DSSR

Discussion Paper No. 123

A Multi-generational Diffusion Model with Social Media Effects and Pre-Launch Forecasting of Smart Phone

Yinxing Li and Nobuhiko Terui

June, 2021

Data Science and Service Research
Discussion Paper

Center for Data Science and Service Research
Graduate School of Economic and Management
Tohoku University
27-1 Kawauchi, Aobaku
Sendai 980-8576, JAPAN

A Multi-generational Diffusion Model with Social Media Effects and Pre-Launch Forecasting of Smart Phone

Yinxing Li and Nobuhiko Terui[1]

Tohoku University

Graduate School of Economics and Management

June, 2021

---

[1] Corresponding Author: Yinxing Li (yinxing.li.a8@tohoku.ac.jp)
Nobuhiko Terui (terui@tohoku.ac.jp) acknowledges a grant by JSPS KAKENHI Grant Number (A)17H01001.

A Multi-generational Diffusion Model with Social Media Effects and Pre-Launch Forecasting of Smart Phone

This paper proposes a diffusion model for multi-generational smart phone by using social media which allows us to forecast sales not only at early stage after launch of specific generation but also before its launch. The model is based on multigenerational generalized Bass model and includes the hierarchical model on the structure of parameters for connecting sequential generations. The social media topics are extracted by the labelled dynamic topic model and they are plugged in the adoption rate function and hierarchical model for parameters as covariates. The model reveals how social media accelerates and decelerates the diffusion in the pre-launch and post-launch phases. Unlike previous multi-generational diffusion models, this model forecasts sales of new-generation products before launch.

The empirical results show that the model forecasts the unlaunched product sales with better precision compared to extensive comparative models including sentiment analysis and non-diffusion model, and social media topics accelerate sales in the pre-launch period and their effects decrease with varying patterns in the post-launch period. In conjunction with its effects on switching and leapfrogging, the model provides useful information for product management over long range time horizon.

Key words: pre-launch forecasting, multi-generational diffusion, hierarchical model, social media, topic model, word-of-mouth

1. Introduction

With the rapid development of social media, the impact of the electronic word-of-mouth effect on the diffusion of products has drawn increasing attention. As more customers tend to read online reviews before purchasing, it becomes important to measure and understand the role of these unstructured data. Compared to one-way communication through traditional media, the diffusion process of a new product has become highly affected by social media. The amount of information about a recently released product is larger and more varied than it has ever been; people can rely on in-person word of mouth (WOM), expert product reviews, YouTube reviews, comments on social media, etc. As a field, we benefit from obtaining a deeper view of how these information sources affect product purchases.

As literature provides empirical evidence that customer WOM has a significant impact on customer purchasing behavior (e.g., Chevalier & Mayzlin, 2006) and a growing number of studies have examined the influence of user-generated content in marketing.

Most WOM research focuses on two types of WOM. The first type is the consumer reviews on retail platforms such as Amazon.com. For example, Lee and Bradlow (2011) show that the automated analysis of consumer reviews promises to support managerial decision making both descriptively and prescriptively instead of the expert guidebook and user survey, and Moe and Trusov (2011) find that WOM valence represented by the rating is helpful for sales forecasting. Additionally, Tirunillai and Tellis (2014) show that the latent topics extracted from consumer reviews are also informative when measuring customer experience such as perceived quality. The second type of WOM exists in the form of the comments posted on social media such as Twitter, Facebook, and bulletin board systems (BBS). For example, Netzer et al. (2012) offer empirical evidence that WOM from consumer forums provides insightful information for firms to investigate the product positioning often at a lower cost in comparison to traditional data sources. In addition, the valence represented by the positivity of WOM is also demonstrated to increase the precision of forecasting such as stock or sales (e.g., Chen et al., 2017; Ritesh et al., 2017). These works focus on the effect of WOM after the product is released, and it is the post-launch effects of social media.

Compared to retail platforms, WOM in social media is not only available after the product is released but also available before the product is launched. The use of social media to predict demand before product launch has aroused considerable recent interest (Gopinath et al., 2013; Marchand et al., 2016; Divakaran et al., 2017; Kim & Hanssens, 2017). It becomes critical for managers to make proper marketing decisions for pre-launched products with very limited data history, or even without any data history – likely overestimating or underestimating demand – which can influence potential profit and cost management (Croxton et al., 2002). These studies showed that WOM before product launch affects future marketing as the pre-launch effects of social media.

Although there are already many implications of social media on marketing, there are mainly three limitations. First, most studies focus on only the roles of social media in the pre-launch or post-launch period. While several previous studies acknowledge that the effect of WOM differs substantially in different situations, such as the types of WOM (customer reviews, social media), there is no study to discuss the different roles of pre- and post-launch WOM effects. Second, pre-launch forecasting approaches incorporating social media are mainly based on non-diffusion models. Fore example, regression models (e.g., Gopinath, 2013) generally start with one-step ahead forecasting, and updates as additional data become available. On the other hand, diffusion models based on the Bass model (Bass, 1969), which has a relatively high explanatory power despite its simple structure, are widely applied in pre-launch forecasting problems (e.g., Kim et al., 2013; Lee et al., 2014). However, they do not incorporate social media information into the diffusion model. Third, previous studies investigate the effect of social media only for single-generation products. Continuous innovation in high-technology industries made product lifecycles shorter and shorter (Shen & Willems, 2012). One notable aspect of short lifecycle products (e.g., iPhone) is that they release upgraded models of the product on a regular schedule, which makes it more challenging and important to forecast diffusion for the new generation.

Motivated by the discussion above, we develop a mulita-generational diffusion model with social media for forecasting the specific generational smart phone sales in even pre-launch period, and explore the roles of social media between the pre- and post-launch periods.

We analyze the product diffusion of the iPhone series (iPhone 5–iPhone 8/X) for empirical analysis. Our data include quarterly sales, pre- and post-launch WOM from social media, and price information for each generation. In addition to price, featured in the extended Bass models of Von Bertalanffy (1957), Mahajan and Muller (1981), Easingwood et al. (1983), Bewley and Fiebig (1988), and Jiang and Jain (2012), we construct the covariates by extracting the features from social media and incorporate them in the adoption rate function as the post-launch effect of social media. Additionally, previous diffusion models assume that key parameters of market potential and imitation rates are independent among generations when the innovator parameter is given. Our model structures generation-specific parameters to hierarchically depict the mechanism underlying parameter shifts between generations. Our structural model of parameter shift across generations facilitates forecasting sales of new-generation products before launch. The social media covariates are also incorporated in this hierarchical model for forecasting pre-launch sales.

Section 2 discusses the effects of social media on multi-generational diffusion in the literature. Section 3 explains our model. Section 4 presents the empirical results for successive generations of iPhones and shows that our model preforms better pre-launch forecasting by using social media covariates. Section 5 discusses the roles of social media on multi-generational diffusion in pre- and post-launch periods in the points of sales, leapfrogging and switching. The managerial implications are provided in sections 6 and we conclude in section 7.

2. Research Background

2.1 Feature Extraction on Social Media

WOM from social media generally requires preprocessing before being incorporated into the model as covariates. Besides volume, represented by the frequency of WOM, valence is also an important factor when explaining the demand for a product. The definition of valence differs in previous studies. Gophinath et al. (2014) defines “attribute-focused,” “emotion focused,” and “recommend-focused” as three distinct WOM valences, and Godes and Mayzlin (2004), Liu (2006), Rui and Whinston (2013), Hennig-Thurau et al. (2015), and Burmester et al. (2015) use the positivity of WOM as valence.

Valence is usually extracted from WOM by various methods, such as binary sentimental classification (Tirunillai & Tellis, 2014), conditional random field (Netzer et al., 2012), and latent Dirichlet allocation (LDA) (Blei et al., 2003). In this study, we use the LDA model since it has been successfully applied to many kinds of marketing problems. For example, Tirunillai and Tellis (2014) use the LDA model to incorporate consumer reviews into five sets of marketing data to extract dimensions (topics) from user-generated content for comparison among markets. They find that some topics resonate across multiple markets and others only in certain markets. Through sentimental analysis, they tag topics for better interpretation and show that multidimensional scaling via LDA captures the dynamics of brand positioning. Ansari et al. (2018) use supervised topic model to identify topics hidden in product reviews that reveal consumer preferences. They employ the variational Bayesian approach for fast and scalable inferences from big data.

As the most related to our study, Li and Terui (2018) incorporate social media effects into a diffusion model to discuss how they influence market potential and internal parameters. Analyzing text data from social media on a smart phone, they extract subjective and objective features by naïve Bayes and topic analysis, respectively, and then use them as covariates for time-varying market potential and imitation parameters. They show how social media affects the single-generation diffusion of smart phone and enhances forecasts.

These studies show that both interpretability and forecasting precision benefit from the latent topics of WOM in marketing.

2.2 Multi-generational Diffusion Model

Research on the diffusion across product generations adopts the framework of multi-generational diffusion models. The main differences among previous models are the assumptions about key parameters and the use of marketing mix variables. Among studies distinguished by their key parameters, Norton and Bass (1987) assume that market potential m for a generation of product depends on innovation and imitation parameters p and q are constant across generations. Mahajan and Muller (1996), Jun and Park (1999), Kim et al. (2000), Danaher et al. (2001), and Jiang and Jain (2012) assume the constancy of p across generations, generation-specific market potential, and imitation parameters. Jiang (2010) and Guo and Chen (2018) assume that all parameters vary across generations. Shi et al. (2014) show that assuming the heterogeneity of the diffusion parameters for the multi-generational diffusion model performs better empirically than the model with a homogeneous structure, even if only with total sales data.

Among diffusion studies distinguished by marketing mix variables, Robinson and Lakhani (1975) and Bass (1980) introduce the price effect term into the adoption rate function. These studies show that price information improves model performance. Horsky and Simon (1983) incorporate an advertising variable directly into sales rather than the adoption rate function, and Horsky (1990) incorporates advertising and price information jointly. The adoption rate function in Bass et al. (1994) includes two marketing mix variables: rates of change in prices and advertising expenditures relative to expenditures at product launch. Jiang and Jain (2012) extend this model to a multi-generational diffusion model.

Table 1 shows a summary of previous literature compared to our framework. Comparing the related approaches, our study has the following distinguished features. (i) Previous studies highlight the importance of pre-launch forecasting for diffusion models (e.g., Bass et al., 2001; Lee et al., 2014). In the literature, some studies (e.g., Dellarocas et al., 2007) forecast later-week revenues quite soon after launch for single-generation products by using data from reviews as early forecasting. However, no study has used social media for pre-launch forecasting while utilizing the diffusion model. Our model employs social media for pre-launch forecasting of the multi-generational smart phone, as well as post launch forecasting. (ii) Social media was used by Li and Terui (2018) as a covariate in a single-generation diffusion model. Their empirical results show that social media topics had a significant impact on the diffusion process, while the sentiment of WOM was not significant in their proposed model. Despite multi-generational products becoming more common and their lifecycles becoming shorter in recent years (e.g., Huang & Tzeng, 2008), social media data have never been applied to multi-generational product diffusion models. In the existing literature, only covariates of price and promotion are used (Jiang & Jain, 2012; Guo & Chen, 2018). (iii) Research considering both pre- and post-launch effects of social media have been conducted using non-diffusion models.

Compared with non-diffusion models, the diffusion models provide more useful insight into real business implications in the following points: (i) Most non-diffusion models could perform better locally, i.e., forecast near-future sales, while the diffusion model is capable of forecasting globally, that is, depicting the entire product lifecycle. (ii) Besides direct effect of social media on sales, we can investigate the impact of social media on sales indirectly by way of potential market and internal parameters. (iii) The proposed model allows us to discuss the leapfrogging and switching in the process of multi-generational diffusion as is shown in section 5. Hence, estimating the diffusion model provides more insightful information with firms for cost management and marketing strategies by forecasting potential sales over long periods and key numbers of imitation rates, leapfrogger and swichter at respective generations by using proposed model.

