ESSAYS ON FORECASTING ONLINE SHOPPING SEARCHES

A Dissertation

Submitted to

The Faculty of the C.T. Bauer College of Business

University of Houston

In Partial Fulfillment of

The Requirements for the Degree of

Doctor of Philosophy

by

Jiang Qian

May 2020

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

Page

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my committee chairs, Rex Du and Ye Hu, and my committee members, Seshadri Tirunillai and Ming Zhao, for their guidance and support throughout the process of completing this dissertation.

In addition, I want to thank Edward Blair, Betsy Gelb, James Hess, Sam Hui, Partha Krishnamurthy, Shijie Lu, Vanessa Patrick, Melanie Rudd, Kitty Wang, and Michael Ahearne for extremely helpful suggestions and comments. I want to also offer my appreciation to my friends, colleagues, and the department staff. They make my time at University of Houston a wonderful experience.

Last but not least, I am grateful for my parents and my wife. They are always there to help me out of my troubles. Without them, this dissertation would not have been possible.

ABSTRACT

My dissertation consists of two essays, both leveraging data from Google Shopping Insights in developing models for forecasting online shopping searches, a critical precursor of sales. The first essay presents a novel Bass-type diffusion model for forecasting online shopping searches of rapid life cycle (RLC) products. The most important innovation of my model is that it allows for imitation/contagion effects to take place through dual channels: a ‘local’ channel of influence mainly through direct, in-person interactions (e.g., schoolmates, colleagues, neighbors) and a ‘national’ channel of influence mainly through social media (e.g., YouTube, Instagram, Facebook). To separate the effects of these two channels of influence, I leverage the fact that data from Google Shopping Insights is at the city level, which allows me to model consumer shopping searches in a particular month and city as a function of not only past searches in the city but also past searches in the rest of the country. The former is treated as a proxy for the amount of local influence, while the latter a proxy for the amount of national influence. Empirical estimates suggest that imitation/contagion can indeed take place through both local and national channels, with their relative importance varying across products. When influence from the national (local) channel dominates, the diffusion curve tends to be steeper (flatter), which provides support for the idea that imitation/contagion through social media, compared to in-person interactions, shortens the product life cycle. Another important feature of my model is that it allows for rapid decay in the influence of prior adopters. Empirical estimates show that, for most RLC products, the influence of prior adopters drops drastically after just one month.

The second essay develops a “Big Data” solution to the so-called ‘cold start’ problem in forecasting, where insufficient longitudinal information prevents one from extrapolating historical patterns into the future with standard time series methods. The innovation of my solution is to mitigate the cold start problem by compensating for the lack of longitudinal data with the abundance of data from a large number of cities and products in Google Shopping Insights that can serve as training samples. My solution adopts a fusion of multiple methods for identifying similar products, and then leverages the spatiotemporal patterns of those similar products in the holdout period to forecast city-level growth in online shopping searches for the focal product. Extensive empirical comparisons suggest that my solution outperforms the benchmarks. Furthermore, I find that a bigger training sample is not always better: a gradual increase in the size of the training sample first improves and then counterintuitively reduces forecasting performance. I attribute this finding to the fact that the incremental predictive value of additional training data diminishes as the sample size increases, while the proportion of noise remains. As a result, methods commonly used for variable/feature selection fail to remove the added noises, resulting in over-fitting and thus deteriorating forecasting performances. This finding cautions that, even in the "Big Data" era, all else being equal, the bigger the training data is not necessarily the better when it comes to forecasting demand growth for new products.

CHAPTER 1. Modeling the Diffusion of Online Search Interests in Rapid Life Cycle Products

Introduction

The wide adoption of social media overcomes physical boundaries and connects people. As a result, new product information in such an interlinked society spreads much faster among consumers (Shapiro and Varian 1998). This creates a positive feedback loop that facilitates the rapid development of new trends and shift of consumer tastes. In such a climate, I have witnessed some fascinating new products. Take the fidget spinner as an example – in a matter of months, it rose to stardom, dominating the air waves and at one point accounting for 20% of the toys and games sold online in the U.S., and then just as rapid as its upsurge, it fell out of consumer favor. Figure 1.1 illustrates the fidget spinner's rise and fall in Google search index (Panel A) and sales (Panel B).

[Insert Figure 1.1 about here.]

Despite their short life spans, such products, often characterized by a greatly accelerated life cycle (Anderson and Zeithaml 1984), are not failures. They are indeed splashing successes – consumers bought tens of millions of fidget spinners, worth hundreds of millions of dollars (The Economist 2017). Although the phenomenon of rapid life cycle (RLC) products isn't new (e.g., Pet Rock 1975-1976) [1], the acceleration of information dissemination in social media has created the soil for making their cycles shorter and more importantly, their occurrences far more frequent, from being merely incidental to predictably recurrent. To name a few examples, over a short period of time between 2015 and 2017, the fidget spinner, adult coloring books (No. 1 and 2 bestselling books on Amazon in April 2015, Flood 2015), and L.O.L. Surprise! (No. 1 toy in

[1] Such a short-lived trend that happened for no apparent reason is often referred to as a "fad" (Bikhchandani, Hirshleifer, and Welch 1992). In this paper, I refer to such products by their key characteristics of rapid life cycles.

the U.S. through November 2017, Semuels 2018) all experienced substantial national level successes and then fizzled out. Consequently, given the scale and regularity of modern-day RLC products, it can become worthwhile for businesses to tap into the constant emergence of such rapid trends and build a business model to plan around them. That said, chasing an explosive trend like the fidget spinner was not risk free. Zing, a toy company located in Portland, Oregon, noticed the trending hashtags related to fidget spinners and jumped onto the bandwagon. Although at the beginning its fidget spinners were sold out immediately, four months after June 2017, the trend had completely faded away, and companies like Zing were left with a large amount of inventory (Nicolaou 2017).

The volatility of RLC products poses various modeling challenges, because the lead time is short, the number of potential leads is high, and their paths to popularity differ markedly. Piecing together the full cycle of an RLC product at its budding stage can be a tall order. The default solution in marketing, often rooted in the Bass model (Bass 1969), intends to resolve the long range forecasting problem for durables. While the three-parameter Bass model doesn't require many periods of observations to implement, its parameters are time invariant, weighing the recent and distant past observations equally. This makes the model insufficiently adaptive to more recent changes when the diffusion process progresses rapidly. In practice, marketing research analysts often have to rely on other analogous products, with their full diffusion cycles known, to fine tune their forecasts in order to ensure performance (Lilien and Rangaswamy 2004). Given the distinct nature of RLC products, however, analogies are not obvious to spot. Recent developments in trend forecasts, such as the quantitative trend spotting method based on dynamic factor analysis (Du and Kamakura 2012), require a relatively long period of observations, a luxury often unavailable for RLC products. In this paper, I aim at filling the gap in the literature by leveraging on a spatiotemporal data from the Google Shopping Insights (GSI, https://shopping.thinkwithgoogle.com/) to develop a parsimonious and yet informative model for tracking the online shopping search interests in RLC products. The GSI data is granular comparing to what's common in diffusion models. It contains a large number of products, instead of the usual one or two. In addition, it tracks hundreds of cities in the U.S. With this data, instead of sales, I focus on the diffusion of shopping search interests, a measure shown in both Figure 1.1 and previous research to be closely associated with product sales (Du, Hu, and Damangir 2015; Hu, Du, and Damangir 2014). Comparing to the Bass model, my model performs well in capturing both the timing and peaks of the shopping search history of various RLC products and is able to significantly improve the forecast accuracy for out-of-sample holdout products.

I arrange the remainder of the paper as follows. First, I detail the full product life cycle of the fidget spinner to provide a general descriptive characterization of RLC products. Then I explain the data, model development, and results, followed by applying the model to forecast the life cycle of two holdout products. Finally, I discuss the managerial implications and caveats of my model.

Characterizing a Rapid Lifecycle Product

As shown in Figure 1.1, the fidget spinner is one of the most well-known among the RLC products. In this section, I use the fidget spinner as an example to illustrate the diffusion process of a rapid cycle product and summarize the key characteristics of such products.

