Measuring the Influence of Multichannel Digital Advertising: A Fractional Calculus-Based Approach

Abstract

While the digital advertising markets have been growing rapidly, the effect of advertising is yet to be properly understood, mainly due to long purchase cycles and unobserved stages in consumers' decision-making process. To address this shortcoming, we first propose a new advertising stock model that accounts for consumers' independent decision-making as well as the word of mouth and network effects. Fractional calculus is used to account for advertising stock erosion due to consumers' memory decay. Based on the proposed advertising stock model, we then develop a generalized proportional hazard model that incorporates different channels' advertising stocks as covariates to measure the effect of these channels on adoptions. The proposed model is tested using a dataset that records individual-level adoptions of a new video game and advertising impressions from different digital channels. The results show that consumers' advertising stocks exhibit a nonmonotonic pattern, contrary to the geometric decay assumed in the literature. Furthermore, we compare the proposed model with two benchmark models and show that our model outperforms them in determining the influence of different channels on adoptions. We also find that consumers' decisions are heavily affected by word of mouth and network effects, which have not been accounted for in prior studies, and that these effects are accentuated post product release. Additionally, our empirical results demonstrate that firms are better off advertising through a few select channels instead of utilizing all available channels. Our findings bear important practical implications and can help advertisers efficiently and effectively plan their advertising campaigns.

Keywords: Advertising stock, advertising channel, digital attribution, fractional calculus, hazard model

Measuring the Influence of Multichannel Digital Advertising: A Fractional Calculus-Based Approach

1. Introduction

Digital advertising is becoming the dominant form of advertising (Todri 2022), with total spending expected to reach $298.4 billion in 2024 (Statista 2023). As a result of such enormous online advertising (or *ad* for short) expenditure, consumers experience frequent exposures to advertisements from various digital channels. Despite the rapid growth of digital advertising market, Gordon, Jerath, Katona, Narayanan, Shin, and Wilbur (2021) report in a comprehensive review article that *ad effect measurement* is yet to be properly addressed. According to them, challenges in advertising effect measurement can be attributed to different reasons, including long purchase cycles and unobserved stages in consumers' decision-making process. Specifically, following the hierarchy of effects model, consumers first gain product awareness based on information received from an advertiser, then potentially develop interest due to actively acquiring and processing relevant information, and finally may act in the form of product purchase due to the developed interest (Barry 1987). While consumers' purchases can be recorded, their information accumulation and interest development are typically not directly observable to the advertiser. Additionally, the common practice of starting advertising *before* product release further complicates the measurement of advertising effect. Particularly, it is challenging to gauge how advertising campaigns influence consumers' purchase decisions before a product is launched. Furthermore, firms often advertise through multiple channels, making it challenging to determine the effect of each channel. We endeavor to address these challenges in this study by developing new advertising stock and digital attribution models, also heeding the recent call by Gordon et al. (2021) to provide an efficient and effective solution to the ad effect measurement problem.

We first introduce a new advertising stock model based on the premise that the contribution of an ad exposure to a consumer's advertising stock is determined by the product information the consumer receives and the subsequent interest in the product the consumer develops over time. Under the new model, depending on the rate of information accumulation and interest development, the advertising stock and subsequently the effect of advertisement either monotonically declines or first increases and then decreases over time, conforming to the two different patterns reported in the literature and enhancing the model's generalizability.

The new model accounts for consumers' independent decision-making as well as the word of mouth and network effects. This renders the model more realistic than existing models that do not incorporate these influences, specifically considering that electronic word of mouth plays a primal role in forming consumers' attitudes and behaviors (Todri, Adamopoulos, and Andrews 2022).

Furthermore, considering that memory plays a pivotal role in the temporal dynamics of consumers' advertising stock, we incorporate consumers' memory decay in the new advertising model as well. In particular, accounting for memory decay is of fundamental importance in cases where firms start their advertising campaigns before releasing their products to the market, expecting to reap the benefit of advertising after product release. In our advertising stock model, we use *fractional calculus* to model the erosion of advertising stock due to consumers' memory decay over time. Fractional calculus, a novel mathematical tool that has been applied in multiple disciplines including management science (Lotfi, Jiang, Naoum-Sawaya, and Begen 2024, Lotfi, Naoum-Sawaya, Lotfi, and Jiang 2024, Lotfi, Jiang, Lotfi, and Jain 2023), extends the classical integration and differentiation operators, making non-integer derivatives and integrals possible. The model assumes a power-law memory decay trend, which closely resembles the Woodworth's classic forgetting curve (Bennedsen and Caspersen 2009).

The prior literature has used experiments or observational data to analyze the effect of advertising, with the latter more widely adopted (Gordon et al. 2021). In this research, we test our model using a video game dataset that records individual-level ad exposures from different channels before and after product release and product adoptions (purchases) after release. This dataset provides a unique opportunity to study the advertising effect because it is quite rare that a single-source dataset contains both advertising exposures and the subsequent conversions (Gordon et al. 2021).

To measure the influence of exposures from different channels on product adoptions, we incorporate changes in advertising stocks resulting from such exposures, as computed by the proposed advertising stock model, into a proportional hazard model. The proposed generalized hazard model can utilize data on individual-level advertising exposures and adoptions, and is also applicable in cases where advertising exposures start to occur before product release.

We compare the proposed generalized hazard model with two benchmark hazard models. In the first benchmark model, the measurement of consumers' stocks of advertisements is conducted using the widely-used Koyck distributed lag model (1954) based on which the effect of an ad exposure decays exponentially over time. It is important to note that, even though the Koyck model is proposed decades ago, this intuitive and parsimonious model is frequently adopted by more recent studies (e.g., Terui, Ban, and Allenby 2011, Zantedeschi, Feit, and Bradlow 2016, and Chae, Bruno, and Feinberg 2019). We develop a second benchmark model in which the Erlang-2 distribution is used to represent each exposure's advertising stock trend (Chatfield and Goodhardt 1973). This model assumes that the effect of each ad exposure peaks, following an Erlang-2 distribution.

Our empirical results demonstrate that the proposed model more accurately fits the training data and delivers a better out-of-sample prediction performance than do the two benchmark models. Additionally, we show that a consumer's advertising stock can exhibit a nonmonotonic trend that first increases and then declines, and that the nonmonotonic pattern emerges due to the word of mouth and network effects.

Moreover, to further adjust the model to cases in which firms start advertising before product release, we develop a model extension which distinguishes between the word of mouth and network effects before and after product release. Our empirical findings based on the extended model demonstrate that the word of mouth and network effects become significantly accentuated after product release, resulting in a much quicker rise in the adstock trends of consumers who embark on their purchase journey post product release.

Finally, our results indicate that firms are better off selecting a subset of the available advertising channels that have the highest impact on consumers' adoption decisions, because viewing advertisements through too many channels may negatively impact consumers' adoption propensity.

This study also offers important practical implications. First, the model we propose explains and predicts the temporal dynamics of advertising effect, which can help firms decide the timing of their advertising campaigns. In particular, since the proposed model can capture the effect of advertising exposures from before product release, it can help firms plan for pre-launch advertising campaigns. Second, firms can implement the proposed modeling framework to distinguish between the word of mouth and network effects before and after product release, allowing them to strategically plan to leverage these effects. Specifically, considering that pre-release word of mouth is expected to be greatly influenced by influencers, firms can plan to efficiently invest in *influencer-based word of mouth* before product release. Third, the proposed model can tease out the effect of different advertising channels on consumer purchases, which can help firms optimize their selection of advertising channels.

The remainder of this paper is organized as follows. We provide a review of the relevant literature in Section 2. We present a review of fractional calculus, including its memory interpretation, in Section 3. Section 4 develops a new advertising stock model and a proportional hazard model for time-to-adoption analysis. Section 5 evaluates the performance of the proposed model against two benchmark models. We introduce and empirically test a few extensions of the new advertising stock model in Section 6. Finally, we provide our concluding remarks in Section 7.

2. Relevant Literature Review

In this section, we briefly review three related bodies of literature on *advertising stock* (or adstock in short), *digital attribution*, and *diffusion of innovations*.

2.1. Advertising Stock

The present research is closely related to the adstock literature that studies how the effect of advertising accumulates and then attenuates over time. Empirical works on the initial and long-term effects of advertising date at least as far back as Koyck (1954). As per Koyck, a consumer accumulates latent adstock from ad exposures over time. Based on a distributed lag structure, Koyck's model captures a process in which the influence of advertising on sales can be characterized by a geometrically decaying process. An exponential decay function is adopted to account for the declining effect of prior ad exposures. Although it is proposed decades ago, the intuitive and parsimonious Koyck model is still frequently adopted by more recent studies to capture the exponentially decaying effect of advertising (e.g., Terui et al. 2011, Zantedeschi et al. 2016, and Chae et al. 2019).

Nerlove and Arrow (1962) defined a stock of goodwill that summarized the effect of past and current advertising spending on consumer demand. Similar to Koyck (1954), they assumed that goodwill depreciates over time. Applying Koyck's distributed lag technique on a dataset collected from the Lydia E. Pinkman Medicine Company over a period of more than 50 years, Palda (1965) demonstrated that incorporating both current and past advertising expenditures into advertising models can produce significant explanatory power. Palda reported that advertising has a pronounced lagged influence on sales. On average, it took almost seven years for the advertising spending to achieve 95% of its sales-generating effect.

Following the standard distributed lag approach by Koyck (1954), Zantedeschi et al. (2016), in an experimental setting with an email channel and a catalogue channel, studied multichannel advertising responses and predicted the short- and long-term influences of different ad channels on consumer purchases. The prior study showed that, by using its method and targeting the most responsive consumers, the firm can increase its predicted advertising returns by almost 70%, compared to targeting based on the well-known recency, frequency, and monetary value measures.

Using a Koyck specification to model adstock, Chae et al. (2019) studied consumers' responses to online advertising repetition. Specifically, they documented consumers' *weariness*, which refers to a situation that there is a point beyond which more ad exposures can have a negative marginal influence. Shapiro, Hitsch, and Tuchman (2021) studied television advertising effectiveness. Specifically, they estimated advertising stock elasticities, which signify the total current and future shifts in sales due to a one-percent increase in advertising at the present time. Their findings demonstrated significantly smaller advertising elasticities compared to the results from the literature and a considerable percentage of negative or statistically insignificant estimates.