Authors	Forecasting Model	Marketing Variable	Social Media Post-Launch	Social Media Pre-launch	Diffusion Model	Multi-Generation
Norton and Bass (1987)	Norton–Bass Model	No	No	No	**Yes**	**Yes**
Bass et al. (2001)	Bass Model	No	No	No	**Yes**	No
Trusov et al. (2009)	Vector Autoregression (VAR)	**Yes**	**Yes**	No	No	No
Jiang and Jain (2012)	Generalized Bass Model (GBM)	**Yes**	No	No	**Yes**	**Yes**
Gopinath et al. (2013)	Linear, log-log, semilog Regression	**Yes**	**Yes**	**Yes**	No	No
Lee et al (2014)	Ensemble Machine Learning Method	**Yes**	No	No	**Yes**	No
Gopinath et al. (2014)	Dynamic Hierarchical Linear Model	**Yes**	**Yes**	No	No	No
Shi et al. (2014)	Extension of Norton-Bass Model	No	No	No	**Yes**	**Yes**
Burmester et al. (2015)	Econometric Model	**Yes**	No	No	No	No
Marchand et al. (2016)	Econometric Model	**Yes**	**Yes**	**Yes**	No	No
Divakaran et al. (2017)	Structural Equation Model (SEM)	**Yes**	**Yes**	No	No	No
Li and Terui (2017)	Extension of Bass Model	**Yes**	**Yes**	No	**Yes**	No
Guo and Chen (2018)	Extension of Norton-Bass Model	**Yes**	No	No	**Yes**	**Yes**
Our study	Hierarchical GBM	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**

Table 1 Comparison of related studies

3. Models

Our mulita-generational diffusion model consists of three parts: (i) feature extraction of social media data, (ii) hierarchical model involving pre-launch social media covariates for prior structure of model parameters, and (iii) marketing mix incorporating post-launch social media and price to define the likelihood by adoption rate function. An overview of our framework is shown in Figure 1.

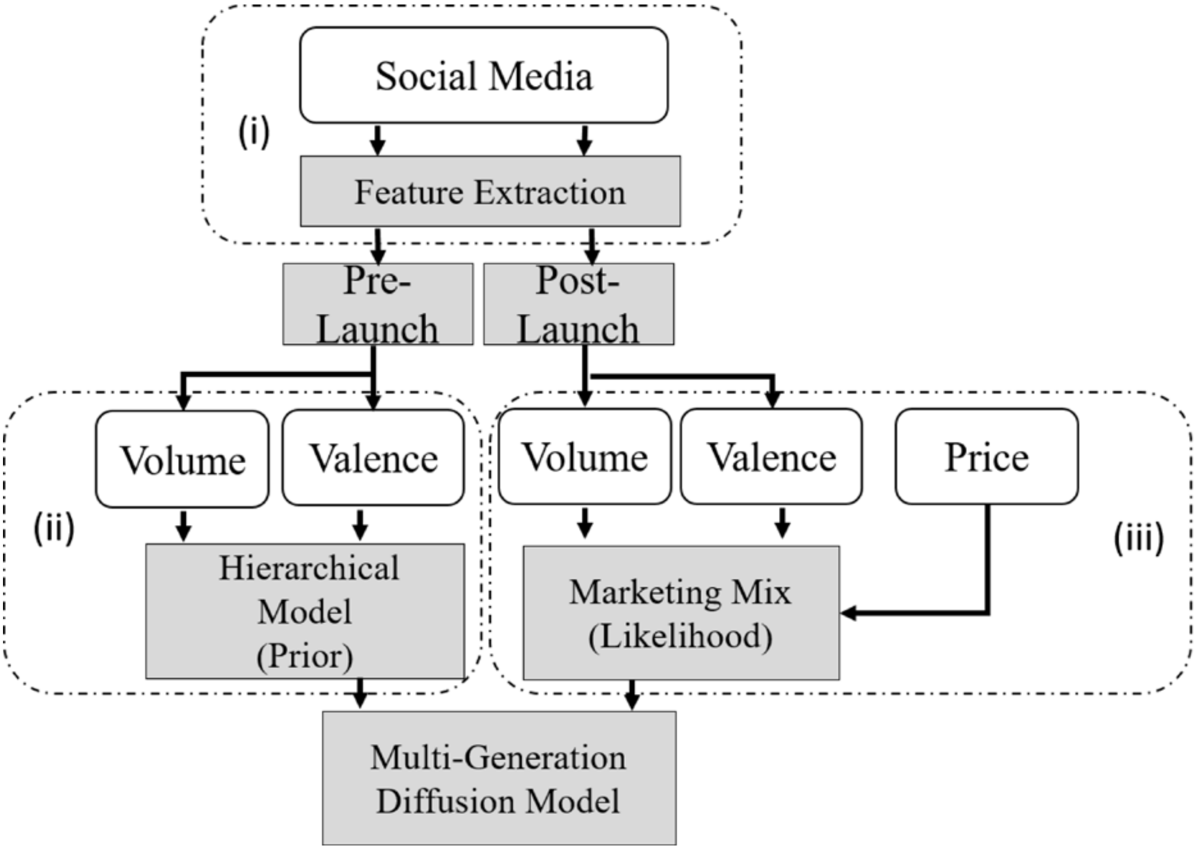

Figure 1 Framework of Model

3.1. Measures of WOM on Social Media

We use both volume and valence as covariates extracted from social media in the pre- and post-launch periods. Volume is defined by the frequency of WOM, and in the similar spirit of related study of Gopinath (2014), valence consists of the latent topics. Note that we define valence in this study not by the positivity of WOM but by the coverage share among social media topics. Our proposed model only considers objective social media topics for two reasons: (i) Li and Terui (2018) proposed the diffusion process of a single generation of the same product and social media site as in our study and their empirical results showed that subjective sentiment terms were not significant, while objective topics had significant effects. (ii) Responding to this result, we adopt the principle of parsimonious modeling and focus on the role of topics in this study.

The topic model (e.g., Blei et al., 2003) extracts the latent topics from WOMs, which distinguishes what the subject of WOM is, such as sports, music, or games. Compared to positivity, social media topics define a multi-dimensional feature vector for each instance of WOM. We apply the labeled dynamic topic model (LDTM), where we assume that the topic proportion θ_t follows a dynamic process according to its hyperparameter α_{θ_t} which follows $\alpha_{\theta_t} \sim N(\alpha_{\theta_{t-1}},\ \sigma_\theta^2)$ (see Blei & Lafferty, 2006), and incorporate seed words – the words that are fixed to specific topics for the dynamic topic model for better interpretability. This model is included in the class of semi-supervised topic models such as the labeled topic model (Daniel et al., 2009) and the seeded-LDA model (Lu et al., 2011), which show higher interpretability compared to the unsupervised topic model (Watanabe & Zhou, 2020).

We set four categories of valence for our proposed LDTM, that is, “attribute,” “emotion,” “recommend,” and “others,” and set seed words for the first three topics by following Gopinath et al. (2014). We describe further details of the seed words and topics in Section 4.1.

The goal of feature extraction is to obtain an N numbers of K-dimensional vector from WOM data by LDTM, where N is the number of WOM and K is the topic number. One example of feature extraction is shown in Figure 2.

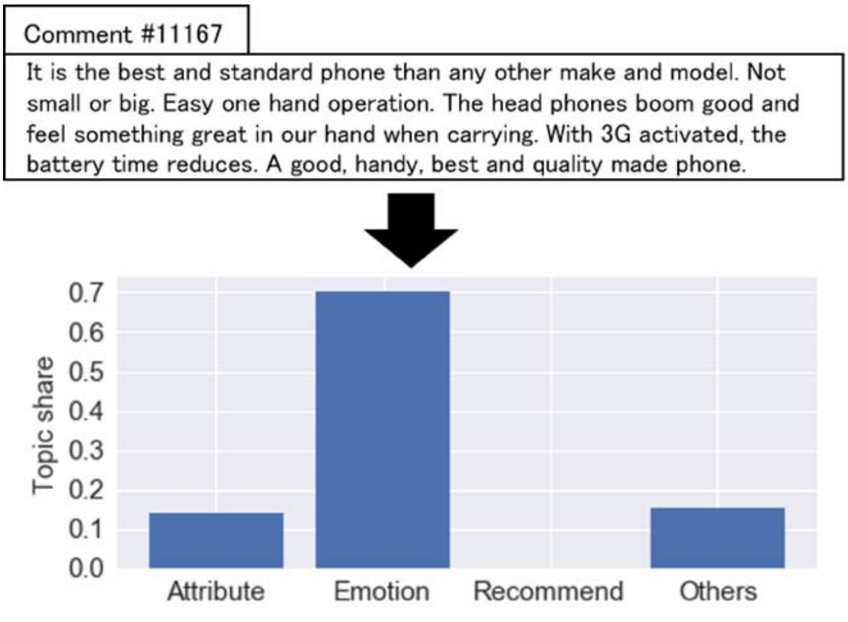

Figure 2 Feature Extraction from WOM

We first identify topics on social media and then incorporate them as covariates into our multi-generational diffusion model to reflect dynamic changes in consumers’ interests. The algorithm is shown in Appendix A.

3.2 Multi-Generational Diffusion Model with Social Media Covariates

We employ the diffusion model generalized by Jiang and Jain (2012), which assumes p is constant across generations, however, q and market potential m are generation-specific; we use the sales function in Srinivasan and Mason (1986) to define our multi-generational diffusion model.

First, we denote m_G, q_G as the market size, imitation rate for G-th generation, and $y_G(t)$ as the number of adaptors at time t with launch time $\tau_G > 0$, and then, following Jiang and Jain (2012), we define our model with additive noise for the adoption of each generation as follows:

Starting with the first generation launched at time 0, we have

$$\begin{cases} y_1(t) = m_1 f_1(t) + u_1(t), & t < \tau_2, \\ y_1(t) = m_1 f_1(t)\left(1 - F_2\left(t - \tau_2\right)\right) + u_1(t), & t \geq \tau_2, \end{cases}$$

$$\begin{cases} y_G(t) = I_G(t) + L_G(t) + S_G(t) + u_G(t), & \tau_G \leq t < \tau_{G+1},\ 1 < G \leq N, \\ \quad \equiv m_G f_G(t-\tau_G) F_G\left(t-\tau_G\right) + y_{G-1}(t) F_G\left(t-\tau_G\right) + Y_{G-1}(t) f_G(t-\tau_G) + u_G(t) & \\ y_G(t) = I_G(t) + L_G(t) + S_G(t) - L_{G+1}(t) + u_G(t), & t \geq \tau_{G+1},\ 1 < G \leq N, \\ \quad \equiv \left(m_G f_G(t-\tau_G) F_G\left(t-\tau_G\right) + y_{G-1}(t) F_G\left(t-\tau_G\right) + Y_{G-1}(t) f_G(t-\tau_G)\right) & \\ \quad\quad \times \left(1 - F_{G+1}\left(t - \tau_{G+1}\right)\right) + u_G(t), & \end{cases} \tag{1}$$

where $I_G(t)$ is the independent adoption of the G-th generation, $Y_G(t)$ denotes cumulative adoptions (sales) of the G-th generation at time t, $f_G(t) = F_G\left(t\right) - F_G\left(t-1\right)$ and $L_G(t)$ and $S_G(t)$ are adoptions from leapfrogging and switching from the previous generation, respectively.

