran Stock Exchange. *Journal of Economics and Business* 95: 47–58. [CrossRef]
- Jara, Alejandro, Timothy Hanson, Fernando Quintana, Peter Müller, and Gary Rosner. 2011. DPpackage: Bayesian Non- and Semi-parametric Modelling in R Software. *Journal of Statistical Software* 40: 1–30. [CrossRef]
- Jedidi, Kamel, Harsharanjeet Sharan Jagpal, and Wayne DeSarbo. 1997. Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science* 16: 39–59. [CrossRef]
- Kleinman, Ken, and Joseph Ibrahim. 1998. A semi-parametric Bayesian approach to generalized linear mixed models. *Statistics in Medicine* 17: 2579–96. [CrossRef]
- Krantz, Matt. 2021. Nine Investors Instantly Make \$16 Billion On GameStop Stock, Squeeze. Available online: <https://www.investors.com/etfs-and-funds/sectors/gme-stock-gamestop-investors-instantly-make-16-billion-gamestop-stock-squeeze/> (accessed on 1 February 2021).
- Lavery, Matthew Ryan, Parul Acharya, Stephen Sivo, and Lihua Xu. 2019. Number of predictors and multicollinearity: What are their effects on error and bias in regression? *Communications in Statistics-Simulation and Computation* 48: 27–38. [CrossRef]
- Lavine, Michael. 1992. Some aspects of Polya tree distributions for statistical modelling. *Annals of Statistics* 20: 1222–35. [CrossRef]
- Lavine, Michael. 1994. More aspects of Polya tree distributions for statistical modelling. *Annals of Statistics* 22: 1161–76. [CrossRef]
- Lee, Jinkook. 2002. A key to marketing financial services: The right mix of products, services, channels and customers. *Journal of Services Marketing* 16: 238–58. [CrossRef]
- Lee, You-Kyung. 2021. Impacts of Digital Technostress and Digital Technology Self-Efficacy on Fintech Usage Intention of Chinese Gen Z Consumers. *Sustainability* 13: 5077. [CrossRef]
- Liao, Shu-hsien, Yin-ju Chen, and Hsin-hua Hsieh. 2004. Mining customer knowledge for electronic catalog marketing. *Expert Systems with Applications* 27: 521–32. [CrossRef]
- Lillo, Fabrizio, Salvatore Micciche, Michele Tumminello, Jyrki Piilo, and Rosario Mantegna. 2015. How news affects the trading behaviour of different categories of investors in a financial market AU-Lillo, Fabrizio. *Quantitative Finance* 15: 213–29. [CrossRef]
- Lin, Shen, Da Ren, Wei Zhang, Yongjie Zhang, and Dehua Shen. 2016. Network interdependency between social media and stock trading activities: Evidence from China. *Physica A: Statistical Mechanics and Its Applications* 451: 305–12. [CrossRef]
- Lin, Yu-Cheng, Chiung-Yao Huang, and Yu-Shan Wei. 2018. Perfectionist decision-making style and ethical investment willingness: A two-factor causal mediation model. *Management Decision* 56: 534–49. [CrossRef]
- Link, William, and Mitchell Eaton. 2012. On thinning of chains in MCMC. *Methods in Ecology and Evolution* 3: 112–15. [CrossRef]
- Liu, Yi, Wei Wang, and Zuopeng Justin Zhang. 2020. The dual drivetrain model of digital transformation: Role of industrial big-data-based affordance. *Management Decision*. [CrossRef]
- Ma, Shaohui, Lu Hou, Wensong Yao, and Baozhen Lee. 2016. A nonhomogeneous hidden Markov model of response dynamics and mailing optimization in direct marketing. *European Journal of Operational Research* 253: 514–23. [CrossRef]
- Maroufkhani, Parisa, Ralf Wagner, Wan Khairuzzaman Wan Ismail, Mas Bambang Baroto, and Mohammad Nourani. 2019. Big data analytics and firm performance: A systematic review. *Information* 10: 226. [CrossRef]
- Martí-Ballester, Carmen Pilar. 2015. Investor reactions to socially responsible investment. *Management Decision* 53: 571–604. [CrossRef]
- Martin, Roderick, and Tahir Nisar. 2007. Activist investment: Institutional investor monitoring of portfolio companies. *Management Decision* 45: 827–40. [CrossRef]
- Miguéis, Vera, Ana Camanho, and José Borges. 2017. Predicting direct marketing response in banking: Comparison of class imbalance methods. *Service Business* 11: 831–49. [CrossRef]
- Mitik, Merve, Ozan Korkmaz, Pinar Karagoz, Ismail Hakki Toroslu, and Ferhat Yucel. 2017. Data Mining Approach for Direct Marketing of Banking Products with Profit/Cost Analysis. *Review of Socionetwork Strategies* 11: 17–31. [CrossRef]
- Moyo, Busani. 2018. An analysis of competition, efficiency and soundness in the South African banking sector. *South African Journal of Economic and Management Sciences* 21: 1–14. [CrossRef]
- Müller, Peter, and Abel Rodriguez. 2013. *Pólya Trees Nonparametric Bayesian Inference*. Beachwood: Institute of Mathematical Statistics, Alexandria: American Statistical Association, pp. 43–51.