Relaxing the assumption that the advertising effect only declines over time, Bass and Clarke (1972) proposed a generalization of the Koyck distributed-lag model that allows responses to advertisements to first intensify and then decline, demonstrating a nonmonotonic behavior. However, Bass and Clarke's study focused on cases in which sales and advertising take place in parallel, making it inapplicable for the cases we study in which advertising exposures start to occur before product release. Furthermore, Bass and Clarke used aggregate-level data on monthly sales and advertising history of a dietary weight-control non-durable product. By contrast, the model we develop can be fit to individual-level consumer data and be used to study the influence of advertising on the adoption of durable products.

2.2. Digital Attribution

The second stream of related research is on digital attribution. The aim of digital attribution research is to distribute the credit for consumer purchases across the digital marketing channels that deliver advertisements to consumers. The digital attribution literature has empirically estimated the effectiveness of advertising based on consumer heterogeneity in terms of exposure to advertising (e.g., Ghose and Todri-Adamopoulos 2016). For example, Shao and Li (2011) proposed a bagged logistic regression model for advertising channel attribution, and cross-validated their model against a simple probabilistic model that directly quantifies the attribution to different channels. The results showed that the bagged logistic regression model and the probabilistic model deliver highly consistent performance. Li and Kannan (2014) developed a new method to determine the value of each online advertising channel using individual-level touch data, and found that the channels' relative effects are significantly different from those reported using metrics presented in previous studies.

Worth particular mentioning are two studies in the information systems literature. Xu, Duan, and Whinston (2014) study the influence of different online advertisement types on conversions by analyzing the interactions among advertisement clicks. The study is motivated by the observation that, while some advertisement clicks may not lead to immediate conversions, they may trigger subsequent clicks on other advertisements that eventually result in conversions. The results demonstrated that a commonly used conversion rate measure overestimates the effect on convergence of search advertisements and underestimates that of display advertisements. Ghose and Todri-Adamopoulos (2016) studied display advertising's effectiveness across various individual behaviors. They empirically showed that exposures to display advertising increase user propensity to engage in both passive search (using received information) and active search (applying effort to collect information). Furthermore, they found that the longer a consumer spends viewing an ad, the more likely she is to engage in direct search behavior (e.g., visiting a website) rather than indirect search behavior (e.g., using search engine inquires).

2.3. Diffusion of Innovations

This study also draws upon the literature on the diffusion of innovations. This stream of literature is best represented by the seminal Bass model (Bass 1969). In one of the directions of extensions of the Bass model, the influence of marketing-mix variables, including advertising, on product purchases has also been studied. Examples of such research include the Generalized Bass Model (GBM) (Bass, Krishnan, and Jain 1994) and the proportional hazard model (PHM) (e.g., Bass, Jain, and Krishnan 2000). Note that the diffusion models proposed in this stream of research are primarily developed to estimate aggregate-level consumer adoptions and are not applicable for cases in which ad campaigns start before product release. However, since the spread of product information during ad campaigns is essentially a diffusion process, this literature provides the theoretical and modeling foundation for the model proposed in this research.

An extensive review of the prior literature has not led to any model that (a) can accommodate the nonmonotonic (first intensifying and then declining) effect of advertising, (b) incorporates the influence of advertising exposures from different channels, and (c) can be used when ad campaigns start before product release. In this study, we aim to develop a model that offers such capabilities.

3. Fractional Calculus and Advertising Memory

In this section, we introduce fractional calculus, which is key to the development of our new adstock model, and discuss its ability in capturing a consumer's product-related memory based on which her adstock changes. Memory is a key part of adstock models as such models are expected to measure the initial and long-term effects of advertising. In fact, it is through memory that advertising demonstrates a carryover effect. Therefore, we carefully incorporate memory, as well as the corresponding depreciation, into our new adstock model.

Prior research suggests that a consumer's advertising stock can be viewed as a *process with memory* (Boltzmann 1876), or more specifically an *economic process with memory* (Baillie 1996, Teyssière and Kirman 2006). According to Tarasov (2018), in an economic process with memory, memory captures the dependence of an output (response variable) at the present time on the history of the changes of an input (impact variable) in a given time frame. Accordingly, memory of advertisement captures the dependence of adstock at the current time on the advertising exposures received during a past time window.

Regarding methods used to study economic processes with memory, Tarasov (2018) points out "*it is known that derivatives of positive integer orders are determined by the properties of the differentiable function only in an infinitesimal neighborhood of the considered point. As a result, differential equations with integer-order derivatives cannot describe processes with memory*." Tarasov then suggests a powerful tool that can be used to describe economic processes with memory. The tool is called *fractional calculus*, which represents a branch of mathematics that generalizes differentiation and integration so that non-integer-order differential and integral operators become possible. Interested readers can refer to Samko, Kilbas, and Marichev (1993), Podlubny (1999), Kilbas, Srivastave, and Trujillo (2006), and Baleanu Diethelm, Scalas, and Trujillo (2012) for comprehensive reviews of the fractional calculus literature.

Drawing on the prior literature (Samko et al. 1993, Kilbas et al. 2006), Tarasov (2018) suggests that an economic process with power-law type memory can be captured by the following fractional integral equation

$$I^{\beta}X(t) := \int_0^t \frac{1}{\Gamma(\beta)}(t-\tau)^{\beta-1}X(\tau)d\tau\,. \tag{1}$$

In this equation, $X(t)$ represents the *input* of the economic process being modeled. The LHS is the *output*, where I^{β} is a non-integer fractional integral of order β ($\beta > 0$). The RHS of this equation is obtained based on the Reimann-Liouville fractional integral (Kilbas et al. 2006), which is a generalization of the standard n-th integral. The term $\frac{1}{\Gamma(\beta)}(t-\tau)^{\beta-1}$ is called the *kernel* of the fractional integral, which is interpreted as a *memory function* that is capable of capturing how the output at the present time depends on the history of the input up to the present time. The Gamma function $\Gamma(x)$, $x > 0$, takes the following form

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt. \tag{2}$$

The memory (kernel) function in Eq. (1) is a power-law type function (Tarasov 2018), which assigns lower weights to earlier events and higher weights to more recent events, as illustrated in Figure 1. The power-law memory shown in Figure 1 closely resembles human memory, making it suitable for capturing memory of advertising. For $0 < \beta < 1$, the higher the β, the better the memory retention.

We develop a new adstock model using the fractional integral presented in Eq. (1). Considering that a consumer's adstock is expected to change as the consumer acquires relevant product information and develops interest in the product, we apply the fractional integral operator on a process representing the consumer's information accumulation and interest development process. Specifically, in our adstock model, $X(\tau)$ in Eq. (1) represents the consumer's information accumulation and interest development process.

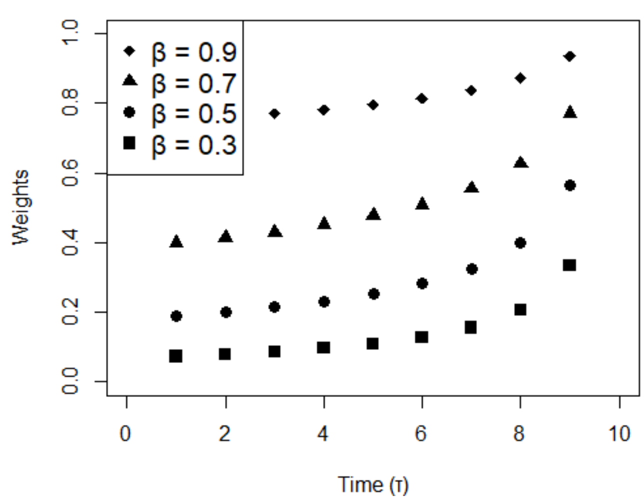

Figure 1. Declining-Toward-Past Memory Function (*t*=10)

4. Model Formulation

In this section, we first propose a new adstock model using the fractional calculus-based operator introduced in the previous section, and then develop a proportional hazard model to capture consumers' timing to product adoptions.

4.1. Fractional Calculus-Based Adstock Model

We propose a model to measure a consumer's adstock in the presence of multiple digital advertising channels. The model incorporates two fundamental components that affect a consumer's adstock: *adstock buildup* and *adstock decay*. The adstock buildup mechanism assumes that a consumer's adstock changes as the consumer acquires information related to a product and develops potential interest in the product. The adstock decay mechanism assumes that a consumer's adstock declines over time due to memory decay.

It has been shown that the flow of related information from different sources influences a consumer's adoption decision (Sawhney and Eliashberg 1996). Specifically, under the hierarchy of effects model of advertising, consumers first gain awareness based on information received from an advertiser, then develop interest due to actively acquiring and processing relevant information, and finally may act in the form of product purchases due to the developed interest (Barry 1987). Therefore, we expect that a consumer's adstock builds up as the consumer acquires relevant information and develops potential interest in the product.

In the meantime, a consumer also forgets product-related information over time. As such information is being forgotten, the developed interest is also expected to die down. The amount of information consumer i maintains at time t depends on the consumer's memory of the information acquired in $\tau \in [t_i, t)$, where t_i is the time at which consumer i begins acquiring relevant information. At t, a consumer with *perfect memory* remembers all the information acquired in $\tau \in [t_i, t)$. The information that a perfect-memory consumer maintains at t equals the cumulative information received until and at t. Mathematically, this leads to the information held by a perfect-memory consumer at t being the first-order integration of information acquisition rate in $[t_i, t]$. Thus, we have

$$Y(t) = \int_{t_i}^{t} y(\tau - t_i)\, d\tau = I^{\beta=1} y(t - t_i), \tag{3}$$

where, for consumer i, $Y(t)$ is the total information held at t and $y(\cdot)$ represents the new information acquired at any point in time in $[t_i, t]$. We name the degree of integral β in Eq. (3) the *memory coefficient*. For a perfect-memory consumer, $\beta = 1$.

While a perfect-memory consumer represents an extreme type, we also consider another extreme type of consumers who only remember the information received at the current time. We name these consumers *amnestic* consumers. For an amnestic consumer, we have

$$Y(t) = I^{\beta=0} y(t - t_i) = y(t - t_i). \tag{4}$$

Most consumers are expected to have a memory capacity that falls between those of amnestic consumers and perfect-memory consumers. Specifically, most consumers can remember a *fraction* of the information they have received in the past. Mathematically, we implement a *fractional* integral of order β of $y(\cdot)$ to represent such *common* consumers

$$Y(t) = I^{\beta} y(t - t_i),\ \ \beta \in (0, 1). \tag{5}$$

For such common consumers, the exact amount of information acquired in $\tau \in [0, t)$ that is still remembered at the current time depends on the formulation we use to operationalize the fractional integral in Eq. (5). In this study, we adopt the Riemann-Liouville integral shown in Eq. (1). Replacing $X(t)$ in Eq. (1) with $y(\tau - t_i)$, we formulate $Y(t)$ as

$$Y(t) = I^{\beta} y(t - t_i) = \int_{t_i}^{t} \frac{1}{\Gamma(\beta)} (t - \tau)^{\beta - 1} y(\tau - t_i) d\tau\ .$$

With this formulation, the common consumers modeled by Eq. (5) have a power-law memory (shown in Figure 1).