In Eq. (1), the error term $u_G(t)$ is assumed to follow a normal distribution $u_G(t) \sim N\left(0, \sigma^2\right)$ independently across generations and time. We assume constant variance across generations because our data do not identify sales by generation. The adoption rate function $F_G\left(t\right)$ for the G-th generation product includes not only marketing mix according to Jiang and Jain (2012) but also social media effects:

$$F_G\left(t\right) = \frac{1 - \exp\left(-\left(p + q_G\right) X_G(t)\right)}{1 + \left(q_G / p\right) \exp\left(-\left(p + q_G\right) X_G(t)\right)}, \tag{2}$$

where the covariate $X_G(t)$ is formulated by the hierarchical model as follows:

$$X_G(t) = t + \alpha_G \log\left(P'_G(t)\right) + \boldsymbol{\beta_G}' \mathbf{log}\left(\boldsymbol{Topic}'_G(\mathbf{t-1})\right) + e_x(t) \tag{3}$$

In the above, $P'_G(t) = P_G(t)/P_G(0)$ and $\boldsymbol{Topic}'_G(\boldsymbol{t-1}) = \boldsymbol{Topic_G}(\boldsymbol{t-1})/\boldsymbol{Topic_G}(\boldsymbol{-1})$. $P_G(0)$ and $P_G(t)$ mean the price of G-th generation product at the launch τ_G and t period after launch respectively, and α_G is the coefficient on the price. $\boldsymbol{Topic}_G(t-1)$ is a $K+1$ dimensional column vector with elements of $Topic_{Gi}(t-1), i = 1, \ldots, K$ and $Freq_G(t-1)$, which means the share of topics in the *i-th* topic and the frequency of comments at *t-1* for generation G, respectively. $\boldsymbol{Topic}_G(-1)$ means the corresponding vector of one period before launch. **log(·)** denotes the operator of taking the log of each element. $\boldsymbol{\beta_G}$ is the $K+1$ dimensional corresponding coefficient vector. We assume that the error term $e_x(t)$ follows $e_x(t) \sim N(0, \sigma_x^2)$.

We use $Topic_{Gi}(t-1)$ rather than $Topic_{Gi}(t)$ to acknowledge the leading property of the social media effect, as it is typical in WOM research (Chintagunta et al., 2010; Godes & Mayzlin, 2004; Gopinath et al., 2014; Li & Terui, 2018).

3.3 Hierarchical Model for Generation-Specific Parameters with Social Media

We assume that diffusion in each generation persists from the previous generation. Specifically, generation-specific parameters are determined by those of previous generations. We first define $\boldsymbol{T_G}$ as the vector of share for each topic before the new generation is released. The k-th element, T_{Gk}, is defined as the share of topics for past ω days before generation G is released, i.e., $\sum_{t=(\tau_G-\omega)}^{\tau_G} Topic_{Gk}(t)/\omega$. We define $\omega = 90$ days according to the interval of sales data (quarterly) in our empirical application.

Then, we define the structural equations for the diffusion parameters (m_G, q_G), and the response parameters (α_G, β_G) in the adoption rate function, respectively, as

$$\begin{cases} m_G^* = \delta_{m0} + \delta_{m1} m_{G\text{-}1}^* + \boldsymbol{\delta}_{mT}{}'\boldsymbol{T}_G + \varepsilon_m \\ q_G^* = \delta_{q0} + \delta_{q1} q_{G-1}^* + \boldsymbol{\delta}_{qT}{}'\boldsymbol{T}_G + \varepsilon_q \end{cases}$$
$$\begin{cases} \alpha_G = \delta_{\alpha 0} + \boldsymbol{\delta}_{\alpha 1}{}'\alpha_{G-1} + \boldsymbol{\delta}_{\alpha T}{}'\boldsymbol{T}_G + \varepsilon_\alpha \\ \beta_G = \delta_{\beta 0} + \delta_{\beta 1}\beta_{G-1} + \boldsymbol{\delta}_{\beta}{}'\boldsymbol{T}_G + \varepsilon_\beta, \end{cases} \tag{4}$$

where $m_G^* = \log(m_G)$, $q_G^* = \log(q_G)$ and we assume $\varepsilon_i(t) \sim N(0, \sigma_i^2)$ for the *i-th* equation and independence across equations. Eq. (4) defines prior distributions for the parameter, and we constitute the posterior distributions when combined with the likelihood function derived by Eqs. (1) - (3).

We note that models by Jiang and Jain (2012) and Norton and Bass (1987) structurally do not allow forecasting of next-generation sales before launch as they assume no structure of connecting successive generations by their parameters. The hierarchical model in Eq. (4) accommodates parameter shifts from previous generations and social media effects on the parameters. The joint posterior density is given in the Appendix.

3.4 Predictive Density

In terms of Eqs. (1) and (2), if we simulate sales for the *G*-th generation product, we need at least one data point of that generation to estimate not only m_G and q_G even when p is assumed constant across generations, but also α_G and $\boldsymbol{\beta_G}$ in the adoption rate function.

Before the *G*-th generation launches, social media provides information about a new generation of the *G*-th product. Then, in terms of the hierarchical model in Eq. (4), implying prior information, we can forecast the adoption of new generations when parameters are given. After the *G*-th generation product launches, the marketing mix (Eq. (3)) enters the adoption rate function to continuously update the forecast with new information such as the latest WOM and price information.

The likelihood and prior functions using social media make it possible to forecast sales before and after the new generation's launch. Bayesian forecasts are formally constituted by the predictive density as

$$
\begin{aligned}
& p\left(y_G(t) \mid m_G, q_G, p, X_G(t), \boldsymbol{T}_G, \boldsymbol{Topic}_G'(t-1), P_G(t)\right) \\
& = \int p\left(y_G(t) \mid m_G, q_G, p, t\right) p\left(m_G, q_G, p \mid m_{G-1}, q_{G-1}, \boldsymbol{T}_G\right) dm_G dq_G dp \quad \text{if } t \le \tau_G \\
& = \int p\left(y_G(t) \mid m_G, q_G, p, X_G(t)\right) p\left(m_G, q_G, p \mid m_{G-1}, q_{G-1}, \boldsymbol{T}_G\right) \\
& \times p\left(X_G(t) \mid \alpha_G, \boldsymbol{\beta}_\mathbf{G}, t, \boldsymbol{Topic}_G'(t-1), P_G{}'(t)\right) \\
& \times p\left(\alpha_G, \boldsymbol{\beta}_\mathbf{G} \mid \alpha_{G-1}, \boldsymbol{\beta}_\mathbf{G-1}, \boldsymbol{T}_G\right) dX_G(t) dm_G dq_G d\alpha_G d\boldsymbol{\beta}_\mathbf{G} dp \qquad \text{if } t > \tau_G
\end{aligned} \tag{5}
$$

These integrations are numerically evaluated with Markov Chain Monte Carlo (MCMC) iterations, as were successfully applied to time series forecasting by Terui et al. (2010) and Terui and Ban (2014). The algorithm for Eq. (5) appears in Appendix B2.

4. Empirical Results

4.1 Data

We use five generations of iPhone products, iPhones 5 through 7, as training data, that is, $G1$ (iPhone 5), $G2$(iPhone 5s, 5c), $G3$ (iPhone 6, 6 Plus), $G4$ (iPhone 6s, 6s Plus), and $G5$ (iPhone 7, 7 Plus). The last generation $G6$ (iPhone 8, 8 Plus, X) is reserved for the test data. All data between 2012Q4 and 2018Q3 were obtained from Statista (www.statista.com), the statistics portal for market data.

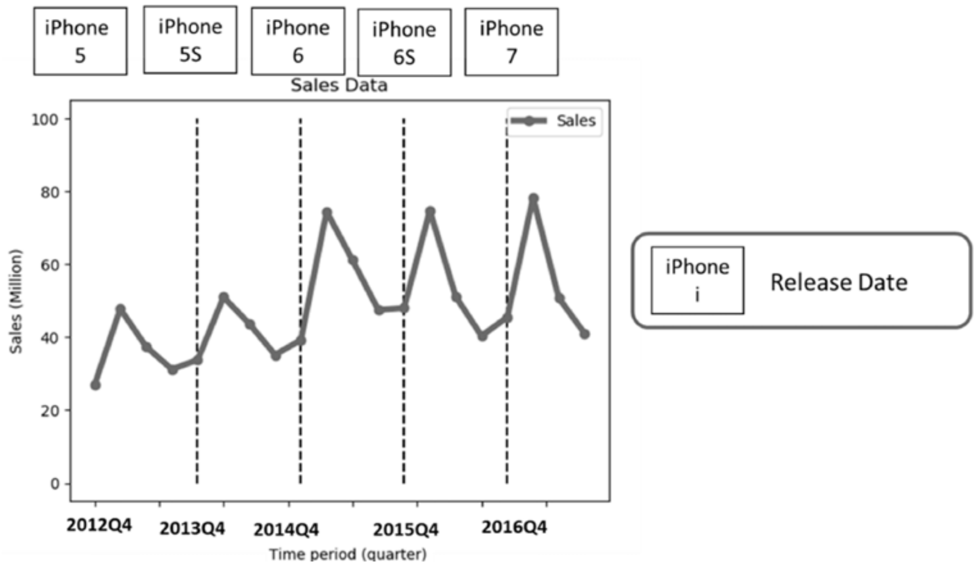

Figure 4 iPhone Sales

Sales data are comprised of the total iPhone sales shown in Figure 4. As there are no unit sales data for each generation, it is harder to optimize all parameters for each generation only based on total sales. However, a related study by Shi et al. (2014) showed that the multi-generation diffusion model works properly for multiple products using only total sales, and the diffusion model with generation-specific heterogeneous diffusion parameters performs better than the model with homogeneity structure, especially for iPhone series.

Historical prices on amazon.com are used as price data in this study. We collect the price at an interval of quarterly to match the sales data by using the price tracker application "Keepa," which provides historical prices for various products on amazon.com. As there are many variations even for one generation (i.e., RAM, color), we trace two types of iPhone that had the highest and lowest prices at release and take mean value as the price of generation G for each time period. The descriptive statistics of price data is shown in Table 2.

unit: US dollar

	G1	G2	G3	G4	G5
mean	533.0	595.0	763.3	835.0	1015.0
std	261.0	223.3	137.0	109.4	115.0
min	280.0	320.0	580.0	740.0	900.0
25%	300.0	397.5	697.5	750.0	937.5
50%	450.0	550.0	755.0	805.0	1000.0
75%	675.0	712.5	802.5	885.0	1077.5
max	1200.0	1050.0	1030.0	1050.0	1160.0

Table 2 Descriptive statistics

Text data were acquired from gsmarena (http://www.gsmarena.com/). This is a well-known bulletin board system (BBS) where users worldwide comment on mobile phones. Although it has a smaller volume of available WOM in comparison to popular social media (i.e., Facebook and Twitter), we employ this site for three reasons. (i) Most WOM in the BBS is generated from aficionados for a specific product, and we expect higher quality and relevancy from this BBS. (ii) Additionally, "gsmarena" has the advantage that a specific generation of iPhone is identifiable for each comment. In contrast, for most popular social media such as Twitter, it is hard to classify each comment according to the specific generation properly as all the WOM for iPhone products are mixed. (iii) Finally, it has a relatively strong influence on potential customers among the smartphone forums (e.g., over 60 million hits for the iPhone 5s topic, over 75 million hits for iPhone 6 topic).

Through conventional natural language preprocessing procedures, punctuations, stop words, and low frequency words (used less than ten times) were removed. After preprocessing, 56,541 comments in total remained for all generations.

4.2 Dynamic Topics

As mentioned in Section 3.1, we assume four topics and define seed words for three topics: "camera," "screen," "battery," and "ram" for topic 1 represent the "attribute" topic; "bad," "good," "great," "worse," "worst," and "best" for topic 2, the "emotion" topic; "buy," "bought," "advice," "recommend," and "should" for topic 3, the "recommend" topic; and no seed words for the last topic, "others." As there is no common standard for choosing the seed words, we employ a weakly supervised topic modeling approach with a smaller number of seed words as minimal prior knowledge, which empirically showed better performance than the standard supervised approach by Lu et al. (2011).

LDTM allows us to detect changes in each topic, implying potential changes in customer interest. Figure 5 shows the volume and composition of dynamic topics through generations. The left panel (a) shows the relationship between sales and volume of WOM. Apparently, in most cases, the volume of WOM becomes the leading indicator of sales. Note that the dashed line represents the release date of the new generation. The right panel (b) shows the compositions of the four topics through all generations, which are generated from the LDTM. The percentage in the figure represents the average share of the four topics for each generation. We observe that the "emotion" topic has the largest average share, and the composition of topics changes through generations. Specifically, "attribute" increases the most, "emotion" follows, "recommend" is relatively stable, and "others" continuously decreases. This may imply that social media users have placed more importance on the three labeled topics for new iPhone products in recent years.