The fidget spinner, originally considered to help children with attention-deficit hyperactivity disorder, stormed the market instead as a novelty toy in 2017. A vast amount of publicity received by fidget spinners was through social media and various online channels (The

Economist 2017). This information diffusion process also manifests itself in the form of online searches. As shown by the data from the Google Trends (https://trends.google.com), Figure 1.1, Panel A illustrates the Google search index for the keyword "fidget spinner" over time. The interests started the major uptick at the beginning of 2017, followed by a sharp surge that peaked in early May of the same year. By the end of 2017, the search interests all but vanished. Figure 1.1, Panel B (note for the different time span from Panel A) illustrates the online sales of the fidget spinner, which had a similar upsurge process and peaked in May 2017. At its apex, all 20 of the top selling toys on Amazon were fidget spinners or their mutations. By any measure, its market size was in the magnitude of hundreds of millions of dollars, enviable even for a large toy manufacturer. Importantly, despite its popularity, the fidget spinner has never seen its days in advertising on national TV (The Economist 2017).

To recap the case of the fidget spinner, I conclude with two characteristics that define an RLC product. First, an RLC product is rarely a necessity, or even a utilitarian product. Like the fidget spinner, the factors that drive the decision to purchase such a product can often be novelty, impulses, or even the social needs to blend in with a group. This is perhaps the main reason why the sales cycle can pick up and spread quickly but is lack of a utilitarian anchor to sustain itself. Consequently, the life cycle of such a product, measured in weeks or months, is usually composed of both a sharp rise and a deep downfall of consumer interests. It is very different from a typical durable good with a more sustained life cycle measured by years or decades. Second, the diffusion of an RLC product is often a grassroots movement, starting from consumers and gradually responded by manufacturers. As a result, social media, instead of mass market advertising campaigns plays a significant role in spreading the trends. Given the borderless accessibility of new media such as Facebook, Instagram, Twitter and YouTube, information diffusion at the national level can be a leading influence for the takeoff of the product life cycle. Because of such grassroots nature, information cascades plays of a major role in the diffusion process (Bikhchandani, Hirshleifer, and Welch 1992). It is not surprising that the crowd would eventually move on to the other novel things.

Data

I acquired the data from the Google Shopping Insights, a business intelligence platform launched in 2014. With this data, I are able to track fifteen RLC products introduced in the U.S. market between January 2014 and March 2018. Google aggregates the shopping related searches for each product at https://google.com and https://google.com/shopping, and creates a "shopping search interests" measure, which is closely associated with the actual search volume.[2] Google aggregates queries like "best fidget spinner", "fidget spinner deals", and "fidget spinner price" all to the product "fidget spinner".[3] Out of the fifteen RLCs products, I use thirteen of them to calibrate the model and leave two as out-of-sample holdouts. I aggregate the search interests at a monthly level. In total, my data covers 51 months between January 2014 and March 2018.

The product search interests are from the most populated 60 Census designated places in the United States with a population size larger than 300,000. The Census designated places are often delineated around concentration of population, housing, and commercial structures. With my relatively high population threshold, they are typically cities (hereafter I refer to my unit of geographic measure as a "city" instead of a Census designated place). The largest city in my

[2] Google doesn't explicitly define the "shopping search interests" measure as a straight sum of the actual shopping related search volume. However, from all the evidence I see, it closely tracks the actual volume. Unlike the Google Trends data, which is indexed to a maximum of 100 for each keyword or set of keywords, the GSI data I use is scaled and comparable across products.

[3] https://shopping.thinkwithgoogle.com/faq

sample is New York City, New York, with a population of 8.37 million and the smallest is Anchorage, Alaska, with a population of 301 thousand. The total population in these cities is approximately 53 million, or 17% of the U.S. population, a caveat that I discuss more extensively in the conclusion.

Not surprisingly, bigger cities tend to have higher search interests, on all the products. Instead of using this absolute measure, I transform search interests into the penetration of search interests, as normalized by the population of the city. In Equation 1.1, I define the penetration of search interests[4], $\mathrm{s_{ijt}}$, as the search interests per 100,000 people for product i (i = 1, …, I) in city j (j = 1, …, J) during month t (t = 1, …, T):

$$\mathrm{s_{ijt}} = \frac{(\text{Search Interests})_{\mathrm{ijt}} \times 100{,}000}{\text{Population}_{\mathrm{j}}} \tag{1.1}$$

The point of initial product introduction or even awareness may sound like an appealing time point to start tracking the diffusion process and producing forecasts for the RLC products, as it seemingly would lead to early market intelligence, something very desirable given my goals. This is however not true. As Golder and Tellis (1997) show, the two common reasons why the launch point isn’t a good point for forecasting are: 1) Many products fail before they have been widely adopted by the public. Those unsuccessful products would only bring more noise to the model. 2) Although some products do succeed in the end, their diffusion processes can be accompanied by a long and slow acceptance pattern at the beginning. The large increase in sales only happens when the number of adopters has accumulated to a certain degree. They define the time point of the transition from the introductory stage to the fast growth stage as the “takeoff”, a critical time point in the life cycle of a new product that is associated with all aspects of

[4] Hereafter I refer to $\mathrm{s_{ijt}}$ as “search interests”.

manufacturing, financing, marketing, product distribution, and inventory management. It also has important implications to whomever responsible for making the decision to promote the new products (Tellis, Stremersch, and Yin 2003).

Therefore, the length of the introductory stage can vary greatly. Since the sales or search interests are not stable in this stage, the data availability is also a problem. Perhaps as a result, the marketing literature on modeling takeoff is very limited. One of the most widely implemented is the aforementioned descriptive definition by Golder and Tellis (1997), that the takeoff is the beginning of a new phase in the sales history of a product, marked by rapid growth. They model the takeoff as a linear function of the baseline sales and growth rate in sales. Specifically, the takeoff pattern is characterized by a large percentage increase in sales when the base level sales is small, or contrariwise, a small percentage increase in sales with a relatively large base level sales. They calibrated the model using consumer durables such as color television, clothes dryer, and microwave oven.

To apply the takeoff model to the RLC products in my sample, I first calculate the monthly growth rate and highlight the big jump in growth after introductory. For most of the RLC products, the threshold rule can seamlessly apply. Figure 1.2, Panel A illustrates the application of the threshold rule to the fidget spinner. The line represents the monthly shopping search interests before it reaches its peak. I observe a sizable jump in search interests between September and October 2016. Consequently, I code the month September 2016 as the takeoff point of the fidget spinner.

[Insert Figure 1.2 about here.]

Universally applying Golder and Tellis (1997) to model the takeoff of all the RLC products, however, may incur two problems. First, they measure the takeoff rule on a yearly basis. The annual adoption growth is relatively stable for durables and their life cycles can last for decades. In my research, the adoption process updates on a monthly basis because of the rapid life cycle. With RLC products' narrow observation time windows, the data becomes even more sparse and susceptible to random shocks. Second, I face the risk of seasonal surges that falsely signal the takeoff. It is especially important to eliminate the temporary sales rises caused by seasonality. Simply using the first large growth as the takeoff signal can increase the risk of false takeoffs.

As an example, Figure 1.2, Panel B shows the takeoff pattern for the adult coloring book. Two sizable jumps in search interests exceeded the threshold after introduction. The first takeoff point, followed by a drop in search interests, would have been defined as the takeoff per Golder and Tellis (1997). There is however, a second takeoff point, from where the search interests never looked back during this time window. Thus, the second time point is a more appropriate choice for the takeoff of the adult coloring books. By modifying the threshold rule concept introduced by Golder and Tellis (1997), I make the framework flexible enough for all the RLC products. Specifically, my approach requires the search interests to increase continuously in the period right after takeoff, not just an isolated uptick. This requirement ensures that seasonality or unexpected events are not the sole cause of the jumps in search interests.

Figure 1.2, Panel C shows, for all the takeoff points, the scatter plot between the search interests and the growth rate measured by the change in search interests. Given the curvilinear relationship between these two, I revise Golder and Tellis (1997)'s linear method to fit a two-parameter curve (Equation 1.2), which yields better performance in forecasting. To define the takeoff point, the actual growth rate of the search interests for product i in city j must be above the curve described by s_{ij0} declining at an exponential rate of -1.1, adjusted by a constant factor

27,000 (Equation 1.2).

$$\text{Growth Rate}_{ij} = 27000 \times s_{ij0}^{-1.1} \tag{1.2}$$

For each RLC product, I only use the twelve-month time window after the takeoff to calibrate the model. Prior to the takeoff, the search interests in many cities tend to be intermittent and unreliable due to sparsity. I truncate the data after the twelfth month of takeoff, as the dominating search interests have all but fizzled out by this point. Table 1.1 reports the takeoff and peak months for each product in my sample. For instance, the fidget spinner took off in September 2016 and the diffusion process peaked in May 2017. The longitudinal history of the search interests is largely consistent with that of online sales (Figure 1.1). The most common takeoff months are July, August and September. The most common peak month is December.