- Müller, Peter, and Fernando Quintana. 2004. Nonparametric Bayesian data analysis. *Statistical Science* 19: 95–110. [CrossRef]
- Müller, Peter, and Riten Mitra. 2013. Bayesian nonparametric inference—why and how. *Bayesian Analysis (Online)* 8: 269–302. [CrossRef] [PubMed]
- Müller, Peter, Fernando Andrés Quintana, Alejandro Jara, and Tim Hanson. 2015. *Bayesian Nonparametric Data Analysis*. Berlin: Springer.
- Müller, Peter, Fernando Quintana, and Gary Rosner. 2004. A method for combining inference across related nonparametric Bayesian models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 66: 735–49. [CrossRef]
- Neal, Radford. 2000. Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics* 9: 249–65.
- Nikunen, Tuulia, Martti Saarela, Eeva-Liisa Oikarinen, Matti Muhos, and Lari Isohella. 2017. Micro-Enterprise’s Digital Marketing Tools for Building Customer Relationships. *Management* 12: 2.

- Olbryś, Joanna, and Michał Mursztyn. 2018a. Assessing accuracy of trade side classification rules. Methods, data, and problems. In *The 12th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Conference Proceedings*. Cracow: Foundation of the Cracow University of Economics.
- Olbryś, Joanna, and Michał Mursztyn. 2018b. Evaluating Trade Side Classification Algorithms Using Intraday Data from the Warsaw Stock Exchange. *Archives of Data Science, Series A (Online First)* 4: 20.
- Olson, David, and Bongsug Kevin Chae. 2012. Direct marketing decision support through predictive customer response modeling. *Decision Support Systems* 54: 443–51. [\[CrossRef\]](#)
- Palacios-Marqués, Daniel, José Merigó, and Pedro Soto-Acosta. 2015. Online social networks as an enabler of innovation in organizations. *Management Decision* 53: 1906–20. [\[CrossRef\]](#)
- Palalic, Ramo, Veland Ramadani, Syedda Mariam Gilani, Shqipe Gërguri-Rashiti, and Leo-Paul Dana. 2020. Social media and consumer buying behavior decision: What entrepreneurs should know? *Management Decision*. [\[CrossRef\]](#)
- Parlar, Tuba, and Songül Kakilli Acaravcı. 2017. Using data mining techniques for detecting the important features of the bank direct marketing data. *International Journal of Economics and Financial Issues* 7: 692–96.
- Pauwels, Koen. 2014. *It's Not the Size of the Data—It's How You Use it: Smarter Marketing with Analytics and Dashboards*. New York: Amacom.
- Petersen, John Andrew, Tarun Kushwaha, and Vipin Kumar. 2015. Marketing communication strategies and consumer financial decision making: The role of national culture. *Journal of Marketing* 79: 44–63. [\[CrossRef\]](#)
- Ranjan, Jayanthi, and Vishal Bhatnagar. 2011. Role of knowledge management and analytical CRM in business: Data mining based framework. *Learning Organization* 18: 131–48. [\[CrossRef\]](#)
- Rizzo, Anthony. 2018. The Relationship of Direct Mail Data and Time Personalization and Response within the Financial Services Industry. Doctoral dissertation, Grand Canyon University, Phoenix, Arizona.
- Rzeszutek, Marcin, Adam Szyszka, and Monika Czerwonka. 2015. Investors' expertise, personality traits and susceptibility to behavioral biases in the decision making process. *Contemporary Economics* 9: 237–352. [\[CrossRef\]](#)
- Saukkonen, Natalia, Teemu Laine, and Petri Suomala. 2018. Utilizing management accounting information for decision-making: Limitations stemming from the process structure and the actors involved. *Qualitative Research in Accounting and Management* 15: 181–205. [\[CrossRef\]](#)
- Scholz, Sören, and Ralf Wagner. 2005. The quality of prior information structure in business planning—An experiment in environmental scanning. In *Operations Research Proceedings 2004*. Berlin and Heidelberg: Springer, pp. 238–45.