Because a consumer's adstock is expected to change in accordance with the consumer's information accumulation and interest development, we model the contribution of ad exposure j from channel k to consumer i's adstock at time t as

$$Adstock_{ijk}(t) = I^{\beta} y(t - t_i), \tag{6}$$

where t_i is the time at which the accumulation of relevant information for consumer i begins.

Note that it is often difficult to determine the exact time at which a consumer begins to collect relevant information (i.e., t_i). Ghose and Todri-Adamopoulos (2016) empirically show that consumers' exposures to display advertising increase their tendency to search for the advertised product and its brand. Furthermore, consumers often engage in both *passive search*, which involves responding to exogenously arriving information, and *active search*, in which the consumer exerts effort to acquire information. In our model operationalization, we assume that a consumer's information accumulation starts when she experiences the first advertising exposure.[1]

Note that the adstock given in Eq. (6) captures the effect of one ad exposure. We next incorporate the buildup effect of multiple ad exposures into the consumer's adstock. Specifically, we use an adstock variable, $z_{ik}(t)$, to represent the aggregate adstock accumulated for consumer *i* at *t* for all exposures from channel *k*

$$z_{ik}(t) = \sum_{j=1}^{n_{i_k}(t)} Adstock_{ijk}(t), \tag{7}$$

in which $n_{i_k}(t)$ is the number of exposures from channel *k* to consumer *i* up to time *t*. Then, a consumer's total adstock at *t*, $AS_i(t)$, is the sum of adstocks from all channels

$$AS_i(t) = \sum_k z_{ik}(t). \tag{8}$$

We next elaborate on the functional form that determines the rate at which a consumer acquires product information. Given that the spread of information is often expected to follow a diffusion process, we adopt the density function of the most well-known diffusion model—the Bass model (Bass 1969), in the following form

$$y(t - t_i) \ = \big(p + qF(t - t_i)\big)(1 - F(t - t_i)), \tag{9}$$

where $F(\cdot)$ represents the cumulative distribution of product information accumulation, and p and q are called the *coefficient of innovation* and *coefficient of imitation*, respectively, in the Bass model. We refer

[1] Our empirical tests demonstrate that relaxing this assumption and allowing the information accumulation's start time to happen either before or after the first exposure does not lead to better empirical performance.

to the adstock model formulated in Eqs. (6), (7), and (8) as the *Fractional Calculus-Based Advertising Stock Model* or *FCASM*.

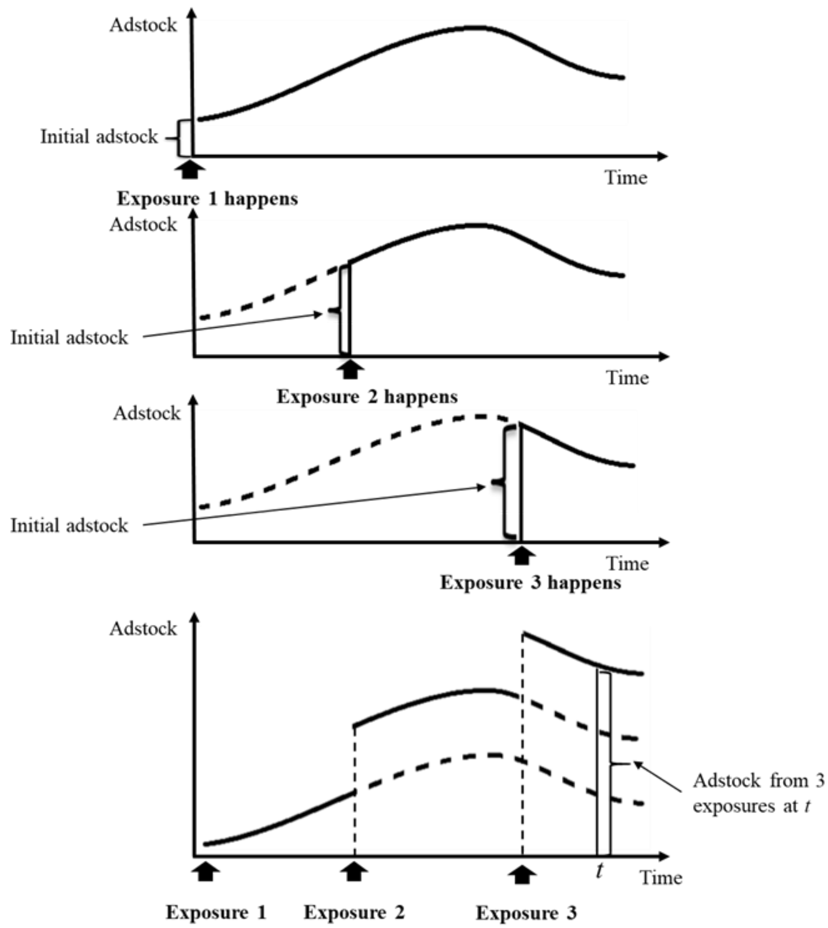

Figure 2. Changes in Adstocks Over Time under a Bell-Shaped Underlying Trend

Using the Bass model to represent the underlying information accumulation and interest development function enables the FCASM to accommodate the monotonic and nonmonotonic advertising trends reported in the literature, thus rendering the FCASM a generalization of existing models. Specifically, when $p < q$, $y(t - t_i)$ demonstrates a bell-shaped trend, resulting in an adstock growth process that can follow the same pattern. This makes the FCASM compatible with the findings of Bass and Clarke (1972), according to which the effect of advertisement can be nonmonotonic. For $p < q$, Figure 2 illustrates how the individual adstocks generated by three exposures that take place at different times and the total adstock (summation of the three individual adstocks) change with time. The individual and the total effects of the three exposures over time are shown by the solid parts of the curves in the figure.

When $p > q$, the Bass density function declines monotonically, in which case the FCASM is compatible with the classic Koyck distributed lag model (Koyck 1954), based on which the effect of an exposure declines exponentially over time. Figure 3 demonstrates the adstock trend corresponding to a monotonically declining underlying information accumulation and interest development trend ($p > q$). Compared to the curves shown in Figure 2, we can see that the adstocks corresponding to the three exposures, as well as their summation, decay at a much faster rate in Figure 3.

The incorporation of the Bass model into the FCASM not only allows it to assume the two main adstock trend shapes discussed in the literature (i.e., monotonic and nonmonotonic), but also leads to further insights. As mentioned earlier, the Bass model in the FCASM represents the process through which consumers acquire product information and develop potential interest in the product, which may subsequently result in product adoptions (purchases). In the Bass model, independent decision making is captured by the parameter *p* and imitation is represented by the parameter *q*. For the FCASM, we interpret the parameter *p* as representing the independently developed interest in the product and the parameter *q* as capturing the interest developed based on imitation or word of mouth effect, as well as the network effect.

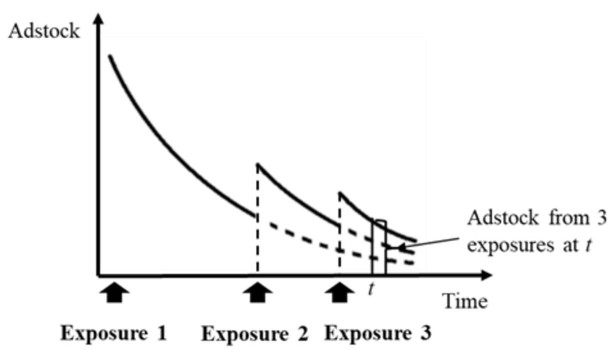

Figure 3. Changes in Adstocks Over Time under A Monotonically Decreasing Underlying Trend

The data we use to evaluate the proposed model is collected for a video game product. The likely existence of a strong network effect for such products suggests *q* being larger than *p*, resulting in a nonmonotonic underlying adstock growth trend instead of the monotonic trend suggested by the Koyck distributed lag model.

4.2. Time to Adoption: Hazard Modeling Framework

One of the primary goals of this research is to estimate the influence of digital advertising on consumers' product purchase (or adoption) decisions. Among various modeling approaches, proportional hazard models (a type of survival models) have been applied to estimate the timing of product purchases. For example, Jain and Vilcassim (1991) introduce a conditional hazard function to study household interpurchase timing while accounting for marketing variables, observed characteristics of households, and unobserved household heterogeneity. However, Jain and Vilcassim study a non-durable product. Qu, Lotfi, Jain, and Jiang (2022) use survival analysis to predict upgrade timing for a product line with successive product generations. However, the effect of advertising on consumers' decision-making process is not considered in Qu et al.'s study.

In the present research, we adopt a hazard modeling framework to disentangle the influences of different digital advertising channels on consumers' timing of adoptions. Consumer's adstocks, estimated based on the FCASM, are incorporated into a proportional hazard model as covariates (i.e., independent variables), resulting in a generalized proportional hazard model.

A hazard model is typically composed of a baseline hazard function and a term that captures the effect of covariates. In this study, we denote the baseline hazard function by $h_0(t)$, and the vector of time-varying covariates for a consumer *i* by $C_i(t)$, which includes the adstocks estimated using the FCASM. The resulting hazard rate function of adoption for consumer *i* is expressed as

$$h_i(t, C_i(t)) = h_0(t)e^{\boldsymbol{C}_i'(t)\boldsymbol{\theta}}, \tag{10}$$

in which $\boldsymbol{\theta}$ is the vector of coefficients that captures the impact of covariates on consumer *i*'s adoption hazard. Eq. (10) can be expanded as

$$h_i(t, C_i(t)) = h_0(t)e^{\sum_k \theta_k z_{ik}(t) + \sum_j \theta_j x_{ij}(t)}. \tag{11}$$

In Eq. (11), $z_{ik}(t)$ represents the adstock for consumer *i* at time *t* generated by exposures from channel *k*, and θ_k is the coefficient that captures the effect of $z_{ik}(t)$ on the adoption hazard. The model can include other covariates in addition to the adstock measures. Specifically, x_{ij} is the jth non-advertising covariate for consumer i and θ_j is the coefficient that captures the effect of this covariate on the adoption hazard.