In Table 1.1, I also report the summary statistics of s_{ijt} for each product. The fidget spinner has the highest peak search interests of 157,739 per 100,000 people. The zero minimum search interests happened because some of the RLC products lost search interests almost entirely towards the end of their twelve-month window. I leave two products, Doc McStuffins Mobile Cart and J-Animals, as the holdout samples to evaluate the performance of my forecasting model. The takeoff month was July 2014 for both products.

Model Development

I extend the Bass model to capture several key characteristics of the diffusion process of RLC products. Consider the classic Bass model, where the probability of purchasing during time period t, f(t), given that no purchase has happened yet is a function of F(t), the cumulative distribution function of f(t).

$$\frac{\mathrm{f(t)}}{1-\mathrm{F(t)}}=\mathrm{p}+\mathrm{qF(t)} \tag{1.3}$$

where the innovation parameter p and imitation parameter q are both constants over time.

First of all, given the rapid pace of the product life cycles I model, an RLC product adopter's attention span is likely to be short as they quickly move on to the next "new thing". Reflected in the Bass model, the imitation effect may no longer be static. Easingwood, Mahajan, and Muller (1983) relax the assumption that the rate of the imitation effect in the Bass model is constant over time. They propose a non-uniform model by allowing the imitation effect to vary. They use a shape factor to capture the degree of influence from early adopters. When the shape factor takes different values, it allows the diffusion process to accelerate, decelerate, or stays constant as in the Bass model. A different approach, taken by Sharma and Bhargava (1994), compares the influence from the distant past adopters to that of the recent past adopters. By allowing unequal weights on their imitation coefficients, they find the adopters in the recent past to be more influential. They further quantify the decay rate of the influence to be 0.25 a year, suggesting that the dominating imitation influence comes from those who adopted during the past year or two. In the same spirit, I differentiate the search interests of the most recent month from the more distant months. I model the imitation effect of as composed of two parts: the lagged search interests right before the focal time period t, during t-1, and the lagged cumulative search interests from the more distant past, up until t-2. The binary discretization of the imitation effect gives the model the flexibility in capturing the time decay of the early adopter's influence, and at the same time avoids over burdening the model by making the number of such parameters increase with each additional period.

Second, when the diffusion process is limited to a few months, the surge of search interests can easily overlap with the holiday season during November and December, particularly for novelty products and toys whose introductions are often planned ahead for sales towards the end of the year (Kurawarwala and Matsuo 1996). With durables, one could aggregate the data to the annual level to remove seasonal fluctuations (Venkatesan, Krishnan, and Kumar 2004). Without meaningful lagged annual seasonal data, as long range forecasting can often depend on, the seasonality of RLC products becomes entangled with their diffusion process. Twelve out of thirteen products in my data sample had their search interests peaks in November or December (Table 1.2). Another approach, used in the generalized Bass model (Bass, Krishnan, and Jain 1994) is to control for seasonality with dummy variables. However, in the context of search interests, the seasonal swings have been expanding overtime (Figure 1.3), making it insufficient to conduct seasonal control through dummy variables. Radas and Shugan (1998) use a transformed-time method to adjust for seasonality. Their method accelerates the product life cycle during peak seasons. When off-peak, the product ages more slowly along the life cycle. Base on this setting, Peers, Fok, and Franses (2012) modify Radas and Shugan (1998) by estimating the seasonal structure based on the proportion that each month contributes during the seasonal peak. Both approaches come at a cost, as they require a complete diffusion cycle to estimate the monthly proportion and diffusion pattern. As the lifespan of an RLC product is usually much shorter than twelve months, a meaningful forecast may need to start as early as two months after the initial takeoff. Since the complete life cycle used in these methods is no longer an option for us, I instead rely on the total category search interests, deducting those of the focal product, to separate the seasonal movements from the diffusion cycle.

[Insert Figure 1.3 about here.]

Third, the information diffusion process of RLC products happens in multiple cities simultaneously. With digital media, the dependence of the neighborhood effect (Bronnenberg and Mela 2004; Mahajan and Peterson 1979; Redmond 1994) on physical proximity weakens. Garber et al. (2004) argue that the formation of a successful spatial diffusion is the joint work of two forces. The external signal (more *national* or global) after arriving at the focal location, must be spread by the internal force (more *local*) in order to complete the diffusion process. Following this line of consideration, my model takes advantage of the spatial granularity of the data from the GSI and untangles the driving force of diffusion as from dual channels: local and national. From the local channel, the influence can come from direct exposure to product adoptions by family, friends, colleagues or neighbors, and can happen in person, by traditional means such as local TV, or from local friends on social media. From the national channel, the influence can be national TV, widely distributed publishing products such as newspaper – online or printed, and particularly social media. Such a dual-channel framework allows us to simultaneously address two layers of variations in diffusion that are interlaced together: for a product, different cities can inherently have different rates of diffusion; and the contamination of information across cities, especially with social media, can happen beyond the border contingency at a national level. With this setting, I can answer some important managerial questions: 1) which channel has a stronger influence on the rate of diffusion? 2) At the early stages of an RLC, local level signals can be spotty. Does including the national channel provide the wisdom of the “crowd” and improve the forecast performance?

Finally, given my goal to create a model valuable for practitioners, I need the model to be able to start picking up signals at an early stage when the data is still limited – with one or two months. Therefore, I set up the model in a hierarchical Bayesian framework to pool information across different RLC products. With my model, product managers can find a convenient tool, taking advantage of the data from a free platform by Google, enabling them to plan on new products long – by RLC standard long – before the full sales potential is realized.

Model

I model the search interests for product i in city j during month t, $\mathrm{s_{ijt}}$, to follow a normal distribution:

$$\mathrm{s_{ijt}} \sim \mathrm{N}(\mu_{\mathrm{ijt}}, \sigma^2) \tag{1.4}$$

where the mean of the distribution μ_{ijt} is guided by a diffusion process, akin to the Bass model, as a function proportional to the overall search interests potential $\mathrm{m_{ij}}$ deducting the lagged cumulative search interests $\mathrm{S_{ij,t-1}}$ (Equation 1.5).

$$\mu_{\mathrm{ijt}} = \left(\mathrm{r_{ij}} \cdot \mathrm{cs_{jt}} + \mathrm{p_{ij}} + \mathrm{Q^L_{ijt}} + \mathrm{Q^N_{ijt}}\right)\left(\mathrm{m_{ij}} - \mathrm{S_{ij,t-1}}\right) \tag{1.5}$$

where the coefficient of proportionality $\left(\mathrm{r_{ij}} \cdot \mathrm{cs_{jt}} + \mathrm{p_{ij}} + \mathrm{Q^L_{ijt}} + \mathrm{Q^N_{ijt}}\right)$ is composed of four components: $\mathrm{cs_{jt}}$ are the search interests for the overall toy category in city j, during month t. It controls for the seasonality that may exist and otherwise confound with the diffusion process of the RLC product. $\mathrm{p_{ij}}$, similar to the Bass model, is the innovation parameter. $\mathrm{Q^L_{ijt}}$ and $\mathrm{Q^N_{ijt}}$ are the constructs for the imitation effect from the local (superscript L) and national (superscript N) channels, respectively. Further, I specify $\mathrm{Q^L_{ijt}}$ and $\mathrm{Q^N_{ijt}}$ as

$$\mathrm{Q^L_{ijt}} = \mathrm{q^L_{1ij}} \cdot \mathrm{s_{ij,t-1}} + \mathrm{q^L_{2ij}} \cdot \mathrm{S_{ij,t-2}} \tag{1.6.a}$$

$$\mathrm{Q^N_{ijt}} = \mathrm{q^N_{1ij}} \cdot \mathrm{s^N_{ij,t-1}} + \mathrm{q^N_{2ij}} \cdot \mathrm{S^N_{ij,t-2}} \tag{1.6.b}$$

where $\mathrm{q^L_{1ij}}$ and $\mathrm{q^L_{2ij}}$ are the local imitation parameters, and $\mathrm{q^N_{1ij}}$ and $\mathrm{q^N_{2ij}}$ are the national imitation parameters. In Equation 1.6a, which defines the local imitation effect, the first component $\mathrm{q^L_{1ij}} \cdot \mathrm{s_{ij,t-1}}$ on the right hand side represents the imitation effect from the recent past, as a function of the search interests during month t-1, $\mathrm{s_{ij,t-1}}$. The second component $\mathrm{q^L_{2ij}} \cdot \mathrm{S_{ij,t-2}}$ represents the imitation effect from the more distant past, as a function of the lagged cumulative search interests up until month t-2, $\mathrm{S_{ij,t-2}}$. Equation 1.6b defines the imitation effect from the national channel, with $\mathrm{s_{ij,t-1}^{N}}$ representing the national level search interests during month t-1 from the 59 cities excluding city j, and $\mathrm{S_{ij,t-2}^{N}}$ is the lagged cumulative national search interests up until month t-2 from the 59 cities excluding city j.