- Schröder, Nadine, and Harald Hruschka. 2012. Response measurement and optimization of direct mailings. *Journal für Betriebswirtschaft* 62: 261–308. [\[CrossRef\]](#)
- Schröder, Nadine, and Harald Hruschka. 2017. Comparing alternatives to account for unobserved heterogeneity in direct marketing models. *Decision Support Systems* 103: 24–33. [\[CrossRef\]](#)
- Schultz, Don E., and James Jimmy Peltier. 2013. Social media's slippery slope: Challenges, opportunities and future research directions. *Journal of Research in Interactive Marketing* 7: 86–99. [\[CrossRef\]](#)
- Shah, Denish, and BPS. Murthi. 2020. Marketing in a data-driven digital world: Implications for the role and scope of marketing. *Journal of Business Research* 125: 772–79. [\[CrossRef\]](#)
- Sharma, Niharika, Arvinder Kaur, Sheetal Gandotra, and Bhawna Sharma. 2015. Evaluation and comparison of data mining techniques over bank direct marketing. *International Journal of Innovative Research in Science, Engineering and Technology* 4: 7141–47.
- Singh, Shubhangi, Marshal Sahni, and Raj Kovid. 2020. What drives FinTech adoption? A multi-method evaluation using an adapted technology acceptance model. *Management Decision* 58: 1675–97. [\[CrossRef\]](#)
- Smith, Katherine Taken. 2012. Longitudinal study of digital marketing strategies targeting Millennials. *Journal of Consumer Marketing* 29: 86–92. [\[CrossRef\]](#)
- Taiminen, Heini Maarit, and Heikki Karjaluo. 2015. The usage of digital marketing channels in SMEs. *Journal of Small Business and Enterprise Development* 22: 633–51. [\[CrossRef\]](#)
- Talay, M. Berk, Billur Akdeniz, Michael Obal, and Janell Townsend. 2019. Stock Market Reactions to New Product Launches in International Markets: The Moderating Role of Culture. *Journal of International Marketing* 27: 81–98. [\[CrossRef\]](#)
- Tang, Heng, Stephen Shaoyi Liao, and Sherry Xiaoyun Sun. 2013. A prediction framework based on contextual data to support mobile personalized marketing. *Decision Support Systems* 56: 234–46. [\[CrossRef\]](#)
- Thwaites, Des, and Sharon CI Lee. 1994. Direct marketing in the financial services industry. *Journal of Marketing Management* 10: 377–90. [\[CrossRef\]](#)
- Tiago, Maria Teresa Pinheiro Melo Borges, and José Manuel Cristóvão Veríssimo. 2014. Digital marketing and social media: Why bother? *Business Horizons* 57: 703–8. [\[CrossRef\]](#)
- Tikkanen, Henriikki, Joel Hietanen, Tuomas Henttonen, and Joonas Rokka. 2009. Exploring virtual worlds: Success factors in virtual world marketing. *Management Decision* 47: 1357–81. [\[CrossRef\]](#)
- Tolani, Kanchan, Ruchi Sao, Pritam Bhadade, and Shravan Chandak. 2020. Money and generations: Financial choices made by Gen X and Gen Y. *International Journal of Management* 11: 657–72.
- Trivedi, Naresh Kumar, Suresh Kumar, Sachin Jain, and Shikha Maheshwari. 2020. KFCM-Based Direct Marketing. In *Rising Threats in Expert Applications and Solutions*. Berlin and Heidelberg: Springer, pp. 495–502.
- Vallejos, Alejandro Antonio Jara. 2008. Bayesian Semiparametric Methods for the Analysis of Complex Data. Doctoral dissertation, Tese (Doutorado em Ciências), Faculdade de Ciências, Universidade Católica de Leuven, Leuven, Belgium.

- Venkatesan, Rajkumar, V. Kumar, and Timothy Bohling. 2007. Optimal customer relationship management using Bayesian decision theory: An application for customer selection. *Journal of Marketing Research* 44: 579–94. [[CrossRef](#)]
- Wansbeek, Tom, and Michel Wedel. 1999. Marketing and econometrics. *Journal of Econometrics* 89: 1–2. [[CrossRef](#)]
- Wienke, Andreas. 2010. *Frailty Models in Survival Analysis*. Amsterdam and Boca Raton: Chapman & Hall/CRC.
- Wu, Jiexing, Kate Jiayi Li, and Jun Liu. 2018. Bayesian Inference for Assessing Effects of Email Marketing Campaigns. *Journal of Business and Economic Statistics* 36: 253–66. [[CrossRef](#)]
- Yamin, Ahmad Bin. 2017. Impact of Digital Marketing as a Tool of Marketing Communication: A Behavioral Perspective on Consumers of Bangladesh. *American Journal of Trade and Policy* 4: 117–22. [[CrossRef](#)]
- Yelkur, Rama, and Maria Manuela Nêveda DaCosta. 2001. Differential pricing and segmentation on the Internet: The case of hotels. *Management Decision* 39: 252–62. [[CrossRef](#)]
- Zantedeschi, Daniel, Eleanor McDonnell Feit, and Eric Bradlow. 2016. Measuring multichannel advertising response. *Management Science* 63: 2706–28. [[CrossRef](#)]