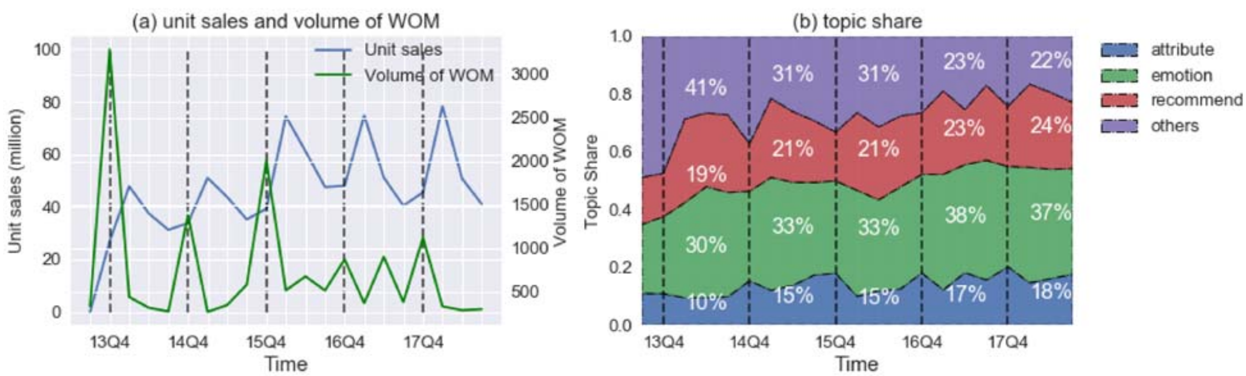

Figure 5 Diffusion and Dynamic Topics

4.3. Model Comparison

We now compare seven models according to the combination of covariates, generation-specific heterogeneity of parameters, and the hierarchical structure of parameters connecting current and previous generations.

Models 1–3

Model 1 is the benchmark multi-generational diffusion model by Norton-Bass (1987). Model 2 uses the specification by Jiang and Jain (2012), which incorporates the price variable. Model 3 adds the topic variables from LDTM to Model 2. Models 1–3 lack the structure to connect generations except as an innovator parameter p (i.e., no transgenerational memory). Thus, we call them zero-order models.

Model 4–Model 5

Models 4–5 include the structure for parameter shifts to the next generation as hierarchical models. Model 4 only has the parameters of previous generations. Model 5 includes additional topic variables. These models can predict the diffusion of the next generation even before launch. These are first-order models.

Model 6–Model 7

Models 6 and 7 are also first-order models. Compared with Models 4 and 5, they have homogeneous coefficients for marketing mix variables between generations, that is, $\alpha_G = \alpha$ and $\boldsymbol{\beta_G} = \boldsymbol{\beta}$. We assume that marketing variables are homogeneous for all generations to achieve parsimony.

Model 8

Model 8 is the extended model of Model 7 by incorporating sentiment analysis. We add two sentiment variables (share of positive comments and negative comments) in the adoption function and hierarchical structure in Eqs. (3) and (4). We employ a pre-trained sentiment classifier model - VADER (Valence Aware Dictionary and sEntiment Reasoner), which uses rule-based values tuned to sentiments from social media by Hutto and Gilbert (2014).

Alternative Model: Vector Autoregression (VAR)

Our model forecasts diffusion for pre-launched products with only aggregated sales data; there are very limited alternative models in the literature. We choose a vector autoregressive (VAR) model by Trusov et al. (2009), which shows that the VAR performs well with the WOM data by accounting for dynamic responses and interactions between marketing variables, topic variables, and sales. This approach forecasts the sales of future generations. To be consistent with a multi-generational structure, we use total sales, total topic shares, and volumes as model variables.

We compare the models using three measures: log of marginal likelihood (LMD), deviance information criteria (DIC), and root mean square error (RMSE) of forecasts for training data ($G1$–$G5$), and test data ($G6$) (pre-launch forecasting). Note that we only use social media before the product is released for pre-launch forecasting to evaluate whether social media is informative for forecasting. We evaluate the out-of-sample fit by forecasting both one step (one quarter) and four steps ahead (one year). We confirmed convergence for all models using Geweke's test (Geweke, 1992) at 95% significance.

The results are shown in Table 3. Zero-order models (Models 1–3) cannot forecast the sales of test data ($G6$) because of lacking structure to accommodate the shift of (m_G, q_G). We evaluate all first-order models by RMSE (Train), log of marginal likelihood, DIC, and RMSE (Test).

	Zeroth-Order			First-Order					
Model	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	VAR
RMSE(Train)	6.980	3.012	2.248	2.358	2.218	3.735	2.612	2.744	6.385
RMSE(Test)	-	-	-	19.425	19.870	15.861	12.114	12.743	21.721
log(ml)	-509.715	-128.642	-114.943	-119.528	-113.476	-179.266	-112.416	-119.458	-307.290
DIC	1027.614	308.255	348.142	362.415	337.978	419.303	293.935	307.567	401.595

Table 3 Model evaluations

Model 5 shows the best in-sample performance as indicated by RMSE, implying that the marketing mix with social media improves forecasting precision, while Model 7 performs best according to log (ml), RMSE for test data, and DIC. Focusing on RMSE for test data, we find that without social media topics, Model 4 and Model 6 generate poor one-step forecasts, as the RMSEs are 21.705 and 114.849 for these models. The proposed model with social media topics in the hierarchical structure performs best for the test data, and this implies that it is not necessarily overfitting, and the social media topics are informative beyond one-step-ahead forecasting. Last, the performance of Model 8 implies that the sentiment analysis does not improve the precision of the diffusion model in our empirical case.

According to the log(ml), the RMSE of test data, and DIC criteria, we chose Model 7 and examine the results of estimation in the following sections.

Robustness Check

We examined the robustness of our best proposed model, Model 7, by comparing with two alternative models. *Alt. Model 1* replaces the LDTM with unlabeled DTM by Blei and Lafferty (2006), when fixing all other parameters, such as topic number, as the same value. *Alt. Model 2* includes two-period lag terms for X_G in the adoption function in Eq. (4), as follows:

$$X_G(t) = t + \alpha_G \log(P'_G(t)) + \boldsymbol{\beta_{1G}}' \mathbf{log}(\boldsymbol{Topic'_G}(\boldsymbol{t-1})) + \boldsymbol{\beta_{2G}}' \mathbf{log}(\boldsymbol{Topic'_G}(\boldsymbol{t-2})) + e_x(t). \quad (6)$$

The results of alternative models are shown in Table 4. We observe two findings. (i) Our proposed model performs better than *Alt. Model 1* for both in-sample and out-of-sample evaluations. In particular, the results of test RMSE (1.394 compared to 7.283) show that the diffusion model with LDTM greatly improves in short-term forecasting, and this implies that our labeled topics are more informative for the diffusion process, compared to the auto-generated, unlabeled topics. (ii) The performance of *Alt. Model 2* is not notably different from the proposed model even when it contains the two-period lag terms, and we employ the simpler proposed model following the principle of parsimonious modeling.

Model		Proposed Model	Alt. Model 1	Alt. Model 2
log(ml)		-112.416	-128.642	-114.048
DIC		293.935	308.255	289.233
RMSE(Train)		2.612	3.042	2.739
RMSE(Test)	1-step ahead	1.394	7.283	2.311
	4-step ahead	12.114	14.180	10.281

Table 4 Robustness Check

4.4. Parameter Estimates

Parameter estimates of Model 7 are shown in Table 5, where number indicates posterior means and the 95% highest posterior density (HPD) interval.

	m_G		p		q_G	
	Mean	*95% HPD*	*Mean*	*95% HPD*	*Mean*	*95% HPD*
G1	164.575	[161.876, 166.218]			1.031	[0.981, 1.080]
G2	4.816	[2.810, 6.605]			1.024	[0.978, 1.072]
G3	66.867	[63.785, 70.108]	0.072	[0.067, 0.081]	1.102	[1.061, 1.143]
G4	0.703	[0.365, 0.948]			1.131	[1.079, 1.191]
G5	3.329	[2.118, 4.631]			1.140	[1.095, 1.202]

Table 5 Parameter Estimates

The columns show that estimates of market size m_G, imitator parameter q_G, the innovator parameter p are constant across generations for the G-th generation's diffusion. Estimates of m_G imply that the market size estimate (164.5) for $G1$ (iPhone 5) is potentially much higher than any other generation. The first generation faces an original higher potential market, and $G2$'s (iPhone 5s) market is estimated much smaller than $G1$. Except for $G1$, the market size of $G3$ (iPhone 6) has the biggest value (66.867) among all the generations, while $G4$ (iPhone 6s) has the smallest market size estimate (0.703). The results of m_G shows that the potential market size increases significantly only in $G3$ (iPhone 6).

A previous empirical study on the diffusion for a variety of products indicates that the average value of the innovation parameter p has been found to be 0.03, and often less than 0.01 as shown by Mahajan et al. (1995). In contrast, the higher estimated value (0.072) in this study suggests that iPhone products are expected to have more innovators than other products.

Estimates for the imitation parameter q_G increase slowly across all generations except $G2$. This finding suggests that consumers tend to become imitators more as generations proceed, and they will use WOM reviews before buying new-generation products.

Table 6 provides the estimates of coefficient parameters for the hierarchical diffusion model and hierarchical adoption rate function in Eqs. (4) and (3), respectively. Note that the coefficients for the post-launch response parameters are homogeneous for Model 7.

	Hierarchical structure						Adoption rate		
	Param.	*Mean*	*95% HPD*	*Param.*	*Mean*	*95% HPD*	*Param.*	*Mean*	*95% HPD*
price	-	-	-	-	-	-	$\hat{\alpha}_G$	-0.201	[-0.430, 0.041]
attribute	$\hat{\boldsymbol{\delta}}_{mT}$	**0.895**	[0.514, 1.375]	$\hat{\boldsymbol{\delta}}_{qT}$	0.256	[-0.572, 0.901]	$\hat{\boldsymbol{\beta}}_G$	**0.463**	[0.264, 0.684]
emotion		**1.948**	[0.328, 3.528]		-0.024	[-1.619, 1.443]		**-0.442**	[-0.765, -0.118]
recommend		**-2.871**	[-4.160, -1.871]		-0.086	[-1.043, 0.952]		**0.180**	[0.024, 0.346]
volume		**3.095**	[1.717, 4.632]		-0.104	[-1.319, 1.393]		**0.300**	[0.103, 0.484]
intercept	$\hat{\delta}_{m0}$	**1.257**	[0.044, 2.538]	$\hat{\delta}_{q0}$	0.027	[-1.200, 1.453]		-	-
lag	$\hat{\delta}_{m1}$	-0.095	[-0.235, 0.046]	$\hat{\delta}_{q1}$	0.006	[-1.972, 1.741]		-	-

Table 6 Hierarchical Parameter Estimates

Hierarchical Diffusion Parameters

For the potential market size parameter m_G, we find that topics "attribute" and "emotion" are positively associated with $m_G{}^*$ with estimates of 0.895 and 1.948, respectively. The estimate of topic "recommend" is -2.871, implying that this topic is negatively associated with $m_G{}^*$. The studies by Kopalle and Lehmann

(2006), Joshi and Hanssens (2009), and Kim et al. (2017) provide empirical evidence that excessive promotion during the pre-launch period can have a negative effect. Our result might correspond to their result for social media in that excessive recommendation by unexperienced users can have a negative impact on potential sales during the pre-launch period. In addition, the estimate of the coefficient for volume is 3.095, meaning that the volume of WOM is positively associated with sales. There is a constant increment of the baseline for market size, as the estimate (1.257) for the intercept term δ_{m0}. Finally, the estimated coefficient of δ_{m1} is not significant in the sense of a 95% HPD interval. This means that the incremental market size m_G is independent of m_{G-1}, that is, a unique market size for a specific generation. Our result supports the specification of previous studies (e.g., Norton & Bass 1987, Jiang & Jain 2012). On the other hand, none of the estimates on the imitation parameter are significant. This implies that the imitation parameter, recognized as the post-launch WOM effect in the literature including the above studies, is highly unpredictable during the pre-launch stage.