The survival function for consumer i after incorporating the vector of covariates is expressed as

$$S_i(t, C_i(t)) = [S_0(t)]^{e^{C_i'(t)\theta}}. \tag{12}$$

We will elaborate on the baseline hazard in Eq. (11), i.e., $h_0(t)$, and our model estimation procedure in the next section.

5. Empirical Analysis

We empirically evaluate the FCASM's performance against two benchmark models using a dataset that records individual-level advertising exposures and product adoptions.

5.1. Data

We use a large-scale individual-level dataset that records consumers' advertising exposures in the form of impressions from six different digital advertising channels, other non-advertising variables, and the consumers' adoptions of the video game featured in the advertisements. This dataset provides a special opportunity to study the effect of advertising on consumer adoptions, because it is rare that companies are able to track both advertising exposures and the corresponding conversions in a single-source consumer data (Gordon et al. 2021). As recorded in the dataset, an impression happens when a consumer views an ad. The dataset includes 89,943 exposure records for 2,233 different consumers. Figure 4 presents a schematic illustration of ad exposures and adoptions recorded in the data. As portrayed in Figure 4, consumers started to experience exposures to advertising through different digital channels before the product was released. After product release, exposures continue to occur, followed by adoptions. The recorded exposures occurred as early as 81 days before the product release. The data also includes 33 days' worth of exposures and product adoptions during the time period after the product release. In addition, the data records non-advertising information, including consumers' *count of other games*, *count of other franchise games*, and *number of previous game days*.

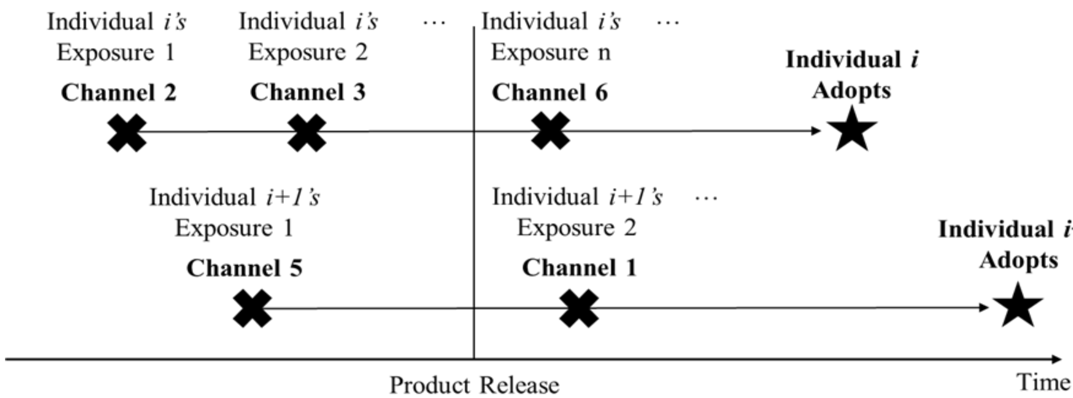

Figure 4. Advertising Exposures and Product Adoptions

The frequency of exposures before and after product release is presented in Figure 5. The vertical dashed line in the figure represents the product release. It can be seen in this figure that there were fluctuations in the observation window, which may reflect ebbs and flows in the company's advertising efforts and consumers' changing online behaviors.

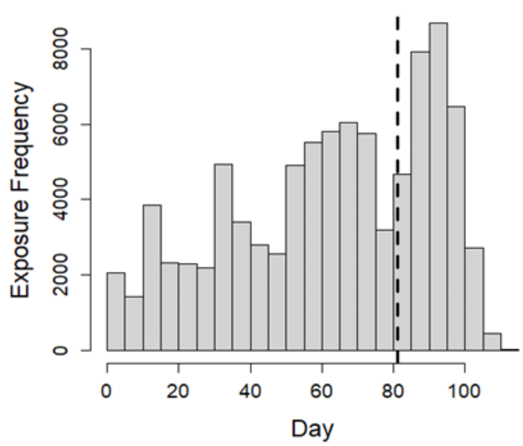

Figure 5. Density of Advertising Exposures

Advertising exposures occurred through six different digital advertising channels. Of all the recorded exposures, 9,573 occurred through channel A, 3,489 through channel B, 55,885 through channel C, 16,799 through channel D, 1,619 through channel E, and 2,587 through channel F. The nonuniform distribution of exposures across different channels could be due to the firm's channel selection strategies and the varying behavior of consumers across different digital channels.

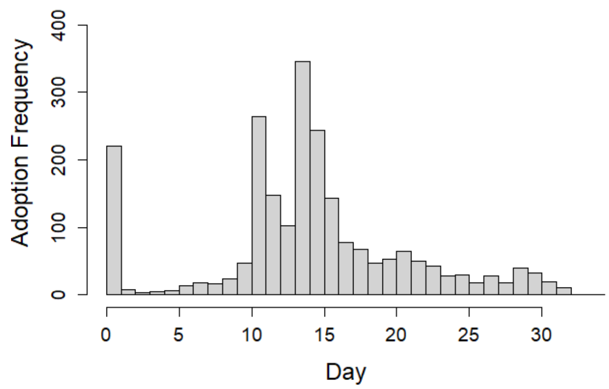

Figure 6. Product Adoptions

Figure 6 shows product adoptions after the product release. The adoption trend exhibits a few spikes. The spike on the product release day likely reflects the adoptions by those who, based on available information about the product, have made their adoption decisions prior to product release. The other spikes could be attributed to price changes.[2]

5.2. Benchmark Models

To evaluate the performance of the proposed model, we consider two benchmark hazard models. In the first benchmark model, the adstock covariates are measured based on the classical Koyck distributed-lag model (Koyck 1954). The model assumes that the effect of an exposure decays exponentially over time. As explained earlier, due to its intuitive and parsimonious form, this classical model has been frequently adopted even in recent studies. For simplicity, we name this model the *Koyck Model*, based on which the contribution of exposure *j* from channel *k* to consumer *i*'s adstock at time *t*, denoted by $K\_Adstock_{ijk}(t)$, is

$$K\_Adstock_{ijk}(t) = e^{-\lambda(t-t_{ijk})},$$

in which t_{ijk} is the time when consumer *i*'s exposure *j* from channel *k* occurs, and λ is the adstock's exponential decay rate. Similar to the FCASM, in the Koyck model, the adstock for consumer *i* at time *t* for exposures from channel *k* is

[2] In the next subsection, we account for these adoption spikes when selecting a hazard modeling framework. Further, we expect that, even without incorporating the price data, our model would not be much affected because assuming that consumers are uniformly influenced by price, the hazard ratio is not impacted by price changes.

$$k\_z_{ik}(t) = \sum_{j=1}^{n_{i_k}(t)} K\_Adstock_{ijk}(t), \tag{13}$$

in which $n_{i_k}(t)$ is the quantity of exposures from channel *k* that consumer *i* experiences up to *t*. Subsequently, the hazard rate of adoption for consumer *i* at time *t* can be expressed as

$$K\_h_i(t, KC_i(t)) = h_0(t)e^{\sum_k \theta_k k\_z_{ik}(t) + \sum_j \theta_j x_{ij}(t)}, \tag{14}$$

in which $KC_i(t)$ is the vector of time-varying covariates for consumer *i*. Similar to Eq. (10), in Eq. (14), θ_k is the coefficient that captures the effect of $k\_z_{ik}(t)$ on the adoption hazard, x_{ij} is the *j*th non-advertising covariate for consumer *i*, and θ_j is the coefficient that captures the effect of the *j*th non-advertising covariate on the adoption hazard.

We propose a second benchmark model that comes from the literature on stochastic models of consumer behavior. This model assumes that each exposure's adstock growth pattern follows the Erlang-2's density function. The Erlang-2 distribution has been used in the literature to describe a consumer's interpurchase timing (Chatfield and Goodhardt 1973, Schmittlein and Morrison 1983). We refer to this benchmark model as Erlang-2 model. In this benchmark model the contribution of exposure *j* from channel *k* to consumer *i*'s adstock at time *t*, denoted by $E\_Adstock_{ijk}(t)$, is

$$E\_Adstock_{ijk}(t) = \lambda^2(t - t_{ijk})e^{-\lambda(t - t_{ijk})},$$

where t_{ijk} is the time at which consumer *i*'s exposure *j* from channel *k* occurs, and λ denotes the rate of the Erlang-2 distribution, reflecting the rate at which an exposure's contribution to adstock first rises and then declines. As opposed to the Koyck model, in the Erlang-2 model, each individual exposure's adstock peaks. Note that under the proposed Erlang-2 model, the adstock for consumer *i* at time *t* generated by exposures from channel *k* and the hazard rate of adoption for consumer *i* at time *t* follow formulations similar to Eq. (13) and Eq (14), respectively.

5.3. Model Estimation and Performance Evaluation

The parameter estimation for a hazard model can be performed in a parametric, non-parametric, or semi-parametric fashion. In this study, we adopt a semi-parametric approach because the adoption trend in our dataset, as shown in Figure 6, exhibits spikes for various reasons, and we expect that the semi-parametric approach is more flexible than the parametric approach in dealing with the spikes. Specifically, we implement the widely-used Cox model (Cox 1972), a semi-parametric model, for performance evaluation. The details of model estimation and the estimation of the parameters' confidence intervals are explained in Appendix A. To operationalize the fractional operator used in the FCASM, we follow the procedure implemented by Lotfi et al. (2023), with corresponding details provided in Appendix B. Note that the generalized hazard model we develop based on the FCASM is more complicated than the base Cox model, resulting in a more computationally intensive model estimation process.

For model performance evaluation, we use Akaike Information Criteria (or AIC). According to Burnham and Anderson (2004), an AIC value is not interpretable by itself because it is influenced by the sample size and has arbitrary constants. Therefore, it is important to rescale AIC to $\Delta_i = \text{AIC}_i - \text{AIC}_{min}$, in which AIC_{min} is the minimum AIC value among *N* different AIC values corresponding to *N* different models being tested. Based on this transformation, the best model has $\Delta_i = 0$, while the remaining models have positive Δ_i values. As formulated by Burnham and Anderson, Δ_i allows for a meaningful interpretation. According to the prior study, models that have $\Delta_i \leq 2$ have substantial support, those with $4 \leq \Delta_i \leq 7$ have considerably less support, and models with $\Delta_i > 10$ essentially have no support.

5.4. Empirical Results and Discussions

We first estimate the proposed FCASM and the two benchmark models using the video game data. Estimation results are shown in Table 1. Next, we conduct tests for multicollinearity among the covariates. The variance inflation factor (VIF; Kutner, Nachtsheim, and Neter 2004) for the FCASM, Erlang-2 model and Koyck model are reported in Table 2, which shows that all VIF values are smaller than 2, thus ruling out the multicollinearity problem.