As discussed earlier, this specification allows us to 1) separate the imitation effects driven by the local channel ($\mathrm{q_{1ij}^{L}}$ and $\mathrm{q_{2ij}^{L}}$) from the national channel ($\mathrm{q_{1ij}^{N}}$ and $\mathrm{q_{2ij}^{N}}$), and 2) capture the rapid nature of RLC products' life cycles by discerning the size of the influence from the recent past ($\mathrm{q_{1ij}^{L}}$ and $\mathrm{q_{1ij}^{N}}$) and the distant past ($\mathrm{q_{2ij}^{L}}$ and $\mathrm{q_{2ij}^{N}}$). Notably, my model is not limited to RLC products; it is able to accommodate the diffusion of durable products as well. One can treat the Bass model as a special case of my proposed model when $\mathrm{r_{ij}} = 0$ (without seasonality control), $\mathrm{q_{1ij}^{L}} = \mathrm{q_{2ij}^{L}}$ (no decay in the imitation effect), and $\mathrm{q_{1ij}^{N}} = \mathrm{q_{2ij}^{N}} = 0$ (no national channel imitation).

From Equations 1.5-1.6, I expect all the parameters to be positive and let them follow lognormal distributions (Clayton and Kaldor 1987). The log-transformed parameters are therefore normally distributed. I denote the set of parameters to be estimated as

$$\lambda_{\mathrm{ij}} = \log\left\{\mathrm{r_{ij}}, \mathrm{p_{ij}}, \mathrm{q_{1ij}^{L}}, \mathrm{q_{2ij}^{L}}, \mathrm{q_{1ij}^{N}}, \mathrm{q_{2ij}^{N}}, \mathrm{m_{ij}}\right\}$$

where I denote each parameter in λ_{ij} as $\lambda_{\mathrm{ij}}^{(\mathrm{k})}$ (k = 1,…,7).[5] Further, I model each parameter in λ_{ij} as a function of product level fixed effect $\alpha_{\mathrm{i}}^{(\mathrm{k})}$, plus city j's profile defined as a linear combination of the city's demographic variables $\mathrm{Z_j}$ and a city level random effect $\mathrm{u_j^{(k)}} \sim \mathrm{N}\left(0, \sigma_{\mathrm{u(k)}}^{2}\right)$:

[5] $\lambda_{\mathrm{ij}}^{(1)} = \log(\mathrm{r_{ij}})$, $\lambda_{\mathrm{ij}}^{(2)} = \log(\mathrm{p_{ij}})$, $\lambda_{\mathrm{ij}}^{(3)} = \log(\mathrm{q_{1ij}^{L}})$, $\lambda_{\mathrm{ij}}^{(4)} = \log(\mathrm{q_{2ij}^{L}})$, $\lambda_{\mathrm{ij}}^{(5)} = \log(\mathrm{q_{1ij}^{N}})$, $\lambda_{\mathrm{ij}}^{(6)} = \log(\mathrm{q_{2ij}^{N}})$, $\lambda_{\mathrm{ij}}^{(7)} = \log(\mathrm{m_{ij}})$.

$$\lambda_{ij}^{(k)} = \alpha_i^{(k)} + Z_j\beta^{(k)} + u_j^{(k)} + \varepsilon_{ij}^{(k)} \tag{1.7}$$

where $\varepsilon_{ij}^{(k)} \overset{\text{i.i.d}}{\sim} N\left(0, \sigma_{(k)}^2\right)$. Following Katona, Zubcsek, and Sarvary (2011), I include the population density ($\mathrm{POPDEN_j}$), the percentage of male population ($\mathrm{MALE_j}$), median age ($\mathrm{AGE_j}$), home ownership ($\mathrm{HOME_j}$), and median household income ($\mathrm{INCOME_j}$) as the city level demographic covariates. With

$$\beta^{(k)} = \begin{bmatrix} \beta_1^{(k)} \\ \vdots \\ \beta_5^{(k)} \end{bmatrix},$$

and $\mathrm{Z_j} = \left[\mathrm{POPDEN_j}\ \ \mathrm{MALE_j}\ \ \mathrm{AGE_j}\ \ \mathrm{HOUSE_j}\ \ \mathrm{INCOME_j}\right]$, Equation 1.7 becomes:

$$\begin{aligned} \lambda_{ij}^{(k)} &= \alpha_i^{(k)} + \beta_1^{(k)} \cdot \mathrm{POPDEN_j} + \beta_2^{(k)} \cdot \mathrm{MALE_j} + \beta_3^{(k)} \cdot \mathrm{AGE_j} + \beta_4^{(k)} \cdot \mathrm{HOME_j} \\ &+\beta_5^{(k)} \cdot \mathrm{INCOME_j} + u_j^{(k)} + \varepsilon_{ij}^{(k)} \end{aligned} \tag{1.8}$$

The $\beta^{(k)}$ coefficients represent the effects of city demographics on the parameter $\lambda_{ij}^{(k)}$. For instance, $\beta_5^{(2)}$ reflects the effect of $\mathrm{INCOME_j}$ on logged p_{ij}, the innovation parameter. If $\beta_5^{(2)}$ is positive and significant, it would suggest that a higher median household income is associated with a higher innovation parameter.

Essentially, I have a multilevel model where the parameters for product i in city j from the first level are the dependent variables of the second level linear regressions with product level fixed effects and city level random effects.

I use a hierarchical Bayesian (HB) framework to estimate the diffusion model specified in Equations 1.4-1.8. Besides the typical advantage of providing accurate inferences of the unobserved heterogeneity in the parameters (Neelamegham and Chintagunta 1999), the HB framework in my model becomes especially useful in forecasting the diffusion process of a holdout product. Considering the importance of timely forecasting an RLC product’s cycle, my forecast is tailored to begin early – right after takeoff. Of course, the catch is at that point of time, there is very limited useable data available for the new product. This is when an HB framework reigns supreme, because it allows us to pool information from the data already available for the calibration products. As the new product's history gradually increases, the shrinkage starts to kick in and the influence from the data of the calibration products slowly eases away. Moreover, although this is not relevant to this paper per se – the HB pooling can be location specific. Let's imagine a scenario where the data for the new product is only available in some but not all the locations. The HB framework can borrow information from the other locations in the sample.

To summarize, the parameters and hyperparameters in the model are

$$\left\{\sigma^2, \sigma^2_{(k)}, \sigma^2_{u(k)}, \lambda^{(k)}_{ij}, \alpha^{(k)}_{i}, \beta^{(k)}_{1}, \beta^{(k)}_{2}, \beta^{(k)}_{3}, \beta^{(k)}_{4}, \beta^{(k)}_{5}\right\}$$

where i = 1, …, I (the number of products) and k = 1, …, 7. I use non-informative priors – normal distributions with mean 0 and variance 100 for the parameters defined on the real domain, and inverse-gamma distributions for those defined on the positive domain. Following Neelamegham and Chintagunta (1999) and Spiegelhalter (1998), I use a Directed Acyclic Graph (DAG) to illustrate my hierarchical Bayes model in Figure 1.4. The circles represent the parameters to be estimated and the squares represent the data. The solid arrows show the deterministic relationships and the dashed arrow indicates a stochastic link. The unlinked entities are independent from each other in my model.

[Insert Figure 1.4 about here.]

Benchmark Models

I compare my proposed model with two benchmark models. All three models are estimated in the same hierarchical Bayesian framework. The only difference is the diffusion model defined in Equations 1.5-1.6. The first benchmark model is based on the Bass model

(1969), where $\mathrm{\mu_{ijt}}$, the mean of the distribution of search interests $\mathrm{s_{ijt}}$, is as follows:

$$\mathrm{\mu_{ijt}} = \left(\mathrm{p_{ij}} + \mathrm{q_{ij}S_{ij,t-1}}\right)\left(\mathrm{m_{ij}} - \mathrm{S_{ij,t-1}}\right) \tag{1.9}$$

Notably, there is no seasonal control, and the imitation parameter $\mathrm{q_{ij}}$ is both local (without the national channel) and static (does not allow decay).