Adoption Rate Parameters

The 95% HPD region indicates that every covariate constructed by volume and valence is significant while the price covariate is not significant. The estimate for topic “attribute” is 0.463, with a positive influence on market size. The topic “emotion” is negatively associated with sales since its estimate is -0.442. Topic “recommend” (0.180) and volume (0.300) have significant positive impacts on sales.

As for the topic “attribute,” the empirical results show that this topic is positively related to sales, which implies that the more users of BBS focus on the “attribute,” the more potential sales can be expected. On the other hand, the result of the topic “emotion” implies that consumers may tend to have a positive attitude toward the new generation of iPhones; however, it generally turns negative after the product is released. This means that users may have positive expectations for the new generation before it is released and disappointed after launch. The results are opposite for the topic “recommend”—only post-launch recommendations are positively associated with sales, and this implies that excessive promotion or recommendation in the pre-launch period may have a negative influence on sales. In contrast, potential customers tend to trust recommendations by experienced users.

As is shown above, we find the different roles of social media between pre- and post-launch periods.

5. Pre- and Post-launch Effects of Social Media

5.1 Social Media Effects on Sales

We showed that volume (frequency) and valence (topics) have a significant influence on diffusion in our empirical results. Next, we investigate how sales are influenced by the topics for each generation by decomposing the sales. We denote the estimated sales in time period *t* as $\hat{y}_G(t)$, then we decompose the estimated sales as

$$\hat{y}_G(t) = \sum_{k=1}^{K} \hat{y}_G^{(k)}(t) + \hat{y}_G^{(others)}(t), \tag{7}$$

where $\hat{y}_G^{(k)}$(k) is the estimated sales influenced by topic *k* in time period *t*, and $\hat{y}_G^{(others)}$ is the sales when excluding the effect of social media. Then, we can further decompose $\hat{y}_G^{(k)}(t)$ as

$$\hat{y}_G^{(k)}(t) = \hat{y}_G^{(k,pre)}(t) + \hat{y}_G^{(k,post)}(t), \tag{8}$$

where $\hat{y}_G^{(k,pre)}(t)$ and $\hat{y}_G^{(k,post)}(t)$ represent the estimated sales in time period *t*, which are influenced by the pre- and post-launch effects of topic *k*, respectively.

First, we calculate $\hat{y}_G^{(k,pre)}(t)$ for the pre-launch effect by controlling the covariates in the hierarchical structure in Eq (4). We denote $\widehat{m}_G^{-k}$ as the estimated potential market size when excluding the covariate term of topic *k* for generation *G*. We note that $\hat{q}_G$ is not considered since the coefficient estimates are not significant for the pre-launch effect. Then we define the pre-launch effect as

$$\hat{y}_G^{(k,pre)}(t) = \mathrm{g}\left(\widehat{m}_G, \hat{X}_G(t)\right) - \mathrm{g}\left(\widehat{m}_G^{-k}, \hat{X}_G(t)\right), \tag{9}$$

where g(·) reflects the comprehensive sales function described in Eq. (1). Next, considering the post-launch effect, topic *k* affects the sales of generation *G* via $X_G(t)$ in Eq. (3). Denote $\hat{X}_G^{-k}(t)$ by the covariate when excluding the covariate term of topic *k*. The post-launch effect can be similarly defined as

$$\hat{y}_G^{(k,post)}(t) = \mathrm{g}\left(\widehat{m}_G, \hat{X}_G(t)\right) - \mathrm{g}\left(\widehat{m}_G, \hat{X}_G^{-k}(t)\right). \tag{10}$$

Note that both $\hat{y}_G^{(k,pre)}(t)$ and $\hat{y}_G^{(k,post)}(t)$ can be negative, as the influence of the topic on sales is not always positive.

Figure 6 shows the differences in sales for each generation when excluding specific topics.

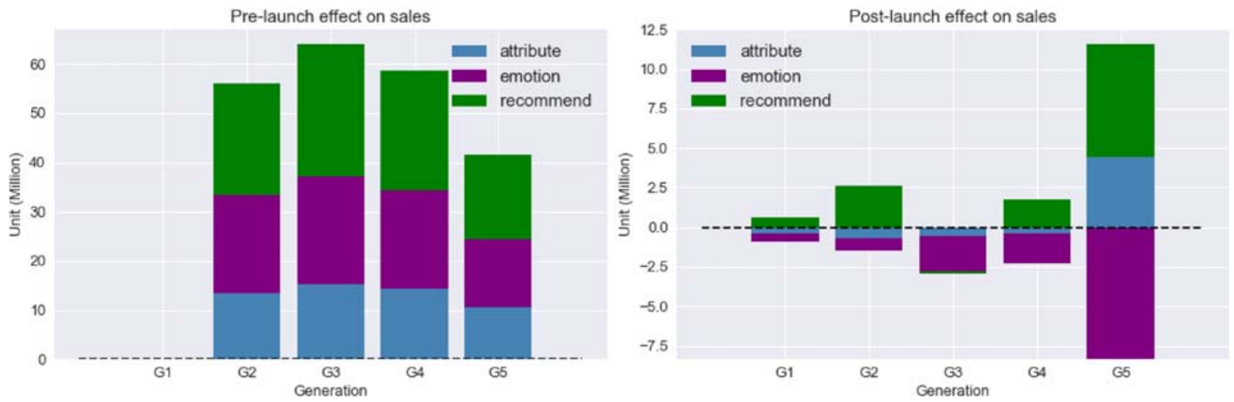

Figure 6 Pre- and Post-launch Effect on Sales

For the expected sales shift caused by the pre-launch effect shown in the left panel, we notice that all the topics accelerate sales and have the highest positive influence for the 3rd generation (iPhone 6). This implies that consumers may generally have positive expectations for the new generation, however, after iPhone 6 was launched, their expectations started to decline. Furthermore, we find that the topic "recommend" has the greatest influence among all the topics and throughout the generations.

On the other hand, the result for the post-launch effect shows a different impact – the topic "recommend" generally accelerates sales, while the topic "emotion" decelerates sales. The topic "attribute" rarely influences sales for the first four generations, and its influence increases in the last generation (iPhone 7).

5.2 Leapfrogging and Switching

Next, we detect leapfrogging and switching to a later generation induced by social media. Following Jiang and Jain (2012), we calculate them as

$$\begin{cases} L_G(t) = y_{G-1}(t)F_G\left(t-\tau_G\right)(1-F_{G+1}(t-\tau_{G+1})), & 1<G<5 \\ L_G(t) = y_{G-1}(t)F_G\left(t-\tau_G\right), & G=5 \end{cases} \tag{11}$$

$$\begin{cases} S_G(t)=Y_{G-1}(t)f_G(t-\tau_G)(1-F_{G+1}(t-\tau_{G+1})), & 1<G<5\ , \\ S_G(t)=Y_{G-1}(t)f_G(t-\tau_G), & G=5 \end{cases} \tag{12}$$

where $L_G(t)$ and $S_G(t)$ are, respectively, denoted as the number of leapfrogs and switches from the previous generation in time period *t*. $L_G(t)$ is induced by the remaining fraction of the previous generation, and $S_G(t)$ is affected by the installed fraction of the previous generation. Similar to Eq. (8), we calculate $\hat{L}_G^{(k,pre)}$, $\hat{L}_G^{(k,post)}$, $\hat{S}_G^{(k,pre)}$, $\hat{S}_G^{(k,post)}$, which are the estimated numbers of leapfrogs and switches induced by the k-th topic in social media for the pre- and post-launch periods, respectively. To measure the impact of topics on leapfrogging and switching for each generation, we calculate the share of leapfrog influenced by each topic, for example, the share of leapfrog $\hat{L}_G^{(k,pre)}$ calculated by $\hat{L}_G^{(k,pre)}/\hat{L}_G$, where $\hat{L}_G$ is the estimated total number of leapfrog for G.

The results of the pre- and post-launch impact on leapfrogging are shown in Figure 7.

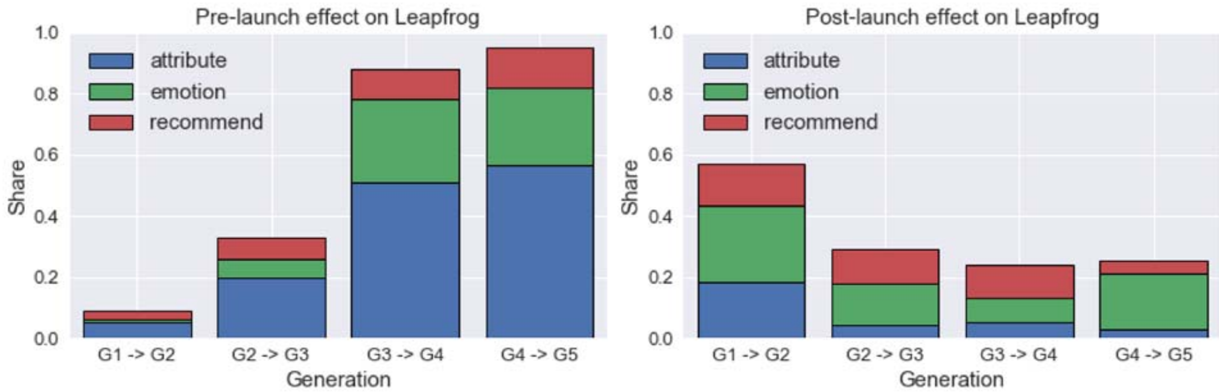

Figure 7 Pre- and Post-launch Effects on Leapfrogging

We observe that, (i) the topic "attribute" has the greatest impact on leapfrog among topics for the pre-launch period, and its share increases through the generations, implying that more consumers decide leapfrogging to the latest generation according to the pre-launch social media topic "attribute"; (ii) the post-launch impact of "attribute" on leapfrogging is relatively small, implying that the topic "attribute" may no longer affect leapfrogging after the product released; (iii) the shares of three topics keep increasing in the pre-launch period while they tend to decrease in the post-launch period, implying that the role of pre-launch information has become more important than the post-launch for leapfrogging.

Next, we show the pre- and post-launch effects of topics on switch in Figure 8.

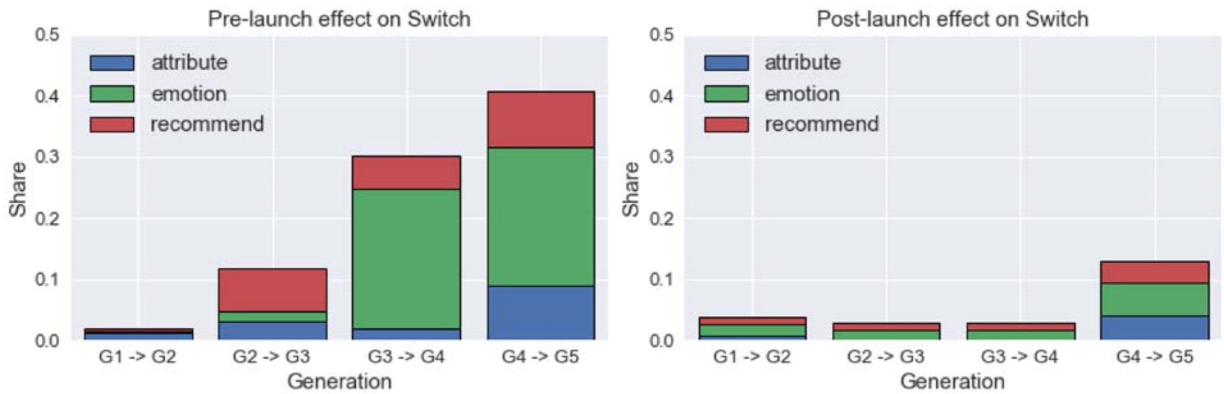

Figure 8 Pre- and Post-launch Effect on Switch

First, we find that the pre-launch effects increase for all topics as G proceeds. Specifically, the impact of the topic "emotion" increases rapidly from $G3$ (iPhone 6). On the other hand, the post-launch effect is relatively small for the switch effect – the total impact of topics on switch is less than 15% for all generations. This implies that switchers are mainly influenced by the pre-launch information related to the topic "emotion."