The results reported in Table 1 demonstrate that the two models that incorporate the nonmonotonic effect of advertising, i.e., the FCASM and Erlang-2 model, lead to smaller AIC values than does the Koyck model under which adstock for an ad exposure declines monotonically. Based on the results reported in Table 1, for the Koyck model we have $\Delta_i = 30329.57 - 30318.98 = 10.59$. Since Δ_i is larger than 10, according to Burnham and Anderson (2004), the Koyck model essentially has no significant support compared to the FCASM, which has the smallest AIC value. Thus, our empirical testing supports the expectation mentioned in Section 4.1 that, in the context of adopting a new video game, it is highly likely that a consumer's adstock does not follow the monotonic trend suggested by the Koyck distributed lag model.

Table 1. Comparison of FCASM, Erlang-2 Model, and Koyck Model

	Koyck	**Erlang-2**	**FCASM**
Channel A	-0.002	-0.074	-0.009
Channel B	0.002	0.436**	0.135
Channel C	0.003	0.048	0.032
Channel D	0.001	0.047	0.186***
Channel E	0.018**	0.409**	0.389**
Channel F	0.007	0.025	-0.010
Count of other games	0.016***	0.016***	0.016***
Count of other franchise games	0.081	0.077	0.081*
Previous game days	0.004***	0.004***	0.004***
p	-	-	0.000001
q	-	-	0.498*
β	-	-	0.23***
λ	0.054	0.093***	-
AIC	30329.57	30325.89	**30318.98**

Note: ***99%, **95%, *90% confidence intervals do not include zero.

For the Erlang-2 model, Δ_i = 30325.89 – 30318.98 = 6.91, barely placing in the [4, 7] range. This suggests that the Erlang-2 model has considerably less support than does the FCASM. Therefore, the results suggest that the FCASM more effectively captures the nonmonotonic effect of advertising than does the Erlang-2 model.

Table 2. VIF Values for FCASM, Erlang-2, and Koyck Model

	Koyck	**Erlang-2**	**FCASM**
Channel A	1.05	1.08	1.04
Channel B	1.15	1.13	1.05
Channel C	1.64	1.7	1.24
Channel D	1.8	1.86	1.22
Channel E	1.01	1.02	1.03
Channel F	1.1	1.12	1
Count of other games	1.11	1.11	1.11
Count of other franchise games	1.15	1.15	1.15
Previous game days	1.07	1.06	1.06

As reported in Table 1, for the FCASM, the estimated value of p is much smaller than that of q, indicating that consumers develop interest in the product and subsequently adstock primarily based on the word of mouth and network effects associated with the product. The presence of strong word of mouth and network effects reveals why the Koyck model and the Erlang-2 model, which do not incorporate these effects, empirically underperform compared to the FCASM.

Furthermore, from Table 1, we can see that the estimated fractional calculus-based memory coefficient (β) of the FCASM is much smaller than 1, suggesting that consumers' adstocks decline with a relatively high rate over time due to a rather weak memory rate. Specifically, with β=0.23, the memory trend is close to the bottom trend curve shown in Figure 1. When β=0.23, for a given ad exposure, only 25%, 15%, and 11% of the adstcoks developed in the past one day, two days, and three days, respectively, are maintained in the current day. This finding is consistent with a report indicating that consumers quickly forget the contents of advertisements and that 80% of them (i.e., four out of five consumers) forget branded content in three days (Inc. 2017).

Table 1 also demonstrates that the three non-advertising covariates under the FCASM are statistically significant, and, as expected, all positively influence the hazard rate of adopting the new video game. This result demonstrates that those who own more games and play more have a higher tendency to purchase the new video game.

Moreover, we compare the FCASM, Erlang-2 model, and Koyck model based on their out-of-sample prediction performance (Norwood, Roberts, and Lusk 2004). Specifically, we fit the models to a training sample and use the estimated parameters to calculate the log-partial likelihood value for the out-of-sample data. As reported in Table 3, with a 75-25 training-testing data split, the FCASM results in a larger log-partial likelihood value than the other two models, indicating a better out-of-sample prediction performance. This result demonstrates that the FCASM does not run the risk of overfitting, even though it has more parameters than the Erlang-2 model and Koyck model.

The FCASM's better empirical performance than the Koyck model and Erlang-2 model can be primarily attributed to the fact that the model benefits from a simple rationale that is also theoretically supported by the literature. Specifically, the FCASM can accommodate the nonmonotonic effect of advertising by assuming that a consumer's adstock growth is affected by a process through which the consumer first learns about the existence of a new product and then begins to accumulate relevant information based on which the consumer develops potential interest in the product (Barry 1987). Under the FCASM, the information accumulation and interest development process can assume a nonmonotonic shape, which in turn results in a nonmonotonic adstock growth trend. In accounting for consumers' information accumulation and interest development, the FCASM explicitly incorporates consumers' independent decision making as well as word of mouth and network effects, generating further insight into consumers' decision-making cycle. We believe that these factors jointly contribute to the FCASM's superior empirical performance.

Table 3. Out-of-Sample Prediction Performance Comparison

	Log-Partial Likelihood
Koyck	-3022.44
Erlang-2	-3022.06
FCASM	-3018.84

It is important to note that in the FCASM, the parameters p and q are grounded in the rich diffusion of innovations literature (Mahajan, Muller, and Wind 2000, Bass 2004) and capture the primary drivers that affect the rate of adoptions for a new product. By connecting the model to the product diffusion literature, users of the FCASM can draw into the huge knowledge base created by pioneering researchers in decades of research endeavors, and derive practical insights that are not possible with the benchmark models. Additionally, the application of fractional calculus to incorporate adstock erosion due to memory decay has resulted in a model that is both theoretically robust and empirically powerful (Tarasov 2018). The FCASM can help researchers and practitioners gain a deeper understanding regarding the short- and long-term effectiveness of advertising, and help them optimize advertising campaigns. Given these advantages, we propose that the FCASM should be the model of choice for measuring the long-term effect of advertising and digital attribution.

5.5. Nonmonotonic Effect of Advertising: Illustration and Explanation

One of the major findings of this study concerns the nonmonotonic adstock growth pattern. Given its importance, we conduct additional exploration to better understand the nonmonotonic effect of advertising. Figure 7 visually demonstrates the nonmonotonic pattern of a random consumer's adstock growth. The consumer's first advertising exposure occurs before product release. The estimated p, q, and β as reported in Table 1 are used to generate the adstock trend. Exposures are randomly generated and shown using vertical dotted lines. The time of product release is represented by the vertical solid line. As formulated in the FCASM, each exposure positively contributes to the consumer's adstock. The grey area in the figure represents the amount of adstock at different times for the consumer. It can be seen that adstock as estimated by the FCASM follows a nonmonotonic pattern, which is consistent with Bass and Clarke (1972), while deviates from the trend proposed by Koyck (1954) that assumes that the effect of ad exposure only declines over time.

Despite the similarity between the finding obtained based on the FCASM and that reported by Bass and Clarke, it is important to point out that the FCASM has several important advantages over the model proposed by Bass and Clarke. First, Bass and Clarke study the advertising effect for a nondurable product at an aggregate level, while the FCASM is developed to capture the influence of advertising on the adoption of a new durable product at the individual consumer level. Second, the FCASM accounts for factors that impact a potential consumer's adoption decision, including the independent decision-making effect as well as the word of mouth and network effects. In contrast, Bass and Clarke's model does not consider these effects. Third, the FCASM can be used in cases in which a company's marketing campaign starts before product release. On the contrary, Bass and Clarke deal with cases where advertising and sales take place concurrently, thus the FCASM is applicable to a broader range of settings.

Furthermore, by predicting consumers' peak adstock level, the FCASM can help firms avoid overspending on advertising. For example, based on the estimation results reported in Table 1, for a consumer who views a single exposure through channel E before the product release, the FCASM predicts the ad's impact on the consumer's purchase tendency will peak at 0.07 on the 27th day after the exposure has occurred. In contrast, the Koyck model as estimated in Table 1 suggests this same consumer after receiving seven more exposures, one every three days, reaches 0.06 purchase propensity 27 days after the first exposure. This implies that, according to the Koyck model, firms need to keep investing in advertising to maintain consumers' adstock levels, resulting in a suboptimal advertising strategy.

Additionally, the estimated FCASM as depicted in Figure 7 is also consistent with an important finding reported by Ghose and Todri-Adamopoulos (2016) that targeting consumers earlier in their purchase funnel can amplify the effectiveness of advertisement. According to the FCASM, ad exposures occurring earlier in the purchase cycle can contribute more to consumers' adstock. In particular, based on the adstock trend demonstrated in Figure 7, receiving exposures before the peak of the adstock trend can lead to a higher adstock peak.

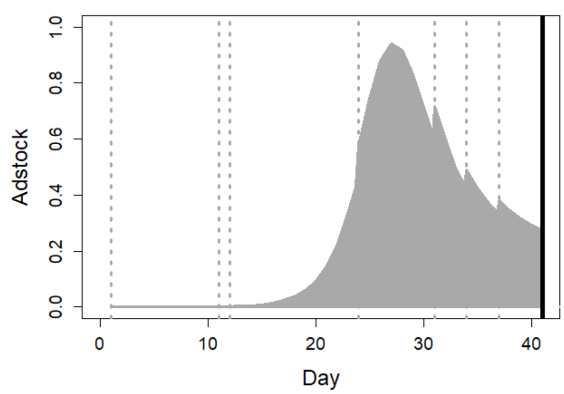

Figure 7. Estimated Adstock Based on FCASM for a Random Consumer

Furthermore, Figure 7 shows that the consumer's adstock reaches its peak before the product release. This means that the firm cannot benefit from the peak adstock because the product has not been made available in the market yet. This observation can help firms optimize the timing of their advertising campaigns.