The second benchmark model extends the Bass model with seasonal adjustment $\mathrm{cs_{jt}}$ and relaxes the static assumption. The model becomes:

$$\mathrm{\mu_{ijt}} = \left(\mathrm{r_{ij}} \cdot \mathrm{cs_{jt}} + \mathrm{p_{ij}} + \mathrm{q_{1ij}s_{ij,t-1}} + \mathrm{q_{2ij}S_{ij,t-2}}\right)\left(\mathrm{m_{ij}} - \mathrm{S_{ij,t-1}}\right) \tag{1.10}$$

Comparing this model to my proposed model, I can get a glimpse of whether allowing the national channel to influence the imitation effect can produce better predictions in the forecasting model.

I fit my proposed model and the benchmark models using the SAS procedure MCMC (Markov Chain Monte Carlo methods). To improve the efficiency of the MCMC sampler, I estimate the model in two steps. First, I estimate the model for each product in the calibration sample separately and produce the posterior distributions of the parameters for each product without information pooling across products. In the second step, I apply the means of the posterior distributions obtained from the first step as the initial values of my proposed model. This two-step approach significantly reduced the time needed for the sampling process to reach convergence.

Forecast Procedure

After calibration, I implement the model to forecast the diffusion life cycles of two holdout products: Doc McStuffins Mobile Cart and J-Animals. Following the convention of Bayesian dynamic modeling, I treat the future search interests ($\mathrm{s'_{ijt}}$ for product i in city j during month t) as a variable with missing values and create an updated forecast posterior for one month at a time. Using the posterior sampling draws of the parameters λ'_{ij} estimated from my calibration sample and the holdout product until t-1, I can produce the posterior inference of s'_{jit}, a procedure that I describe using the following integration:

$$p\left(s'_{ijt}\middle|\mathrm{Data},\mathrm{Data}'_{t-1}\right)=\int p\left(s'_{ijt}\middle|\lambda'_{ij}\right)p\left(\lambda'_{ij}\middle|\mathrm{Data},\mathrm{Data}'_{t-1}\right)d\lambda_{ij} \tag{1.11}$$

where “Data” represents the data from the calibration sample and Data'_{t-1} represents the data from the holdout until t-1. On the left hand side, $p\left(s'_{ijt}\middle|\mathrm{Data},\mathrm{Data}'_{t-1}\right)$ is posterior of the forecast. On the right hand side, $p\left(\lambda'_{ij}\middle|\mathrm{Data},\mathrm{Data}'_{t-1}\right)$ is the posterior distribution of λ'_{ij}, estimated from the data. Essentially, the integration calculates the posterior distribution of s'_{ijt} as the weighted average over the posterior distribution of the diffusion parameters λ'_{ij}.

To ensure I create a true forecast scenario, as in reality without the benefit of hindsight, I produce the posterior forecast of $p\left(s'_{ijt}\middle|\mathrm{Data},\mathrm{Data}'_{t-1}\right)$ step-by-step, one month at a time (Figure 1.5). Specifically, I start the forecast at month t = 3 after takeoff, at which point the only new data available is search interests during months 1 and 2. With meagerly two observations for each city, I forecast the search interests in month 3. At this point, given the extremely limited amount of information from the holdout product, the forecast would rely heavily on the calibration sample. Then after I create the forecast for month 3, I update the posterior distribution of the parameters $p(\lambda'_{ij}|\mathrm{Data},\mathrm{Data}'_{t-1})$ using the forecast I have just created and forecast month 4. One month at a time, with these steps, I eventually forecast the search interests for the holdout products up until eight months after the takeoff.

Results

I start by comparing the goodness-of-fit performance of the proposed model to the benchmarks. Model 1 is a direct adaption from the Bass model, which does not include seasonality control, the decay of the imitation effect, or informative spillover from the national channel to the imitation effect. Model 2 falls between Model 1 and Model 3, my proposed the model. It extends Model 1 by including a search trend in the toy category to control for seasonality and differentiating the imitation effects from the recent past (t-1) and the distant past (cumulative until t-2) in the diffusion process. Model 3 is my proposed model. Comparing to Model 2, it distinguishes the two sources of information channels, local and national, for the imitation effect. Following Neelamegham and Chintagunta (1999), in Table 1.2, I compare the performances of the three models using the overall – for all thirteen calibration products across the 60 cities – the root mean square errors (RMSEs) and mean absolute errors (MAEs).[6] By both performance metrics, Model 3 performs the best. Comparing to Model 1, Model 2 improves the performance substantially, improving RMSE from 7,655 to 5,332 and MAE from 1,540 to 1,077 (both a 30% decrease). Model 3, adding the dual information channel imitation effect, improves further compared to Model 2, with RMSE dropping from 5,332 to 5,287 (a 31% decrease compared to Model 1) and MAE from 1,077 to 1,006 (a 35% decrease compared to Model 1).

[Insert Table 1.2 about here.]

Figure 1.6 shows the actual vs. predict monthly search interests for each product. The dots represent the actual search interests and the line the predictions from my proposed model.[7] I have adjusted the vertical axis scales with the peak of each product's cycle. Their magnitudes are not visually comparable. Overall, despite the conspicuously different paths and magnitudes of

[6] $\mathrm{RMSE} = \sqrt{\frac{\sum_{i,j,t}\left(\widehat{s_{ijt}} - s_{ijt}\right)^2}{I \times J \times T}}$ and $\mathrm{MAE} = \frac{\left|\widehat{s_{ijt}} - s_{ijt}\right|}{I \times J \times T}$ where $\widehat{s_{ijt}}$ is the predicted value of s_{ijt} from the model.

[7] Since the posterior mean search interests predictions from the model $\overline{s_{ijt}}$ is at product-city-month level, I aggregate it across the cities for each product to produce the predicted search interests in Figure 1.6.

the search interests in the RLCs of the products, the proposed model proves capable; it captures the rally, peak, and downfall of most products quite well. By distinguishing the recent past and distant past imitation effects, the proposed model has generally avoided the overshooting issue[8] common among forecasting models.

[Insert Figure 1.6 about here.]

Next, I examine the estimated seasonality, innovation, market potential, and imitation parameters, interpret how the city demographic variables are associated with them, and demonstrate the results from applying my model to forecast the search interests of two holdout RLC products.

Seasonality, Innovation, and Market Size

In Table 1.3, I report the mean posterior estimates of the seasonality coefficients r_{ij}, the innovation parameters p_{ij}, and the marketing penetration potentials m_{ij}. The parameters reported are the averages across the cities for each product and log-transformed. The pattern in the seasonality coefficients $\overline{\log(r_\imath)}$ shows that the search interests of more popular products tend to be less associated with the category-level seasonality movement, with a strong negative correlation between these estimates and the average search interests for each product from Table 1.1 (correlation = -0.627). This is perhaps not surprising, as the search interests for more popular products tend to be more driven by the interests in themselves than the seasonal interests shift in general. The higher the innovation parameter estimates $\overline{\log(p_\imath)}$, the more the early search interests arises without previous "adopters". Such positive association often reveals as a steep ramp up in the early months after takeoff. The highest innovation parameters belong to the fidget

[8] Often manifested as: the carryover effect from the peak in the data pushes the predicted peak one or more periods delayed after the actual peak.

spinners, the Luvabella dolls, and the Fingerlings monkey. And the lowest ones are the adult coloring books, hideaway pets, and emoji joggers. And finally, the market penetration potential estimates $\overline{\log(m_\imath)}$ are closely associated with the average search interests in Table 1.1 (correlation = 0.639). Such relationship reflects a combination of the size of the overall RLC and the rates of ramp up and decline. A relatively "flat" life cycle would indicate a high potential even if the peak may not be as high. For example, the Fingerlings monkey and self-balancing scooters have similar average search interests at 497 and 484, respectively. The flatter life cycle of self-balancing scooters (peak at 1,554, lower than 1,829 of Fingerlings monkey) would suggest an eventual higher market penetration potential.

[Insert Table 1.3 about here.]