Through the results of sections 5.1 and 5.2 above, we find that the topic "recommend" affects sales the most, while the topic "attribute" has the biggest impact on leapfrogging, and the topic "emotion" has the biggest influence on switching, which shows that different topics influence diffusion in different ways. Furthermore, our results also show that the topic "attribute" generally has a strong impact on leapfroggers, while it has a weak impact on switchers. Goldenberg and Oreg (2007) mentioned that leapfroggers skip generations mainly to reach the latest technology. Our results are consistent with their statement in the sense that the topic "attribute" is the key factor for leapfroggers to adopt the new generation. In contrast, switchers who already owned the previous generation do not care about this topic.

6. Managerial Implications

Our framework with a hierarchical diffusion process and dynamic topics extracted from social media provides many useful implications in real business situations.

The first is pre-launch forecasting for a new generation. Short-life cycle products such as smartphones face the problem of predicting demand. However, as the possibility of refinement is quite limited after the product launch, managers benefit from better pre-launch forecasting (Divakaran et al., 2017). Using estimated hierarchical structure and diffusion model parameter estimates, we can forecast sales of new generations before launch. In the empirical application, we examined the pre-launch forecast of the new-generation $G6$ and compared it with actual sales from 2017Q4 to 2018Q3. The RMSE for the hold-out sample is described in section 4.3, where the proposed model performs better than the comparative models. In addition, Figure 9 indicates the performance of the forecast with a box plot diagram by comparing actual observations from one through four steps ahead. The forecasting results show that all predictions except the third step are relatively close to the actual observation of the sales unit, which implies that social media is useful and reliable for pre-launch forecasting. Managers can form expectations of their new-generation sales based on prior information from social media. Furthermore, the forecasting can be continuously updated when the sales, price, and WOM data of the next period are available during the post-launch period, as shown in Figure 10.

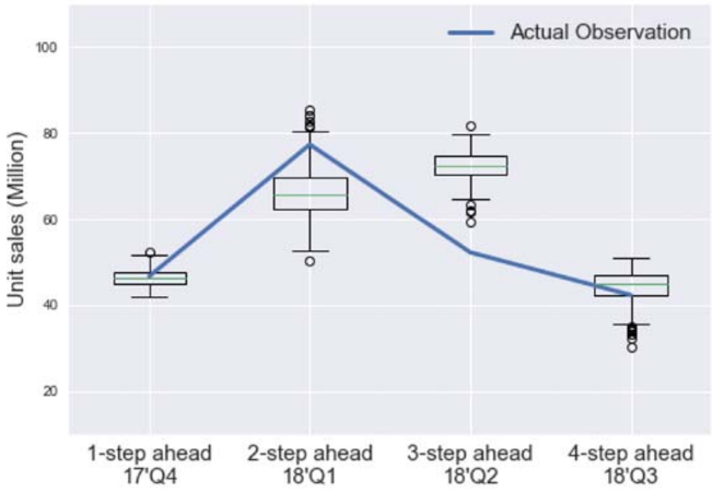

Figure 9 Pre-launch Forecasting of G6 (iPhone 8/X)

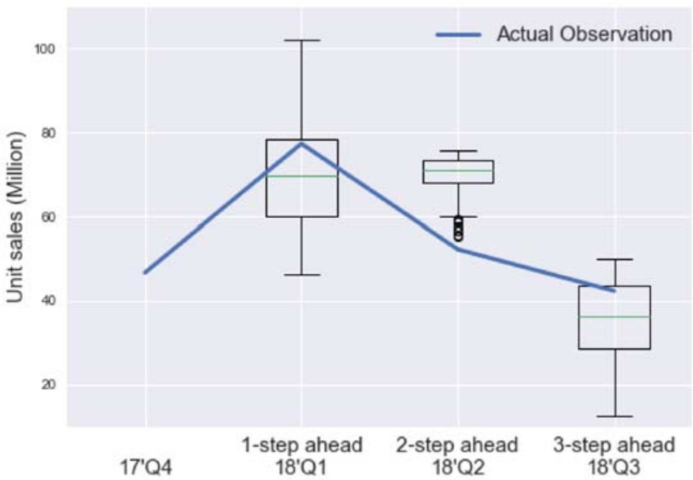

Figure 10 Post-launch Forecasting of G6 (iPhone 8/X)

Second, our approach provides useful insights into understanding the role of social media during the pre- and post-launch periods. While many studies emphasize the importance of marketing promotion in the pre-launch period, such as marketing campaigns (Elberse & Anand, 2007) and WOM advertising (Kim and

Hanssens, 2017), our empirical results show that post-launch WOM also significantly affects the diffusion process. In particular, our implication indicates that the impact of post-launch social media on sales keeps increasing throughout generations, implying that firms may be beneficial by investing resources in post-launch WOM advertising.

Third, contrary to previous studies that use positivity for the valence of WOM, we employ topic proportion as the valence of WOM and have shown that valence has a significant impact on diffusion. In the last section, we found that each topic affects sales, leapfrog, and switch in different ways in the in pre-launch and post-launch periods. Thus, our approach helps managers to not only quantify the impact of each topic on product sales, but also to optimize WOM advertising by focusing on the topics that impact the target generation.

7. Conclusion

This study proposed a multi-generational diffusion model with social media information, where we set a hierarchical structure connecting diffusion parameters of successive generations. We also proposed a labeled dynamic topic model (LDTM) to extract dynamic features hidden in social media.

Unlike previous multi-generational diffusion models, our model makes it possible to forecast sales of new-generation products before launch. Our empirical result shows that the performance of high precision and useful insight for firms on their management over product lifecycle.

The post-launch effects of social media on sales were directly measured in the adoption rate functions, and the pre-launch effects were evaluated indirectly by shifts in market size and imitation parameters of the multi-generational diffusion model in the hierarchical structure. The empirical results show that social media has a significant positive impact on sales in the pre-launch period. Social media effects decrease after launch but can gradually increase with different roles, including becoming a negative influence. We conducted a similar discussion on leapfrogging and switching behaviors.

Our study highlights the significance of the valences of WOM represented by latent topics, and shows that the topics play different roles in the diffusion process. Furthermore, we seek the contribution ratio of the topics in the diffusion process by decomposing the sales. For example, the topics have the biggest impact on leapfrog in the pre-launch period, while the topics have little influence on switching in the post-launch period.

Further issues remain. First, as the upgrade of the iPhone product follows a cyclical pattern, it is challenging to investigate the performance of the proposed model in more “irregular” markets. The treatment of possible feedback from sales to social media is another challenging problem. Second, our proposed model only considered objective social media topics as valence. It can be further extended by incorporating the positivity of topics, such as sentiment LDA (Li et al., 2010). Third, some explanations for the empirical results such as the role of topic “emotion” in pre- and post-launch period is difficult. As there is no relative literatures focus on the topic share in the diffusion process, the role of topic deserves more research attention in the future. We leave these issues for future research.

References

Ansari A., Li Y. and Zhang J.Z. (2018), “Probabilistic Topic Model for Hybrid Recommender Systems: A Stochastic Variational Bayesian Approach,” *Marketing Science*, 37(6), 987-1008.

Bass F.M. (1969), “A New Product Growth for Model Consumer Durables,” *Management Science,* 15(5), 215-227.

Bass F.M. (1980), “The Relationship between Diffusion Rates, Experience Curves, and Demand Elasticities for Consumer Durable Technological Innovations,” *The Journal of Business*, 53(3), 51-67.

Bass F.M., Krishnan T.V and Jain D.C (1994), “Why the Bass Model Fits without Decision Variables,” *Marketing Science*, 13(3), 203-326.

Bass F.M., Gordon K., Ferguson T.L. and Githens M.L. (2001), "DIRECTV: Forecasting Diffusion of a New Technology Prior to Product Launch," *Applied Analytics*, 31(3), 82-93.

Bewley R. and Fiebig, D. (1988), “Estimation of Price Elasticities for an International Telephone Demand Model,” *Journal of Industrial Economics*, 36(4), 393-409.

Blei, D.M., Ng A.Y. and Jordan, M.I. (2003) “Latent Dirichlet Allocation,” *Journal of Machine Learning Research*, 3, 993-1022.

Blei D.M. and Lafferty J.D (2006), “Dynamic Topic Model,” *ICML '06 Proceedings of the 23rd international conference on Machine learning*, 113-120.

Burmester A.B., Becker J.U., Heerde, H.J. and Clement Michel (2015), "The impact of pre- and post-launch publicity and advertising on new product sales," International Journal of Research in Marketing, 32(4), 408-417.

Chevalier J.A and Mayzlin, D. (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43(3), 345-354.

Chen K., Luo P. and Wang H. (2017), "An influence framework on product word-of-mouth (WoM) measurement," *Information & Management*, 54(2), 228-240.

Chintagunta P.K., Gopinath S., Venkataraman S. (2010) "The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets," *Marketing Science*, 29(5), 944–957.

Croxton K.L, Lambert D.M., Garcia-Dastugue S. and Rogers D. (2002) "The Demand Management Process," The International Journal of Logistics Management, 13, 51-66.

Danaher P.J., Hardie B.G.S. and Putsis W.P. (2001), "Marketing-Mix Variables and the Diffusion of Successive Generations of a Technological Innovation," *Journal of Marketing Research*, 38(4), 501-514.

Daniel R., Daniel H., Ramesh N., Christopher D.M. (2009), "Labeled LDA: A Supervised Topic Model for Credit Attribution in Multi-Labeled Corpora," *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, 248-256.

Dellarocas C., Zhang X.M. and Awad N.F. (2007), "Exploring the Value of Online Product Reviews in Forecasting Sales: The Case of Motion Pictures," *Journal of Interactive Marketing*, 21(4), 23-45.

Divakaran P.K.P., Palmer A., Sondergaard H.A. and Matkovskyy R. (2017), "Pre-launch Prediction of Market Performance for Short Lifecycle Products Using Online Community Data," *Journal of Interactive Marketing*, 38, 12-28.

Easingwood C.J., Mahajan V. and Muller E. (1983), "A Nonuniform Influence Innovation Diffusion Model of New Product Acceptance," *Marketing Science*, 2(3), 273-295.

Elberse A. and Anand B. (2007), "The effectiveness of pre-release advertising for motion pictures: An empirical investigation using a simulated market," *Information Economics and Policy*, 19, 319–343.

Geweke J. (1992). "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments," *In Bayesian Statistics 4*, Bernardo, J.M., Berger, J. O., Dawid, A.P. and Smith, A.F.M. (eds.), 169-193. Oxford: Oxford University.

Godes D., Mayzlin D. (2004) "Using online conversations to study word-of-mouth communication," *Marketing Science*, 23(4), 545–560.

Goldenberg J. and Oreg S. (2007), "Laggards in Disguise: Resistance to Adopt and the Leapfrogging Effect," *Technological Forecasting & Social Change*, 74, 1272-1281.

Gopinath S., Chintagunta P.K. and Venkataraman S. (2013), "Blogs, Advertising, and Local-Market Movie Box Office Performance," *Marketing Science,* 59(12), 2635-2654.

Gopinath S., Thomas J.S. and Krishnamurthi L. (2014), " Investigating the Relationship Between the Content of Online Word of Mouth, Advertising, and Brand Performance," *Marketing Science*, 33(2), 241-258.

Guo Z.L. and Chen J.Q. (2018), "Multi-Generation Product Diffusion in the Presence of Strategic Consumers," *Information Systems Research*, 29(1), 206-224.

Hennig-Thurau T., Wiertz, C. and Feldhaus F. (2015), "Does Twitter matter? The impact of microblogging word of mouth on consumers' adoption of new movies," *Journal of the Academy of Marketing Science*, 43(3), 375–394.