6. Model Extensions

The primary contributions of our research are (i) developing a new adstock model (i.e., the FCASM) that incorporates consumer memory as well as the word of mouth and network effects, thereby accurately measuring consumers' adstocks over time, and (ii) determining the effect of each advertising channel in a multichannel setting using a generalized hazard modeling framework. We develop and empirically test model extensions to further enhance these two contributions. Specifically, concentrating on (i) we test whether the word of mouth and network effects change from before to after product release, and focusing on (ii) we examine whether the count of channels through which consumers view advertising has any effect on their hazard rate of adoption.[3]

6.1. Varying Word of Mouth and Network Effects Before and After Product Release

The results reported in Table 1 reveal that the nonmonotonic adstock trend emerges due to consumers' decisions being influenced by word of mouth and network effects. In the base FCASM, since the parameter q remains unchanged across the observation window, an implicit assumption is that the word of mouth and network effects are the same before and after product release. However, once the product is released in the market, these effects may change significantly due to more product information becoming available. In this subsection we investigate whether word of mouth and network effects have varying influences on consumers before and after product release. To this end, we develop a model extension in which the parameter q in the FCASM's underlying Bass model is time-dependent

$$q = \begin{cases} q_0 & t < product\ release\ date \\ q_0 + q_1 & t \geq product\ release\ date \end{cases}. \tag{15}$$

In Eq. (15), q_0 represents the influence of word of mouth and network effects for consumers who receive their first ad exposures before product release and q_1 captures the possible changes in these effects for consumers whose first ad exposures occur after the product release. We name this model extension the *Extended Fractional Calculus-Based Advertising Stock Model* or *EFCASM*. The empirical results corresponding to the EFCASM are reported in Table 4. As detailed in the table, q_1 is positive and

[3] To further test the robustness of the fractional calculus-based memory formulation used in the new adstock model (i.e., the FCASM), we develop and empirically test two model versions: one with a time-varying memory rate and another with an exposure-influenced memory rate. Our empirical results demonstrate that the two model versions do not perform better than the base FCASM, thereby confirming the robustness of the memory formulation used in the base FCASM. Further details are provided in Appendix C.

statistically significant, indicating that the word of mouth and network effects are stronger for consumers whose first ad exposures happened after product release. Specifically, based on the estimated parameters of the EFCASM as reported in Table 4 (i.e., the estimated p, q_0, q_1, and β), the adstock trend for a consumer with a single ad exposure occurring after product release peaks at around four days. However, consistent with the adstock trend shown in Figure 7 based on the FCASM, the adstock trend under the EFCASM for a consumer whose single ad exposure occurred before product release peaks at around 27 days. This result could be attributed to the fact that, after product release, consumers who have adopted the product can share more concrete and convincing information, thus having a strong impact on consumers who have just started to collect information about the product. In contrast, the word of mouth and network effects might be weaker before product release because consumers' opinions are to a great extent based on speculations.

In sum, the empirical findings based on the EFCASM suggest that (i) consistent with the findings from the base FCASM, consumers' adstocks are developed predominantly based on the word of mouth and network effects, which are not explicitly modeled in the prior literature, and (ii) these effects grow stronger when more information about the product becomes available. Furthermore, consistent with the results from the FCASM, the estimated value for the EFCASM's memory parameter reflects relatively low consumer memory. Additionally, comparing the EFCASM's AIC value as reported in Table 4 with that of the FCASM reported in Tables 1 and 4 demonstrates that the EFCASM fits the data better than does the FCASM. Therefore, considering that the EFCASM generates more insight into the nonmonotonic behavior of advertising and that it empirically performs better than the FCASM, we select the EFCASM over the FCASM for cases in which advertising starts before product release.

The findings from the EFCASM bear practical implications. Specifically, the varying magnitude of word of mouth and network effects as captured by the model helps firms more accurately predict purchase timing for consumers who begin their purchase journey at different times. Additionally, word of mouth before product release is expected to be predominantly shaped by influencers who may have access to a pre-production version of the product, while word of mouth after product release is expected to be largely formed by consumers who have experienced the actual product. Therefore, the proposed model enables firms to compare the effect of *influencer-based word of mouth* with that of *consumer experience-based word of mouth,* aiding firms in planning their investment in influencer-based marketing. Additionally, consistent with the results from the FCASM, the findings from the EFCASM suggest that firms should time their advertising campaigns optimally to make the most of the adstock peak period for the consumers who start their purchase cycle before product release.

6.2. Advertising Channel Versatility

Firms often advertise through multiple channels, resulting in some consumers viewing ads through more than one channel. Therefore, a primary focus of our research is dedicated to developing a generalized hazard model that incorporates the effect of multiple channels, i.e., Eq. (11). In a multichannel setting, some inquiries naturally arise: Does the number of channels through which a consumer is exposed to advertising impact the consumer's product adoption behavior? If so, how many channels should a firm utilize for advertising? We extend the proposed proportional hazard model to answer these questions and help firms improve their advertising strategy. Specifically, we include *Count of channels* as a covariate in the adoption hazard. To account for the possibility that the association between *Count of channels* and the hazard rate of adoption is nonlinear, we incorporate the quadratic form of *Count of channels* into the hazard model presented in Eq. (11)

$$h_i(t, C_i(t)) = h_0(t)e^{\sum_k \theta_k z_{ik}(t) + \theta_c Count\ of\ channels + \theta_{c2} Count\ of\ channels^2 + \sum_j \theta_j x_{ij}(t)}. \qquad (16)$$

In Eq. (16), θ_c and θ_{c2} are coefficients for linear and quadratic terms of *Count of channels,* respectively. Considering that we selected the EFCASM over the FCASM in the previous subsection, to measure consumers' adstocks from different channels, we use the EFCASM.

The estimation results for the model version that incorporates the quadratic form of *Count of channels* are reported in Table 4. As can be seen from the table, the estimated coefficient for *Count of channels* is positive and significant, while that for *Count of channels*^{2} is negative and significant, indicating a convex relationship between the hazard rate of adoption and *Count of channels*.[4] Additionally, the estimated values for the adstock models' parameters (i.e., p, q_0, q_1, and β) remain mostly unchanged after including *Count of channels*, reflecting consistency of the results. Furthermore, the AIC value improves after incorporating *Count of channel*, due most likely to having more covariates in the model. Therefore, we select the model version with *Count of channels* because it casts light on the influence of the number of channels through which consumers view advertising on their adoption hazard and that it empirically performs better than the model version that does not incorporate *Count of channel*.

Figure 8 further illustrates the effect of the number of channels on adoption hazard. Here, we assume $CountEffect = e^{\theta_c Count\ of\ channels + \theta_{c2} Count\ of\ channels^2}$, which multiplicatively influences the baseline hazard of adoption ($h_0(t)$) in Eq. (18). Figure 8 demonstrates that delivering advertising through one or two channels contributes to the highest adoption propensity, while utilizing more than two channels may lead to a decline in adoption tendency.

Table 4. Comparison of Model Extensions with FCASM

	FCASM	**EFCASM**	**EFCASM** with Count of Channels
Channel A	-0.009	-0.056	-0.028
Channel B	0.135	0.107	0.109
Channel C	0.032	0.042	0.039**
Channel D	0.186***	0.239***	0.180***
Channel E	0.389**	0.353	0.356**
Channel F	-0.010	-0.007	0.022
Count of channels	-	-	0.932**
Count of channels^2	-	-	-0.309***
Count of other games	0.016***	0.016***	0.016***
Count of other franchise games	0.081*	0.08*	0.087*
Previous game days	0.004***	0.004***	0.004***
p	0.000001	0.00001	0.000005
q	0.498*	-	-
q_0	-	0.414**	0.453**
q_1	-	3.336***	3.627**
β	0.23***	0.207***	0.295***
AIC	30318.98	30313.02	30270.99

Note: ***99%, **95%, *90% confidence intervals do not include zero.

[4] Our empirical testing demonstrates that the associations of *Count of channels* and *Count of channels*^{2} with adoption hazard may be time dependent. Therefore, when computing *Count of channels* and *Count of channels*^{2}, we multiply them with a time-transformation of 1/time. This transformation implies that *Count of channels* has the strongest impact on the product adoption hazard shortly after the product release.

The primary practical takeaway from this result is that exposing consumers to multiple advertising channels may prove counter-productive because it may cause fatigue, annoyance, or confusion to consumers. Therefore, firms may be better off concentrating on a few select channels with more substantial influences on consumers' purchase decisions. Based on the results detailed in Table 4, not all channels have a statistically significant effect in our data sample. This result, coupled with the observation from Figure 8 that viewing advertising through more than two channels can reduce consumers' adoption propensity, suggest that the firm should concentrate on two of the statistically significant channels that have a larger impact on consumers' adoption tendency, which are channels D and E.

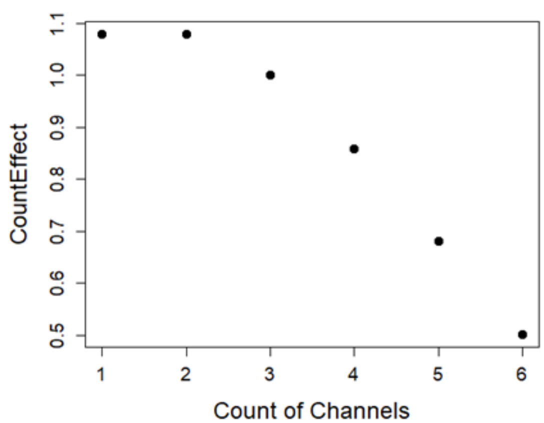

Figure 8. Effect of Count of Channels on Adoption Hazard

7. Conclusions and Directions for Future Research

The primary goal of this research is to measure the effect of advertising in a multichannel setting, which stands as one of the main ongoing challenges concerning online advertising. Challenges in the measurement of ad effect can be attributed to factors such as long purchase cycles and unobserved stages in consumers' decision-making process (Gordon et al. 2021). According to the literature, the effect of advertising may exhibit a monotonic or nonmonotonic temporal trend. While the possible nonmonotonic effect of advertising has been reported in the literature, to the best of our knowledge, no formal modeling framework has been introduced to explain this behavior at the individual level. In this research, we propose a new advertising stock model named the *Fractional Calculus-Based Advertising Stock Model* (FCASM) to fill this void. Under the proposed model, a consumer's adstock can either monotonically decline or peak in the middle of a time window, conforming to one of the two main advertising behaviors reported in the prior literature.

An innovative feature of the new adstock model is that it uses fractional calculus to tackle the challenging problem of modeling the decaying ad effect due to consumers' memory decay, thereby capturing the lasting effect of ad exposures. Specifically, the proposed model incorporates a fractional calculus-based power-law memory trend that resembles human memory to account for consumers' memory decay and the corresponding decline in their advertising stocks. This innovation has a solid theoretical foundation and improves model interpretability and empirical performance.