Imitation Effect

In Table 1.4, I report the mean posterior estimates of the local imitation parameters of the recent past q^{L}_{1ij} and distant past q^{L}_{2ij}, and the national imitation parameters of the recent past q^{N}_{1ij} and distant past q^{N}_{2ij}. The parameters reported are the averages across the cities for each product and log-transformed. Namely, I report $\overline{\log(q^{L}_{1\imath})}$, $\overline{\log(q^{L}_{2\imath})}$, $\overline{\log(q^{N}_{1\imath})}$, and $\overline{\log(q^{N}_{2\imath})}$ in columns 1, 2, 4, and 5, respectively. For example, the fidget spinner's log-transformed national channel imitation coefficients are -11.14 for the recent past (column 4) and -17.55 for the distant past (column 5). The ratio between the two is 476.19 (column 6), suggesting much stronger imitation effect from the recent past. In column 3 and 6, the ratios between recent past and distant past imitation parameters for the local and national channels, $\frac{\exp\left(\overline{\log(q^{L}_{1\imath})}\right)}{\exp\left(\overline{\log(q^{L}_{2\imath})}\right)}$ and $\frac{\exp\left(\overline{\log(q^{N}_{1\imath})}\right)}{\exp\left(\overline{\log(q^{L}_{2\imath})}\right)}$ respectively, suggest an almost universal decay of the imitation effect over time and most of them very strong (with the exception of Magic Tracks' national channel).

[Insert Table 1.4 about here.]

Recall that Sharma and Bhargava (1994) find an average *annual* decay rate of 0.25 for the durables, translating into a ratio of four compared to the numbers in columns 3 and 6. My ratios, at a *monthly* rate, have twelve out of thirteen cases above four, most of them many times above four, for either the local or the national channel. These results suggest that compared to conventional durables, the RLC products in my sample, plausibly RLC products in general, have a much higher decay rate over time in the imitation effect. Many of these ratios are above ten and some measured by hundreds, which suggests in those cases, the cycles move so rapidly that only the most recent month's search interests are relevant as far as the imitation effect is concerned.

Given that the dominant imitation effect comes from the recent past at month t-1, I focus on the comparison between the recent past effects from the national and local channels and report the ratio $\frac{\exp\left(\overline{\log(q_{11}^{N})}\right)}{\exp\left(\overline{\log(q_{11}^{L})}\right)}$ in column 7 of Table 1.4. I observe a very wide range in the ratios, from the lowest at 0.03 for the adult coloring books to the highest at 169.27 for the darn yarn. A high ratio means the imitation effect from national channel is dwarfing that from the local channel. Moreover, the opposite is true for a low ratio. The simulation conducted by Garber et al. (2004) can explain the possible process behind such different ratios. Specifically, social media doesn't have boundaries; it is conceivable that the national channel can "light up" a new trend almost instantly. When the national channel dominates the local, it usually means the RLC is heavily "top-down" and lack of local imitation support after the initial interests, which manifests as a rapid upsurge accompanied by a sharp decline in the life cycle – even by the RLC product's standard. On the contrary, when $\frac{\exp\left(\overline{\log(q_{11}^{N})}\right)}{\exp\left(\overline{\log(q_{11}^{L})}\right)}$ is small, it often suggests the new RLC product has attracted imitation support from the local channel and as a result, the life cycle may become more sustainable. The Google Trends interests for the darn yarn and adult coloring books appear to support this interpretation of the implications from $\frac{\exp\left(\overline{\log(q_{11}^{N})}\right)}{\exp\left(\overline{\log(q_{11}^{L})}\right)}$, where the life cycle of adult coloring books both rose and declined at a slower pace (Figure 1.7).

[Insert Figure 1.7 about here.]

City Demographics

In Table 1.5, I present the estimated posterior mean coefficients for the city-specific demographic variables. These results, estimated from Equation 1.8, allow us to gain further insights into how the profile of a city may be associated with the parameters in the diffusion process. For example, what is the association between the median household income and the innovation parameter? Out of the 35 $\beta^{(k)}$ coefficients, 16 are significant.[9]

[Insert Table 1.5 about here.]

The estimates in the first column are for the seasonality control r_{ij}. Three demographic variables, median age (mean = -0.13), home ownership (mean = 0.12) and median household income (mean = 0.17) are significant. They suggest that 1) understandably, cities with an older population composition are less likely to have seasonal search interests for the toy category. 2) Home ownership, representing the ratio of stable city dwellers, and median household income, representing the wealth and disposable income that can be spent on discretionary products such as toys, are both positively associated with seasonality.

Column 2 presents the estimates on the innovation parameter p_{ij}. Population density (mean = 0.23), percentage of male population (0.08), median age (-0.12), and household income (0.16) are significant. A higher population density is connected with a stronger innovation effect,

[9] I use “significant” or “significance” to refer to the Bayesian results when the 95% confidence interval of the posterior estimate does not cover zero.

likely holdout by the association between a high population density and large metropolitan areas. And large cities tend to have more concentrated earlier adopters and trend setters (Johnson 2011). It also appears that a higher percentage of male population is associated with more early search activities among the RLC products in my sample, possibly holdout by the gender difference in the decision-making processes (Venkatesh, Morris, and Ackerman 2000). Unsurprisingly, a higher median age is associated with a lower innovation effect. Finally, higher median household tends to be positively related to the innovation effect.

The estimates for the imitation parameters q_{1ij}^{L}, q_{2ij}^{L}, q_{1ij}^{N}, and q_{2ij}^{N} are sparsely significant, suggesting that the demographic variables I are able to include do not have many linkages to the imitation effects. Much of the variations at the city level go to the unobservable. Particularly, as discussed earlier, since the near past imitation effects q_{1ij}^{L} and q_{1ij}^{N} tend to dominate in strength, the more practically meaningful demographic variable results would be related to them. I have only two demographic variables significant: home ownership (mean = 0.13) and median household income (mean = 0.15) on the recent past local channel q_{1ij}^{L}, which can be interpreted as a stability of residency and disposable income explanation. There is also a hint of suggestion, from home ownership and median household income carrying negative coefficients on q_{2ij}^{N} (mean = -0.11) and q_{2ij}^{L} (mean = -0.32) respectively, that in the cities where people have stable residency and more disposable income, the distant past imitation effect decays faster.

Recall that my dependent measure, the search interests normalized on a per 100,000 capita basis, represents the market penetration. For the total market penetration potential m_{ij}, the population density carries a negative coefficient (mean = -0.15), suggesting for more densely populated cities (tend to be large cities), it is more difficult to gain a higher rate of market penetration. Intuitively, this is plausible considering bigger cities usually have more distractions competing for the residents' attention. Moreover, home ownership is negatively associated with market penetration potential (mean = -0.23) and median household income is positive (mean = 0.06).

Forecasting

A direct benefit of my model, estimated in an HB framework, is its ability to start forecasting when a new RLC product is still in the early stages of its life cycle. Specifically, my model needs only two data points after the takeoff to forecast the life cycle of a product. In this section, I present the forecast results of two holdout products Doc McStuffins Mobile Cart (Table 6, Panel A) and J-Animals (Table 1.6 Panel B), both of which had takeoff month in July 2017 as identified using Equation 1.2. I also compare the holdout performance of the proposed model to that of the two benchmarks.

For each holdout product, I report three forecasts, conducted using different starting months: two months, three months, and four months after the takeoff month. As one can imagine, the earlier month the forecast starts, the less information the forecasting model can rely on from the holdout product. Instead, it has to lean more towards pooling information from the calibration sample to produce the forecast. As I gradually have more months of data from the holdout product, the forecast tends to get more shrinkage towards it. I forecast a total of eight months out of the observation window of two, three, or four months.

[Insert Table 1.6 about here.]

In Table 1.6, I report the MAEs of the forecasts, calculated as the absolute value of the difference between the actual and forecast search interests of each product during each month. Overall, my proposed model consistently outperforms the benchmark models by big margins, with the exception of J-Animals forecast at three months after the takeoff. Comparing to Model 1, the biggest improvement from my model happens when the data is the scantiest, forecasting at merely two months after the takeoff. The forecasts from my model have a 94% decrease in average MAE for Doc McStuffins Mobile Cart (592 vs. 10,399) and a 53% decrease for J-Animals (498 vs. 1,070). Comparing the MAEs on a month-by-month basis, Model 3 outperforms Model 1 in 44 out of 48 contrasts, with the difference significant at $p < 0.0001$ in a two-tail paired t-test. Considering that Model 1 is a more parsimonious model with fewer parameters, it is surprising that the proposed model beats it when the available data points are few. Indirectly, it suggests that my extensions to the Bass model, namely the seasonal control, the decay of the imitation effect, and the dual-channel local/national influence on the imitation effect, are picking up the true signals pertinent to the diffusion of the RLC products.