Horsky D. and Simon L.S. (1983), "Advertising and the Diffusion of New Products," *Marketing Science*, 2(1), 1-17.

Horsky D. (1990), "A Diffusion Model Incorporating Product Benefits, Price, Income and Information," *Marketing Science*, 9(4), 342-365.

Huang C.Y. and Tzeng G.H. (2008), "Multiple Generation Product Life Cycle Predictions Using a Novel Two-Stage Fuzzy Piecewise Regression Analysis Method," *Technological Forecasting and Social Change*, 75, 12-31.

Hutto, C.J. & Gilbert, E.E. (2014), "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text," *Eighth International Conference on Weblogs and Social Media (ICWSM-14).*

Jiang Z.R. (2010), "How to Give Away Software with Successive Versions," *Decision Support Systems*, 49(4), 430-441.

Jiang Z.R. and Jain D.C (2012), “A Generalized Norton–Bass Model for Multigeneration Diffusion,” *Management Science*, 58(10), 1887-1897.

Joshi A.M. and Hanssens D. (2009), “Movie Advertising and the Stock Market Valuation of Studios: A Case of ‘Great Expectations?’” *Marketing Science*, 28(2), 239-250.

Jun D.B. and Park Y.S. (1999) “A choice-based diffusion model for multiple generations of products,” *Technology Forecasting and Social Change*, 61(1), 45-58.

Kim N. Chang, D. and Shoker, A. (2000), “Modeling Intercategory Dynamics for a Growing Information Technology Industry,” *Management Science*, 46(4), 496-512.

Kim T., Hong J., Koo H. (2013), “Forecasting diffusion of innovative technology at prelaunch: A survey based method,” *Industrial Management and Data Systems,* 113 (6), 800-816.

Kim H. and Hanssens D.M. (2017), “Advertising and Word-of-Mouth Effects on Pre-launch Consumer Interest and Initial Sales of Experience Products,” *Journal of Interactive Marketing*, 37, 57-74.

Kopalle P.K. and Lehmann. D.R. (2006), “Setting Quality Expectations When Entering a Market: What Should the Promise Be?” *Marketing Science*, 25(1), 8-24.

Lee H., Kim S.G., Park. H.W. and Kang P. (2014), "Pre-launch New Product demand forecasting using the Bass model: A statistical and machine learning-based approach," *Technological Forecasting and Social Change*, 86, 49-64.

Lee T.Y. and BradLow, E. (2011), "Automated marketing research Using Online Consumer Reviews," *Journal of Marketing Research*, 48, 881-893.

Li Y. and Terui N. (2018) “Social Media and the Diffusion of an Information Technology Product,” J. Chen et al. (Eds.) *Knowledge and Systems Sciences,* vol.949, 171-185, Springer.

Liu Y. (2006), “Word of mouth for movies: Its dynamics and impact on box office revenue,” *Journal of Marketing*, 70(3), 74–89.

Lu B., Ott M., Cardie C. and Tsou B. (2011), “Multi-aspect Sentiment Analysis with Topic Models,” *Proceedings of the 2011 IEEE 11th International Conference on Data Mining*, 11, 81-88.

Marchand A., Hennig-Thurau T., Wiertz C. (2016), "Not all digital word of mouth is created equal: Understanding the respective impact of consumer reviews and microblogs on new product success," *International Journal of Research in Marketing*, 34(2), 1-19.

Mahajan V. and Muller, E. (1981), "A Nonsymmetric Responding Logistic Model for Forecasting Technological Substitution," *Technology Forecasting and Social Change*, 21(3), 199-213.

Mahajan V. and Muller E., Bass F.M. (1995), "Diffusion of New Products: Empirical Generalizations and Managerial Uses," *Marketing Science, 14(3)*, 79-88.

Mahajan V. and Muller E. (1996), "Timing, Diffusion, and Substitution of Successive Generations of Technological Innovations: The IBM Mainframe Case," *Technological Forecasting and Social Change*, 51(2), 213-224.

Moe W. and Trusov M. (2011), "The Value of Social Dynamics in Online Products Ratings Forums," *Journal of Marketing Research*, 49, 444-456.

Netzer O., Feldman R., Goldenberg J., and Fresko M. (2012), "Mine Your Own Business: Market-Structure Surveillance Through Text Mining," *Marketing Science*, 31 (3), 521-543.

Norton J.A. and Bass F.M. (1987), "A Diffusion Theory Model of Adoption and Substitution for Successive Generations of High-Technology Products," *Management Science,* 33(9), 1069-1086.

Ritesh B.R., Chethan R., Harsh S.J., Sheetal V.A. (2017), "Stock Market Prediction Using News Articles," *Journal of Emerging Technologies and Innovative Research*, 4(03), 153-155.

Rui H., Liu Y. and Whinston A. (2013), "Whose and what chatter matters? The effect of tweets on movie sales," *Decision Support Systems*, 55(4), 863–870.

Robinson B. and Lakhani C. (1975), "Dynamic Price Models for New-Product Planning," *Management Science*, 21(10), 1089-1214.

Shen and Willems (2012), "Strategic sourcing for the short-lifecycle products," *Int. J. Production Economics* 139, 575-585.

Shi X., Fernandes K. and Chumnumpan P. (2014), "Diffusion of Multi-Generational High-Technology Products," *Technovation*, 34, 162-176.

Srinivasan V. and Mason C.H. (1986), "Technical Note—Nonlinear Least Squares Estimation of New Product Diffusion Models," *Marketing Science*, 5(2), 169-178.

Terui N., Ban M. and Maki T. (2010), "Finding Market Structure by Sales Count Dynamics - Multivariate

Structural Time Series Models with Hierarchical Structure for Count Data," *Annals of the Institute of Statistical Mathematics*, **62**, 92-107.

Terui N. and Ban M. (2014), "Multivariate Structural Time Series Models with Hierarchical Structure for Over-dispersed Discrete Outcome," *Journal of Forecasting*, **33**, 376-390.

Tirunillai S. and Tellis G.J. (2014), "Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation," *Journal of Marketing Research*, 51(4), pp. 463-479.

Trusov M., Bucklin R.E. and Pauwels K. (2009), "Effects of Word-of-Mouth versus Traditional Marketing: Findings from an Internet Social Networking Site," Journal of Marketing, 73(5), 90-102.

Von Bertalanffy L. (1957), "Quantitative Laws in Metabolism and Growth," *The Quarterly Review of Biology*, 32(3), 217-231.

Watanabe K. and Zhou Y. (2020), "Theory-Driven Analysis of Large Corpora: Semisupervised Topic Classification of the UN Speeches," *Social Science Computer Review*, 2020, 1-21.

Appendix.

A. Labeled Dynamic Topic model

The DTM is base model for LDTM as detailed in Blei and Lafferty (2006). Assuming we define a set of seed words $seed\_words = \{"camera": 1, "like": 2\}$, means word "camera" is fixed to the 1st topic, while "like" is fixed to the 2nd topic. $seed\_words[word]$ denotes the corresponding topic of the related word. Based on DTM, The algorithm of LDTM can be written as

For t = 1,2, …, T

1. Draw Topics

   $\beta_{t,k}|\beta_{t-1,k} \sim N(\beta_{t-1,k}, \sigma^2 I)$

2. Draw a_t

   $\alpha_t|\alpha_{t-1} \sim N(\alpha_{t-1}, \delta^2 I)$

For each document:

3. Topic proportion

   $\eta_{i,d}|\alpha_t \sim N(\alpha, a^2\mathbf{I})$

   $\theta_{i,d}|\eta_{i,d} \sim \pi(\eta_{i,d})$

   For each word:

   if word in *seed_words*:

   $z_{t,d,n} = seed\_words[word]$

   else:

4. Topic-word assignment

   $z_{t,d,n}|\theta_{t,d} \sim Multinomial(\theta_{t,d})$

   $w_{d,n}|z_{d,n}\{\beta_{t,k}\} \sim Multinomial\left(\pi(\beta_{t,z_{d,z}})\right)$

Here $\pi(\cdot)$ is a softmax function that can fix the sum of $\beta_{t,z_{d,z}}$ to 1.

B. MCMC method for the Bass model

B.1 Prior Settings for the Bass model

Parameter	Setting
$m_G{}^{*} \sim N(\mu_{m0}, \tau_{m0}{}^{-1})$	$\mu_{m0} = 0, \tau_{m0} = 0.1$
$q_G{}^{*} \sim N(\mu_{q0}, \tau_{q0}{}^{-1})$	$\mu_{q0} = 0, \tau_{q0} = 0.01$
$p^{*} \sim N(\mu_{p0}, \tau_{p0}{}^{-1})$	$\mu_{p0} = 0, \tau_{p0} = 0.01$
$X_G(t) \sim N(\mu_X, \tau_X{}^{-1})$	$\mu_{X0} = 0, \tau_{X0} = 0.01$
$\sigma \sim IG(a, b)$	$\mathrm{a} = 0.5, \mathrm{b} = 0.5$
$\alpha_G \sim N(\mu_{\alpha0}, \tau_{\alpha0}{}^{-1})$	$\mu_{\alpha0} = 0, \tau_{\alpha0} = 0.01$
$\boldsymbol{\beta_G} \sim N(\mu_{\beta0}, \tau_{\beta0}{}^{-1})$	$\mu_{\beta0} = 0, \tau_{\beta0} = 0.01$
$\sigma_x \sim IG(a, b)$	$\mathrm{a} = 0.5, \mathrm{b} = 0.5$
$\boldsymbol{\Theta} \sim N(\mu_{\Theta0}, \tau_{\Theta0}{}^{-1})$	$\mu_{\Theta0} = 0, \tau_{\Theta0} = 0.01$
$\boldsymbol{\Xi} \sim IG(a, b)$	$\mathrm{a} = 0.5, \mathrm{b} = 0.5$

* $\boldsymbol{\delta}$ is the vector of $(\Delta_m, \Delta_q, \sigma_\alpha, \sigma_\beta)$, $\boldsymbol{\Xi}$ is the vector of $(\sigma_m, \sigma_q, \sigma_\alpha, \sigma_\beta)$.

B.2 Posterior Density for Model Parameters

The joint posterior density of our model parameters is represented by

$$
\begin{aligned}
& p\left(\{m_G, q_G, p\}, \{X_G(t)\}, \sigma, \{\boldsymbol{\alpha}_G, \beta_G\}, \sigma_x, \{\boldsymbol{\Delta_m}, \boldsymbol{\Delta_q}, \boldsymbol{\Delta_\alpha}, \boldsymbol{\Delta_\beta}\}, \{\sigma_m, \sigma_q, \sigma_\alpha, \boldsymbol{\sigma_\beta}\} \mid \{\boldsymbol{y}_G, t\}, \{\mathbf{Topic}'_{\mathbf{G}}(t-1), \boldsymbol{T}_G, P'_G(t)\}\right) \\
&\quad = p\left(\{m_G, q_G, p\}, \{X_G(t)\} \mid \{\boldsymbol{y}_G, t\}, \{\boldsymbol{\Delta_m}, \boldsymbol{\Delta_q}\}, \{\boldsymbol{\alpha}_G, \beta_G\}, \sigma\right) p\left(\sigma \mid \{m_G, q_G, p\}, \{\boldsymbol{y}_G, t, X_G(t)\}\right) \\
&\quad \times p\left(\boldsymbol{\alpha}_G \mid \{X_G(t), t, \mathbf{Topic}'_{\mathbf{G}}(t-1), P'_G(t)\}, \boldsymbol{\Delta_\alpha}, \sigma_x\right) \\
&\quad \times p\left(\beta_G \mid \{X_G(t), t, \mathbf{Topic}'_{\mathbf{G}}(t-1), P'_G(t)\}, \boldsymbol{\Delta_\beta}, \sigma_x\right) p\left(\sigma_x \mid \{\boldsymbol{\alpha}_G, \beta_G\}, \{X_G, t, \mathbf{Topic}'_{\mathbf{G}}(t-1), P'_G(t)\}\right) \\
&\quad \times p\left(\boldsymbol{\Delta_m} \mid \{m_G, m_{G-1}\}, \{\boldsymbol{T}_G\}, \sigma_m\right) p\left(\sigma_m \mid \{m_G, m_{G-1}\}, \{\boldsymbol{T}_G\}, \boldsymbol{\Delta_m}\right) \\
&\quad \times p\left(\boldsymbol{\Delta_q} \mid \{q_G, q_{G-1}\}, \{\boldsymbol{T}_G\}, \sigma_q\right) p\left(\sigma_q \mid \{q_G, q_{G-1}\}, \{\boldsymbol{T}_G\}, \boldsymbol{\Delta_q}\right) \\
&\quad \times p\left(\boldsymbol{\Delta_\alpha} \mid \{\alpha_G, \alpha_{G-1}\}, \{\boldsymbol{T}_G\}, \sigma_\alpha\right) p\left(\sigma_\alpha \mid \{\alpha_G, \alpha_{G-1}\}, \{\boldsymbol{T}_G\}, \boldsymbol{\Delta_\alpha}\right) \\
&\quad \times p\left(\boldsymbol{\Delta_\beta} \mid \{\beta_G, \beta_{G-1}\}, \{\boldsymbol{T}_G\}, \boldsymbol{\sigma_\beta}\right) p\left(\boldsymbol{\sigma_\beta} \mid \{\beta_G, \beta_{G-1}\}, \{\boldsymbol{T}_G\}, \boldsymbol{\Delta_\beta}\right) \qquad \text{(B.1)}
\end{aligned}
$$

where $\boldsymbol{\Delta_i}$ is the coefficient of the vector of each equation in Eq. (4).