Based on the new adstock model, we develop a generalized proportional hazard model to estimate the influence of exposures from different digital advertising channels on consumers' timing of product adoptions. We test our model using a dataset that includes consumers' exposures to advertising through different channels and their product adoptions. This dataset provides a great opportunity to study the effectiveness of advertising because companies rarely are able to monitor advertising exposures and the corresponding consumer conversions in a single-source dataset (Gordon et al. 2021). Consistent with the literature, we find that the consumers' adstocks demonstrate a nonmonotonic pattern—it first increases to a peak before declining. Furthermore, in a comparison with two benchmark models, we show that the proposed model outperforms them in model fitting, out-of-sample prediction, as well as interpretability.

Notably, the new model accounts for the effect of independent decision-making, as well as the word of mouth and network effects. To the best of our knowledge, no model from the previous literature, including the widely-used Koyck model, accounts for these effects. Our empirical findings demonstrate that the nonmonotonic effect of advertising emerges primarily due to the word of mouth and network effects, further underscoring the importance of including these effects. Considering that our research primarily concentrates on studying the common practice of starting advertising campaigns before product release and to further investigate the role of word of mouth and network effects, we develop a model extension, named the *Extended Fractional Calculus-Based Advertising Stock Model* (EFCASM), which distinguishes between these effects before and after product release. The empirical results based on the EFCASM uncovers a significant leap in the word of mouth and network effects post product release, most likely due to more product information becoming available.

Our empirical findings can benefit the advertising practice. Specifically, using the new model, firms can accurately estimate consumers' adstocks over time and predict the timing of consumer purchases, enabling them to effectively time their advertising campaigns and product releases. Furthermore, the precise estimation of consumers' adstocks over time can help firms prevent overspending on advertising. For instance, the proposed model enables firms to predict consumers' peak adstock periods during which advertising spending can be reduced, thus avoiding excessive expenditure. Moreover, considering that the EFCASM distinguishes between word of mouth and network effects pre- and post-product release and that word of mouth before product release is expected to be primarily influencer-based, the EFCASM enables firms to optimally invest in *influencer-based word of mouth*. Finally, our results suggest that firms may benefit from selecting a smaller number of the advertising channels that have the highest impact on consumers' purchase tendency. The proposed model can help identify the most efficient channels and determine the optimal number of channels to employ in advertising campaigns.

There exist multiple avenues for future research. For example, our dataset records only advertising through digital channels. With a richer dataset, a future study could analyze the integration of digital and traditional channels. Furthermore, the advertising channels are unknown in our data, a dataset with more information on advertising channels can lead to further insights.

Acknowledgements

The authors acknowledge the supportive role of the Wharton Customer Analytics Initiative (WCAI) for granting access to the dataset used in this research. Furthermore, the authors are thankful for the support of the company that provided the data, particularly the company's data science team that collaborated with the authors in the early stages of the project.

References

Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of econometrics*, 73(1), 5-59.
Baleanu, D., Diethelm, K., Scalas, E., & Trujillo, J. J. (2012). *Fractional calculus: models and numerical methods* (Vol. 3). World Scientific.
Barry, T. E. (1987). The development of the hierarchy of effects: An historical perspective. *Current issues and Research in Advertising*, *10*(1-2), 251-295.
Bass, F. M. (1969). A new product growth for model consumer durables. *Management science*, *15*(5), 215-227.
Bass, F. M. (2004). Comments on "a new product growth for model consumer durables the bass model". *Management science*, *50*(12_supplement), 1833-1840.
Bass, F. M., & Clarke, D. G. (1972). Testing distributed lag models of advertising effect. *Journal of Marketing Research*, *9*(3), 298-308.
Bass, F. M., Jain, D., & Krishnan, T. (2000). Modeling the marketing-mix influence in new-product diffusion. *New-product diffusion models*, 99-122.
Bass, F. M., Krishnan, T. V., & Jain, D. C. (1994). Why the Bass model fits without decision variables. *Marketing science, 13*(3), 203-223.
Bennedsen, J., & Caspersen, M. E. (2009). Recalling Programming Competence. 9th Koli Calling International Conference on Computing Education Research.
Boltzmann, L. (1876). On the theory of the elastic aftereffect. *Poggendorff's Ann. Phys. Chem*, *7*, 624-645.
Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research*, *33*(2), 261-304.
Chae, I., Bruno, H. A., & Feinberg, F. M. (2019). Wearout or weariness? Measuring potential negative consequences of online ad volume and placement on website visits. *Journal of Marketing Research*, *56*(1), 57-75.
Chatfield, C., & Goodhardt, G. J. (1973). A consumer purchasing model with Erlang inter-purchase times. *Journal of the American Statistical Association*, *68*(344), 828-835.
DR, C. (1972). Regression models and life tables. *JR Stat Soc*, *34*, 248-275.
Ghose, A., & Todri-Adamopoulos, V. (2016). Toward a digital attribution model. *MIS quarterly*, *40*(4), 889-910.
Gordon, B. R., Jerath, K., Katona, Z., Narayanan, S., Shin, J., & Wilbur, K. C. (2021). Inefficiencies in digital advertising markets. *Journal of Marketing*, *85*(1), 7-25.
Inc. (2017). *How to Prevent 80% of Viewers from Forgetting Your Branded Content*. https://www.inc.com/peter-roesler/how-to-prevent-80-of-viewers-from-forgetting-your-branded-content.html, Retrieved, August, 12, 2023.
Jain, D. C., & Vilcassim, N. J. (1991). Investigating household purchase timing decisions: A conditional hazard function approach. *Marketing Science*, *10*(1), 1-23.
Kilbas, A. A., Srivastava, H. M., & Trujillo, J. J. (2006). *Theory and applications of fractional differential equations* (Vol. 204). elsevier.
Koyck, L. M. (1954). *Distributed lags and investment analysis* (Vol. 4, p. 145). Amsterdam: North-Holland Publishing Company.
Kutner, M. H., Nachtsheim, C. J., & Neter, J. (1983). Applied linear regression models. Richard D. Irwin, Inc.
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.

Lotfi, A., Jiang, Z., Naoum-Sawaya, J., & Begen, M. A. (2024). Modeling Sales of Multigeneration Technology Products in the Presence of Frequent Repeat Purchases: A Fractional Calculus-Based Approach. *Production and Operations Management*, 10591478241240744.
Lotfi, A., Naoum-Sawaya, J., Lotfi, A., & Jiang, Z. (2024). To Skim or not to Skim: Studying the Optimal Pricing Strategy for Technology Products. *Omega*, *127*, 103079.
Lotfi, A., Jiang, Z., Lotfi, A., & Jain, D. C. (2023). Estimating life cycle sales of technology products with frequent repeat purchases: a fractional calculus-based approach. *Information Systems Research*, *34*(2), 409-422.
Mahajan, V., Muller, E., & Wind, Y. (Eds.). (2000). *New-product diffusion models* (Vol. 11). Springer Science & Business Media.
Nerlove, M., & Arrow, K. J. (1962). Optimal advertising policy under dynamic conditions. *Economica*, 129-142.
Norwood, B., Roberts, M. C., & Lusk, J. L. (2004). Ranking crop yield models using out-of-sample likelihood functions. *American Journal of Agricultural Economics*, *86*(4), 1032-1043.
Palda, K. S. (1965). The measurement of cumulative advertising effects. *The Journal of Business*, *38*(2), 162-179.
Podlubny, I. (1999). *Fractional differential equations*, Academic Press.
Qu, X., Lotfi, A., Jain, D. C., & Jiang, Z. (2022). Predicting upgrade timing for successive product generations: An exponential-decay proportional hazard model. *Production and Operations Management*, *31*(5), 2067-2083.
Sawhney, M. S., & Eliashberg, J. (1996). A parsimonious model for forecasting gross box-office revenues of motion pictures. *Marketing Science*, *15*(2), 113-131.
Samko, S.G., Kilbas, A.A., & Marichev, O.I. (1993). *Fractional integrals and derivatives-theory and applications*. Gordon and Breach. Linghorne, PA.
Shao, X., & Li, L. (2011). *Data-driven multi-touch attribution models*. Presented at Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York: ACM, pp. 258-264.
Shapiro, B. T., Hitsch, G. J., & Tuchman, A. E. (2021). TV advertising effectiveness and profitability: Generalizable results from 288 brands. *Econometrica*, *89*(4), 1855-1879.
Statista. (2023). *Digital advertising - United States*. https://www.statista.com/outlook/dmo/digital-advertising/united-states, Retrieved January 15, 2024.
Tarasov, V. E. (2018). Generalized memory: Fractional calculus approach. *Fractal and Fractional*, *2*(4), 23.
Terui, N., Ban, M., & Allenby, G. M. (2011). The effect of media advertising on brand consideration and choice. *Marketing science*, *30*(1), 74-91.
Teyssière, G., & Kirman, A.P. eds. 2006. *Long memory in economics*. Springer Science & Business Media.
Todri, V., Adamopoulos, P., & Andrews, M. (2022). Is distance really dead in the online world? The moderating role of geographical distance on the effectiveness of electronic word of mouth. *Journal of Marketing*, *86*(4), 118-140.
Todri, V. (2022). Frontiers: The impact of ad-blockers on online consumer behavior. *Marketing Science*, *41*(1), 7-18.
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.
Zantedeschi, D., Feit, E. M., & Bradlow, E. T. (2017). Measuring multichannel advertising response. *Management Science*, *63*(8), 2706-2728.
Zhan, L., Guo, D., Chen, G., & Yang, J. (2018). Effects of repetition learning on associative recognition over time: Role of the hippocampus and prefrontal cortex. *Frontiers in human neuroscience*, *12*, 277.

Online Appendix

Appendix A. Model Estimation

In this Appendix, we explain parameter estimation for the Cox model and the estimation of the confidence intervals for the model parameters.

A.1. Model Estimation

The Cox model leaves the baseline hazard rate $h_0(t)$ unspecified. Specifically, the hazard ratio of two consumers i and i' with the vectors of covariates C_i and $C_{i'}$, respectively, and the vector of coefficients θ can be written as

$$\frac{h_i(t,C_i)}{h_{i'}(t,C_{i'})} = \frac{h_0(t)\exp(C_i\beta)}{h_0(t)\exp(C_{i'}\beta)} = \frac{\exp(C_i\theta)}{\exp(C_{i'}\theta)}. \tag{A1}$$

It can be seen in Eq. (A1) that the hazard ratio is independent of the baseline hazard $h_0(t)$. We next elaborate on our model estimation procedure.