An unanticipated yet interesting result is the inconsistent performance of Model 2, which extends Model 1 to include the seasonal control and decay of the imitation effect but without adding the national channel influence. For Doc McStuffins Mobile Cart, its performance is much better than Model 1, worse than but at the same magnitude as Model 3 when forecasting two or three months after takeoff (From Models 1, 2, and 3, the average MAEs are: 10,399 vs. 822 vs. 592; and 4161 vs. 925 vs. 794). Under other conditions and for J-Animals, however, the forecasts from Model 2 are much worse than either Model 1 or Model 3. The wildly inconsistent performance of Model 2 suggests that the national channel, being from dozens of cities, can be crucial in stabilizing the imitation signal and improve forecast output.

Overall, the proposed model produces significantly improved forecast accuracy compared to the benchmarks. It is evident that when implemented in the field, my model can yield highly useful marketing intelligence, especially at a very early stage after the takeoff, for managing the life cycle of RLC products in a wide spectrum of business functions, from manufacturing, inventory, supply chain management, to social media and marketing mix planning.

Concluding Remarks

Enabled by social networks, a constantly recurring theme in recent years is the rise and fall of new products that complete their life cycles within a short period of a few months. These products, such as the fidget spinner, create hundreds of millions of dollars in revenue every year despite their short-lived popularity – they are very successful by any new product standard. Yet, the volatile nature of their RLC makes them a challenge for practitioners to plan ahead. In this paper, I extend the Bass model to develop a framework, estimated in a hierarchical Bayesian setting, to track and forecast the life cycles of RLC products.

Compared to the benchmark, a direct adaption from the Bass model, my model performs well in in-sample goodness-of-fit, lowering the MAE by 35%. More remarkably, when applied to forecast the life cycles of two holdout RLC products, my model shrinks the MAEs from the benchmark by 94% and 53%, respectively. The best performance in the holdout forecast happens when the lead time window for forecasting is the shortest at only two months of observations after the takeoff, making my proposed model a particularly useful tool for generating early marketing intelligence.

I attribute the superior performance and practical value of my model to the following components integrated into the diffusion model. First, I control for the seasonality in the life cycle by using the product category level search interests. This may sound like a mundane extension from the Bass model, but it is essential for RLC products, whose life cycles run the complete A to Z within a year, very often leaving the sales or search signals extensively confounded with the holiday season. Second, I allow the imitation effect to decay, distinguishing the influence of the recent past from the distant past. I expect this extension to make a difference for RLC products. And it does. Different from durables where the imitation effect from existing adopters may last for years – and even when the decay of the imitation effect is considered, it's measured by years (Sharma and Bhargava 1994) – the RLC products in my data demonstrate a very fast decay of the imitation effect. For all but one of the products, the most recent search interests during month t-1 are the dominating forces in the imitation effect. Not surprisingly, such a decay is perhaps the fundamental reason why RLC products tend to run through the life cycle so quickly. Third, I allow two channels to influence the imitation effect: local and national. The relative strengths of the two channels vary by products. The results show that a relatively flat life cycle – by RLC standard – is associated with a stronger strength in the local channel. I interpret this as an indication that in order for the trends to sustain, interests from the local channel have to pick up. Fourth, the HB paradigm I use for estimating the model is vital for generating forecasts early in the life cycle when input from the holdout sample is limited. Bayesian pooling lets the model borrow information from the calibration sample. As the observed periods of the holdout gradually increase, more shrinkage automatically kicks in, allowing the holdout itself to have added weight in the estimated posterior results. Fifth, my model leverages on the search data from the GSI, a relatively open platform by Google. While sales data may be the Holy Grail for measuring product diffusion, they are also difficult to acquire timely. The search data I use, shown to be closely related to sales (Hu, Du, and Damangir 2014; The Economist 2017), are updated almost on a real time basis and contain great granularity spatially.

My analysis of the city demographics adds an additional layer of output that can be informative in explaining how spatial factors influence the diffusion parameters. The results reveal that median age, whenever significant, appears to be consistently negatively associated with the diffusion parameters. Given the novelty and rapid nature of the RLC products, this makes sense. The results also show that a higher level of median household income to be associated with a quicker decay in the imitation coefficient. With the demographic variables available to us, I do not have a full set of profile-generating demographics that practitioners can potentially use for geo-targeting purpose. I caution this as a limit.

Last but not least, I acknowledge that a few caveats exist, and some provide potential directions for future research. First, when selecting the cities in my study, I choose a population threshold of 300,000. This is a relatively high threshold for a Census designated place – in the end I have 60 cities in my sample, accounting for 17% of the U.S. population. Ideally, I would like to set this threshold lower and include a higher share of the population. However, not every successful RLC product peaks at the same level as the fidget spinner, therefore leading to data sparsity with lesser successful products in small cities. The potential issue arises from the representativeness of the population in my sample. For instance, the rural and suburban areas tend to have lower population density. As a result, out data sample likely doesn't cover the full range of possible population density. Therefore, some of the results related to the demographic variables may not be amenable to extrapolation into the range out of the sample. That said, since my model can easily scale up to many more cities, using a high threshold to fight off data sparsity itself isn't a modeling concern. Perhaps most importantly, a high level of consistency exists between the RLCs aggregated from my 60-city sample and the U.S. overall (correlation = 0.999), which means the aggregated diffusion parameters at the product level should be reasonably representative.

Second, since I extracted my main data from the GSI, I can only observe the manifestation of social influences as reflected in the information acquisitions in the form of online searches. I do not have direct observations of the RLC products related activities on social media platforms such as YouTube, Twitter, or Facebook. Having such information would certainly help enrich the findings, especially with the possibility of attributing the search outcomes to various types of media, the information they portray, and how the messages spread in a social network (Bakshy et al. 2012).

Third, given the data available, I limit my research to studying physical products – all the RLC products in my data are. The framework, however, is extendable to estimate the life cycles of more intangible things, or shall I call them RLC activates. For instance, the ice bucket challenge,[10] aiming at advocating for donations to the ALS[11] Association, involves dumping a bucket of ice and water over a person's head. It came out of nowhere and spread like wildfire on social media in the summer of 2014, successfully collecting $115 million in donations to the ALS Association (Rogers 2016). The diffusion process is extremely similar to the RLC products I study. I speculate that such activities, with an even stronger social component to it than physical products, maybe reveal to be more reliant on the national channel than my study. Besides activities, digital contents such as YouTube videos, Twitter Tweets, Instagram photos can all suddenly become "trending" and run through an even more rapid life cycle, measured by days or even hours.

Fourth, to estimate the RLC, I use a two-step process. First, I use a takeoff model (Golder and Tellis 1997) to define the period of observations. Then I extend the Bass model to capture the diffusion cycle. This is less than elegant a solution, but probably unavoidable. As I have discussed earlier, prior to reaching the takeoff point, because of the sporadic nature of the signals in search, there isn't a reliable way – at least not in the Bass-style diffusion paradigm – to identify and extract these erratic signals. Without reaching the critical mass, many of these pre-takeoff signals may stay idle for an unpredictably long period of time and even die out without