The second line of Eq. (B.1) captures the product of conditional posterior density for parameters in the diffusion model in Eqs. (1) and (2). The third and fourth lines are joint posterior density of parameters in the marketing mix in Eq. (3). The fifth to eighth lines define joint posterior density for parameters in the hierarchical structure connecting the (*G*-1) generation to the *G* generation in Eq. (4).

We employ MCMC to estimate parameters because the procedure for hierarchical models is well-established and some of necessary conditional posterior densities are available in closed form. The sampling scheme of MCMC for estimating the model is a hybrid of Metropolis-Hastings and Gibbs sampling for other parameters in hierarchical models.

B.3 Conditional Posterior Distributions

(1) ${m_G}^* \mid \{y_G(t), t, X_G(t)\}, \{p^*, {q_G}^*\}, \sigma$

For *iter* (=1, …, R) of MCMC iterations, we use Metropolis-Hastings with a random walk algorithm for each generation G,

$$m_G^{*(iter)} = m_G^{*(iter-1)} + \lambda_m;\ \lambda_m \sim N(0,0.1), \tag{B.2}$$

where the acceptance probability is

$$\alpha = \min\left(1, \frac{p\left(m_G^{*(iter)} \mid \{y_G(t), t, X_G(t)\}, \{p^*, {q_G}^*\}, \sigma\right)}{p\left(m_G^{*(iter-1)} \mid \{y_G(t), t, X_G(t)\}, \{p^*, {q_G}^*\}, \sigma\right)}\right), \tag{B.3}$$

where $t = 1, \ldots, N$ and $G = 1, 2, \ldots, 5$.

(2) $p^* \mid \{y_G(t), t, X_G(t)\}, \{{m_G}^*, {q_G}^*\}, \sigma$

For p and q_G, we also use Metropolis-Hastings sampling, which is the same as m_G above.

$$p^{*(iter)} = p^{*(iter-1)} + \lambda_p;\ \lambda_p \sim N(0,0.01). \tag{B.4}$$

The probability of acceptance is

$$\min\left(1, \frac{p\left(p^{*(iter)} \mid \{y_G(t), t, X_G(t)\}, \{m_G, q_G\}, \sigma\right)}{p\left(p^{*(iter-1)} \mid \{y_G(t), t, X_G(t)\}, \{m_G, q_G\}, \sigma\right)}\right). \tag{B.5}$$

(3) $q_G \mid \{y_G(t), t, X_G(t)\}, \{m_G, p\}, \sigma$

$$q_G^{*(iter)} = q_G^{*(iter-1)} + \lambda_q;\ \lambda_q \sim N(0,0.01). \tag{B.6}$$

The acceptation probability is

$$\min\left(1, \frac{p\left(q_G^{*(iter)} \mid \{y_G(t), t, X_G(t)\}, \{m_G, p\}, \sigma\right)}{p\left(q_G^{*(iter-1)} \mid \{y_G(t), t, X_G(t)\}, \{m_G, p\}, \sigma\right)}\right). \quad (B.7)$$

(4) $X_G(t) \mid \{m_G, q_G, p\}, \sigma$

$$X_G{}^{(iter)}(t) = X_G{}^{(iter-1)}(t) + \lambda_X;\ \lambda_X \sim N(0{,}0.01). \quad (B.8)$$

The acceptation probability is

$$\min\left(1, \frac{p\left(X_G(t)^{(iter)} \mid \{y_G(t), t, \}, \{m_G, q_G, p\}, \sigma\right)}{p(X_G(t)^{(iter-1)} \mid \{y_G(t), t\}, \{m_G, q_G, p\}, \sigma)}\right). \quad (B.9)$$

(5) $\sigma \mid \{y_G(t), t, \boldsymbol{X_G(t)}\}, \{m_G, q_G, p\}$

If we define estimated sales in period t for the generation *G* as

$$\hat{y}_G(t) = f(\{y_G(t), t, X_G(t)\}, \{m_G, q_G\}), \quad (B.10)$$

we can update σ by

$$IG\left(\mathrm{a} + \frac{\mathrm{n}}{2}, \mathrm{b} + \frac{\sum_{t=1}^{n}(\mathrm{y}_G(t) - \sum_{G=1}^{M}\hat{y}_G(t))^2}{2}\right), \quad (B.11)$$

M stands for number of generations in this equation.

Marketing Mix.

(6) $\alpha_G \mid \{\boldsymbol{\beta_G}\}, \{X_G(t), t, \boldsymbol{Topic'_G(t-1)}, P'_G(t)\}, \sigma_x$

$$N\left(\left(n\sigma_\beta + \sigma_{\beta 0}\right)^{-1}\left(\sigma_x \sum_{i=1}^{n}(X_G(t) - \boldsymbol{\beta_G}\,'\,\boldsymbol{Topic'_{Gi}(t-1)})P'_G(t)^{-1} + \mu_{\beta 0}\sigma_{\beta 0}\right), \left(n\sigma_x + \sigma_{\beta 0}\right)^{-1}\right) \quad (B.12)$$

(7) $\boldsymbol{\beta_G} \mid \{\alpha_G\}, \{X_G(t), t, \boldsymbol{Topic'_G(t-1)}, P'_G(t)\}, \sigma_x$

For each topic *j* the posterior of coefficient β_{Gj} can be derived from a normal regression equation from

$$N(n\sigma_x + \sigma_{\alpha 0})^{-1}\left(\sigma_x \sum_{i=1}^{n}\left(X_G(t) - \alpha_G \cdot P'_G(t) - \sum_{k \neq j} Topic'_{Gk}(t-1) \cdot \beta_{Gk}\right) Topic'_{Gj}(t-1)^{-1} + \mu_{\alpha 0}\sigma_{\alpha 0}\right),$$

$$, (n\sigma_x + \sigma_{\alpha 0})^{-1})$$

$$(B.13)$$

(8)$\sigma_x \mid \{\alpha_G,\ \boldsymbol{\beta_G}\}, \{X_G(t), t, \boldsymbol{Topic'_G(t-1)}, P'_G(t)\}$

$$IG\left(\alpha + \frac{n}{2}, \beta + \frac{\sum_{i=1}^{n}(X_G(t) - \alpha_G \cdot P'_{Gi}(t) - \boldsymbol{\beta_G}\,'\,\boldsymbol{Topic'_{Gi}(t-1)})^2}{2}\right). \tag{B.14}$$

Hierarchical Structure.

(9) $\boldsymbol{\Delta_m} \mid \{m_G, m_{G-1}\}, \{\boldsymbol{T}_G\}, \sigma_m$

Assuming D as data matrix for hierarchical structure for *m,* and $\Delta_{\boldsymbol{m}}$ as coefficient vector, we can derive the posterior of coefficient Δ_{mj} by

$$N\left((n\sigma_m + \sigma_{m0})^{-1}\left(\sigma_m \sum_{G=1}^{M}\left(m_G - \sum_{z=1}^{K\neq j}(D_{Gz} \cdot \delta_{mz})\right)D_{Gj}^{-1} + \mu_{m0}\sigma_{m0}\right), (n\sigma_m + \sigma_{m0})^{-1}\right). \tag{B.15}$$

(10) $\sigma_m \mid \{m_G, m_{G-1}\}, \{\boldsymbol{T}_G\}, \boldsymbol{\Delta_m}$

$$IG\left(\alpha + \frac{n}{2}, \beta + \frac{\sum_{G=1}^{M}(m_G - \sum_{z=1}^{Z}(D_{Gz} \cdot \delta_{mz}))^2}{2}\right), \tag{B.16}$$

where Z means the number of coefficients in the hierarchical structure. As the sampling methods are the same among all the hierarchical structures, we can sample for other parameters ($\boldsymbol{\Delta_q}, \boldsymbol{\Delta_\alpha}$ *and* $\boldsymbol{\Delta_\beta}$) as well.

Forecasting

(10) $y(t) \mid m_G, q_G, p, X_G(t), \boldsymbol{T_G}, \boldsymbol{Topic'_G(t-1)}, P'_G(t)$

For $iter = 1, \ldots,$ ITER, where ITER means the max iteration, after sampling all the parameters using MCMC, the forecasting unit sales for the G-th generation $y_G(t)$ can be written as

$$\begin{cases} y_1(t)^{(iter)} = m_1^{(iter)} f_1(t), & t < \tau_2, \\ y_1(t)^{(iter)} = m_1^{(iter)} f_1(t)\left(1 - F_2\left(t - \tau_2\right)\right), & t \geq \tau_2, \end{cases}$$

$$\begin{cases} y_G(t)^{(iter)} = m_G^{(iter)} f_G(t-\tau_G) F_G\left(t - \tau_G\right) & \tau_G \leq t < \tau_{G+1},\ 1 < G \leq N, \\ \qquad + y_{G-1}(t)^{(iter)} F_G\left(t-\tau_G\right) + Y_{G-1}(t)^{(iter)} f_G(t-\tau_G), & \\ y_G(t)^{(iter)} = (m_G^{(iter)} f_G(t-\tau_G) F_G\left(t-\tau_G\right) & t \geq \tau_{G+1},\ 1 < G \leq N, \\ \qquad + y_{G-1}(t)^{(iter)} F_G\left(t-\tau_G\right) + Y_{G-1}(t)^{(iter)} f_G(t-\tau_G)) \times \left(1 - F_{G+1}\left(t - \tau_{G+1}\right)\right), & \end{cases}$$

(B.16)

where

$$F_G\left(t \mid p^{(iter)}, q_G^{(iter)}\right) = \frac{1-\exp\left(-\left(p^{(iter)}+q_G^{(iter)}\right)X_G(t)^{(iter)}\right)}{1+\left(q_G^{(iter)}/p^{(iter)}\right)\exp\left(-\left(p^{(iter)}+q_G^{(iter)}\right)X_G(t)^{(iter)}\right)}. \tag{B.17}$$

Note that

$$\begin{cases} X_G(t)^{(iter)} = t, & t < \tau_G \\ X_G(t)^{(iter)} = t + \alpha_G^{(iter)} \log\left(P_G'(t)\right) + \boldsymbol{\beta}_{\boldsymbol{G}}^{(\boldsymbol{iter})'} \mathbf{log}(\boldsymbol{Topic}_{\boldsymbol{G}}'(\boldsymbol{t}-\mathbf{1})), & t \geq \tau_G \end{cases} \tag{B.18}$$

then total sales $y(t)^{(iter)}$ can be calculated by

$$y(t)^{(iter)} = \sum_G y_G(t)^{(iter)} . \tag{B.19}$$