We follow Cox's (1972) proposed partial likelihood to estimate the vector of coefficients θ, which in the FCASM and the benchmark models include θ_ks and θ_js, reflecting the influence of advertising and non-advertising covariates, respectively, on adoption. Cox proposes a partial likelihood method for θ without involving the baseline hazard. Assume that X_i is an adoption time censored random variable and C_i is a set of covariates. Suppose that $\tau_1 <\ldots< \tau_K$ are K distinct adoption times and assume that there are no tied adoption times, meaning that each adoption takes place at a different time. Let $R(t)=\{i : X_i \geq t\}$ be the set of consumers who are *at risk* of adoption at time t. At each adoption time X_j, the contribution to the likelihood is

$$\frac{h_0(X_j)e^{\theta' C_j}}{\sum_{l\epsilon R(X_j)} h_0(X_j)e^{\theta' C_l}}.$$

The partial likelihood for all adoptions is

$$L(\theta) = \prod_{j=1}^{K} \frac{h_0(X_j)e^{\theta' C_j}}{\sum_{l\epsilon R(X_j)} h_0(X_j)e^{\theta' C_l}}$$

$$= \prod_{j=1}^{K} \frac{e^{\theta' C_j}}{\sum_{l \epsilon R(X_j)} e^{\theta' C_l}}. \tag{A2}$$

For data with censoring we can write

$$L(\theta) = \prod_{j=1}^{n} \left[\frac{e^{\theta' C_j}}{\sum_{l \epsilon R(X_j)} e^{\theta' C_l}} \right]^{\delta_i}, \tag{A3}$$

where δ_i=1 for adopting consumers, δ_i=0 for censored consumers at time X_j, and n is the number of consumers in the study. Note that Eq. (A2) and Eq. (A3) are equal. Now assume that there are ties in adoption times, meaning that there are adoptions that occur on the same day. One popular method to handle ties is suggested by Breslow (1972)

$$L(\theta) = \prod_{j=1}^{K} \prod_{i=1}^{d_j} \frac{e^{\theta' C_j^i}}{\sum_{l \epsilon R(X_j)} e^{\theta' C_l}} \tag{A4}$$

where d_j is the number of adoptions at τ_j and C_j^i is the vector of covariates for the ith consumer who adopts at time τ_j. Considering that our covariates are time-varying, we modify Eq. (A4) as

$$L(\theta) = \prod_{j=1}^{K} \prod_{i=1}^{d_j} \frac{e^{\theta' C_j^i(t)}}{\sum_{l \epsilon R(X_j)} e^{\theta' C_l(t)}}, \tag{A5}$$

where $C_j^i(t)$ shows the values at t of the vector of time-varying covariates for the ith consumer who adopts at time τ_j. Taking the log of Eq. (A5) we get the log-partial likelihood as

$$LL(\theta) = \sum_{j=1}^{k} \sum_{i=1}^{d_j} (\theta' C_j^i(t) - \log\left(\sum_{l \epsilon R(X_j)} e^{\theta' C_l(t)}\right)). \tag{A6}$$

To estimate θ, we can use an optimization method to maximize Eq. (A6) by changing θ. We conduct model estimations in R. We use the *maxlik* function to maximize the log-partial likelihood for the models we estimate.

A.2. Estimating Confidence Intervals

We use the profile likelihood approach for estimating the confidence intervals of the estimated parameters. Profile-likelihood-based confidence intervals are specifically useful for nonlinear models (Royston 2007). Assume θ_i is an element of θ and is a scalar parameter of particular interest. We would like to construct a profile-likelihood function for θ_i. Let the true value of θ_i in the population be θ_{i_0} and θ_i's profile likelihood function be $l_p(\theta_i)$. What we need to conduct to compute the $l_p(\theta_i)$ and the confidence interval corresponding to θ_{i_0} is to fix θ_i's value and find the maximum likelihood estimation for the remaining parameters of θ. To find the confidence interval, this process is repeated for an appropriate set of values of θ_i until the range $(\theta_{i_{left}}, \theta_{i_{right}})$ is found, which satisfies the equality

$$2\left\{l\left(\widehat{\theta_\iota}\right) - l_p(\theta_{i_{left}})\right\} = 2\left\{l\left(\widehat{\theta_\iota}\right) - l_p(\theta_{i_{right}})\right\} = C_{1;1-\alpha}$$

in which $C_{1;1-\alpha}$ is the (1- α)th quantile for the χ^2 distribution with 1 degree of freedom. In the models we test, $l\left(\widehat{\theta_\iota}\right)$ is the maximum log-partial likelihood estimated by changing all model parameters. For the models we examine, we examine the 90%, 95%, and 99% confidence intervals of all model parameters including those for the covariates incorporated into the Cox model and the parameter(s) of the adstock model based on which the consumers' adstocks for impressions from different channels are calculated.

References

Breslow, NE, 1972. Contribution to the discussion of Cox (1972), *Journal of the Royal Statistical Society* B, 34: 216-217.

Cox, D.R., 1972. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), pp.187-202.

Royston, P., 2007. Profile likelihood for estimation and confidence intervals. *The Stata Journal*, *7*(3), pp.376-387.

Appendix B. Operationalization of the Fractional Calculus-Based Operator used in FCASM

Estimating the FCASM using the Riemann-Liouville integral shown in Eq. (1) may be computationally challenging. Therefore, we use a substitute mathematical operator that mimics the Riemann-Liouville integral and is computationally less demanding. Specifically, following Lotfi, Jiang, Lotfi, and Jain (2023), we use the operator $I_{n,k}^{\beta}$ in which k and n are the approximate operator's parameters. The developed approximate operator uses the n-point Gauss quadrature formula proposed for integrals (DeVore and Scott 1984) and Spouge's approximated formulation for the Gamma function (Spouge 1994). Increasing the values of k and n leads to $I_{n,k}^{\beta}$ converging to the original operator I^{β}, i.e., the Riemann-Liouville integral. Lotfi et al. define $I_{n,k}^{\beta}$ as

$$I_{n,k}^{\beta} y(t) := \frac{y(0)t^{\beta}}{G_k(1+\beta)} + \frac{y'(0)t^{1+\beta}}{G_k(2+\beta)} + I_{n,k}^{2+\beta} y''(t), \tag{B1}$$

where β is the order of the operator, $y(t)$ is the function on which the operator is applied, and

$$I_{n,k}^{2+\beta} y''(t) := \frac{1}{G_k(2+\beta)} \frac{t}{2} \sum_{i=1}^{n} w_i \left(\frac{t}{2}(1-x_i)\right)^{1+\beta} y''\left(\frac{t}{2}(x_i+1)\right). \tag{B2}$$

In Eq. (B2), x_i and w_i denote the quadrature nodes and weights (DeVore and Scott 1984), and $G_k(.)$ is defined as

$$G_k(x) = (x-1+h)^{x-\frac{1}{2}} e^{-(x-1+h)} \sqrt{2\pi} \left[c_0 + \sum_{i=1}^{k} \frac{c_i(h)}{x-1+i}\right],$$

in which $c_0 = 1$, the parameter h is real, $k = \lceil h \rceil - 1$, and

$$c_i(h) = \frac{(-1)^{i-1}(-i+h)^{i-\frac{1}{2}}}{\sqrt{2\pi}(i-1)!} e^{-i+h}.$$

The detailed derivation of $I_{n,k}^{\beta}$ and the proof of its convergence to the original operator are provided in Lotfi et al. (2023).

References

DeVore, R.A. and Scott, L.R., 1984. Error bounds for Gaussian quadrature and weighted-L^1 polynomial approximation. *SIAM journal on numerical analysis*, *21*(2), pp.400-412.

Lotfi, A., Jiang, Z., Lotfi, A., & Jain, D.C. 2023. Estimating Life Cycle Sales of Technology Products with Frequent Repeat Purchases: A Fractional Calculus-Based Approach. *Information Systems Research*, 34(2): 409-422.

Spouge, J.L., 1994. Computation of the gamma, digamma, and trigamma functions. *SIAM Journal on Numerical Analysis*, *31*(3), pp.931-944.

Appendix C. Validating Robustness of Fractional Calculus-Based Advertising Memory

In this appendix we develop and empirically test model extensions to evaluate the robustness of the fractional calculus-based memory formulation used in the FCASM. Specifically, we examine whether the memory rate (i) changes over time or (ii) is influenced by ad exposures.

C.1. Time-Varying Memory Rate

In the base FCASM, the memory rate captured by the memory coefficient β is constant. However, as the consumers are collecting relevant information to form their opinion of the product, they likely receive repetitive information. It is known that repetition learning can improve memory performance (Zhan, Guo, Chen, and Yang 2018). Based on this expectation, we develop a model version in which memory rate increases over time. Considering that it is expected that the memory rate improves as the information collection and interest development process progresses, for consumer i, we consider a time-varying coefficient of memory $\beta_i(t)$ that increases as the cumulative amount of information received increases

$$\beta_i(t) = \beta \int_{t_i}^{t} y(\tau - t_i)\, d\tau. \tag{C1}$$

In Eq. (C1), $y(.)$ represents the new information acquired at a given point in time in $[t_i, t]$ and t_i denotes the beginning of the information collection for consumer i, which in our model operationalization is assumed to coincide with the consumer's first ad exposure. Considering that in the FCASM, we use the bell-shaped Bass density function for $y(.)$, $\int_{t_i}^{t} y(\tau - t_i)\, d\tau$ becomes an S-shaped trend that runs between 0 and 1, resulting in a time-varying memory rate that increases over time from 0 to β. Based on Eq. (C1), β is the highest memory rate consumers can achieve as they progress in their purchase cycle. Our empirical testing shows that this model extension does not result in an AIC value smaller than that of the FCASM reported in Table 1. Therefore, in the absence of empirical support in favor of an increasing memory as formulated in Eq. (C1), we select the simpler FCASM.

C.2. Exposure-Influenced Memory

In the base FCASM, advertising exposures are assumed to have no influence on memory. In this subsection, we introduce an extension of the FCASM in which the adstock memory is influenced by exposures. Specifically, in this model version, with more ad exposures occurring, the memory grows stronger. The rationale behind this model version is similar to the one used in developing the extension based on Eq. (C1). Particularly, more exposures may serve as a repetition of product-related information acquired in the past, resulting in a decline in the memory decay rate. Therefore, we formulate the memory rate as

$$\beta = \beta_0 + 1 - \exp(-\sigma * n_i(t)), \tag{C2}$$

where β_0 is the base memory, $n_i(t)$ is the number of ad exposures for consumer *i* until time *t* and σ represents the rate at which memory improves with more exposures. We limit the memory rate β shown in Eq. (C2) to run between 0 and 1. In our empirical testing the value of σ approaches zero, indicating that the extended model reduces to the base FCASM with constant memory rate reported in Table 1. Therefore, we do not find any empirical evidence that memory improves with the occurrence of more ad exposures.