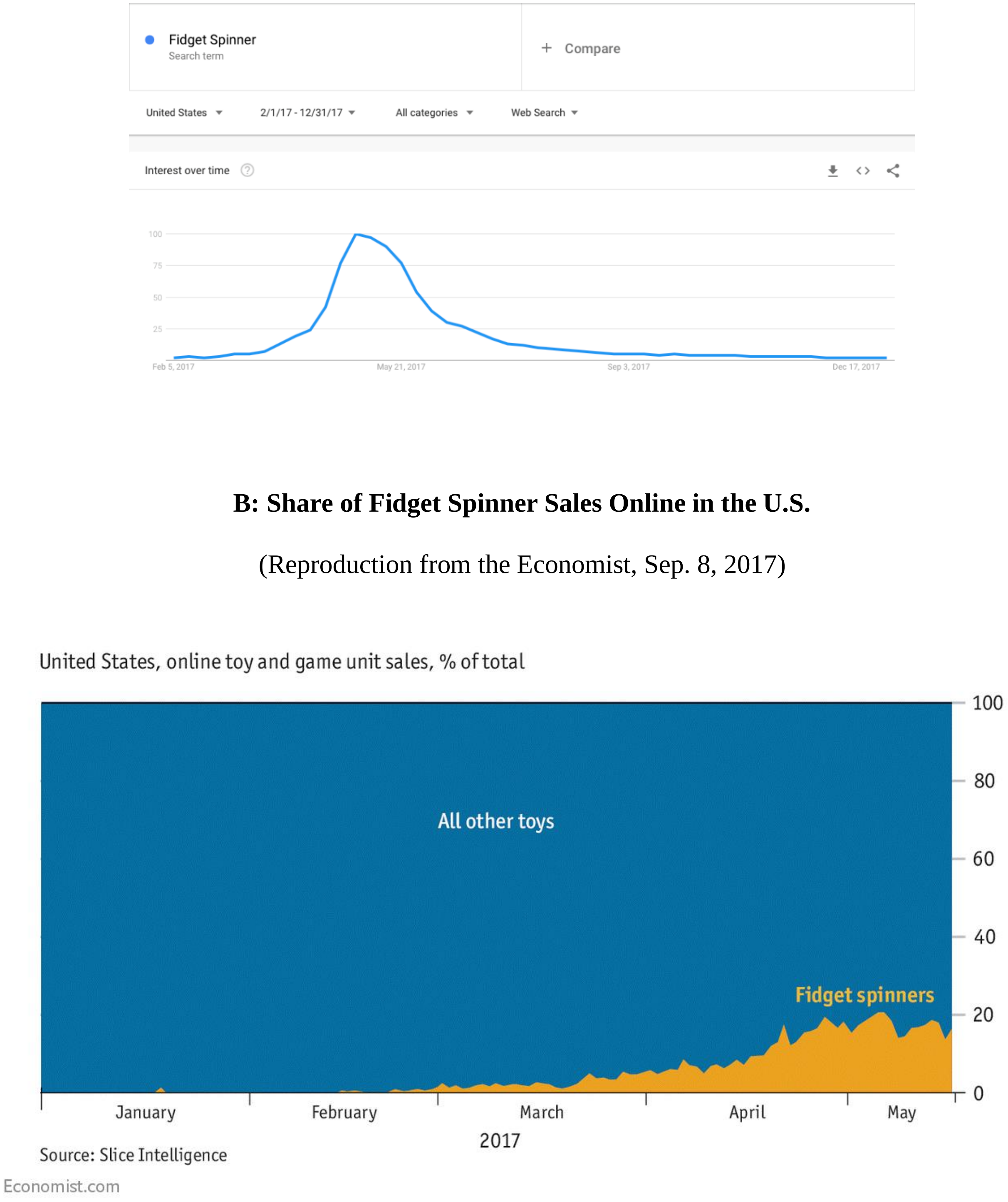
[10] http://www.alsa.org/fight-als/ice-bucket-challenge.html
[11] ALS: amyotrophic lateral sclerosis, or Lou Gehrig's disease, a progressive neurodegenerative disease.

ever really taking off (Valente 1996). I identify the recent development in machine learning models as a potential solution to this. My data is granular: it contains thousands of places. My data is also sparse: the search interests in small cities are like occasional flashes of fireflies; they are not a constant flow but a slow drip. If a machine learning model is able to aggregate the sparse signals across thousands of Census designated places, including the small ones with only a few thousand residents, the quality of the output signal may improve significantly. At the end of the day, the bottom line is whether a well-constructed machine learning model can take advantage of the spatial granularity of the data to overcome its sparseness. This is a future research direction that I are pursuing.

Fifth, like the Bass model, the perspective of my model is reactive. That is, the goal of the proposed model is to track and forecast a product already in the market, sometimes for a while. While environmental factors such as social pressure and competition may play crucial roles in the diffusion of new RLC products, the product itself is obviously an indispensable part of the successful – or failed – life cycle. Therefore, adding coded characteristics of the product can potentially turn the model from reactive to proactive. However, RLC products are different from automobiles (Du, Hu, and Damangir 2015), and their characteristics are usually very nonstandard (e.g., imagine finding comparable commonalities between the fidget spinner and. the adult coloring books). The solution cannot be simply to extract features. Instead, it may again involve relying on machine learning to abstract the hidden traits that make a product fulfill its RLC potential. If successful, the method can be instantly applied to the fast fashion industry by companies such as Zara (Choi 2013), to not only follow successful trends, but also identify the designs that are more likely to become the *next* trend.

I hope my model, given its tested performance and practical applications, can become a useful tool in the field and inspire future research on the increasingly rapid pace of information diffusion through the massive social networks.

Figure 1.1: The Life Cycle of the Fidget Spinner



A: Google Search Index of “Fidget Spinner” in the U.S.

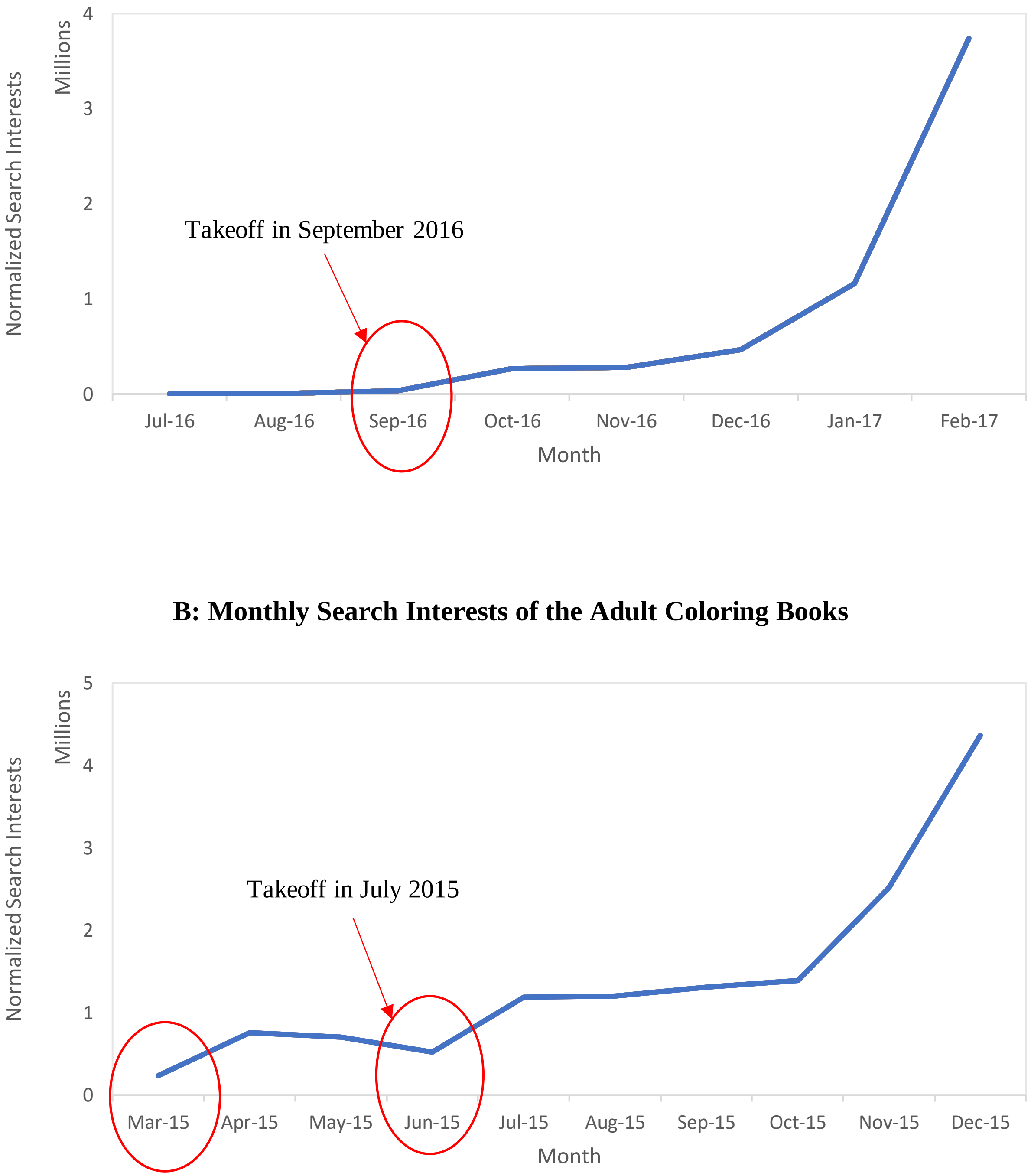
B: Share of Fidget Spinner Sales Online in the U.S.

(Reproduction from the Economist, Sep. 8, 2017)

Source: Slice Intelligence

Economist.com

Figure 1.2: Takeoff Threshold Rule



A: Monthly Search Interests of the Fidget Spinner

B: Monthly Search Interests of the Adult Coloring Books

C: Takeoff Threshold Curve

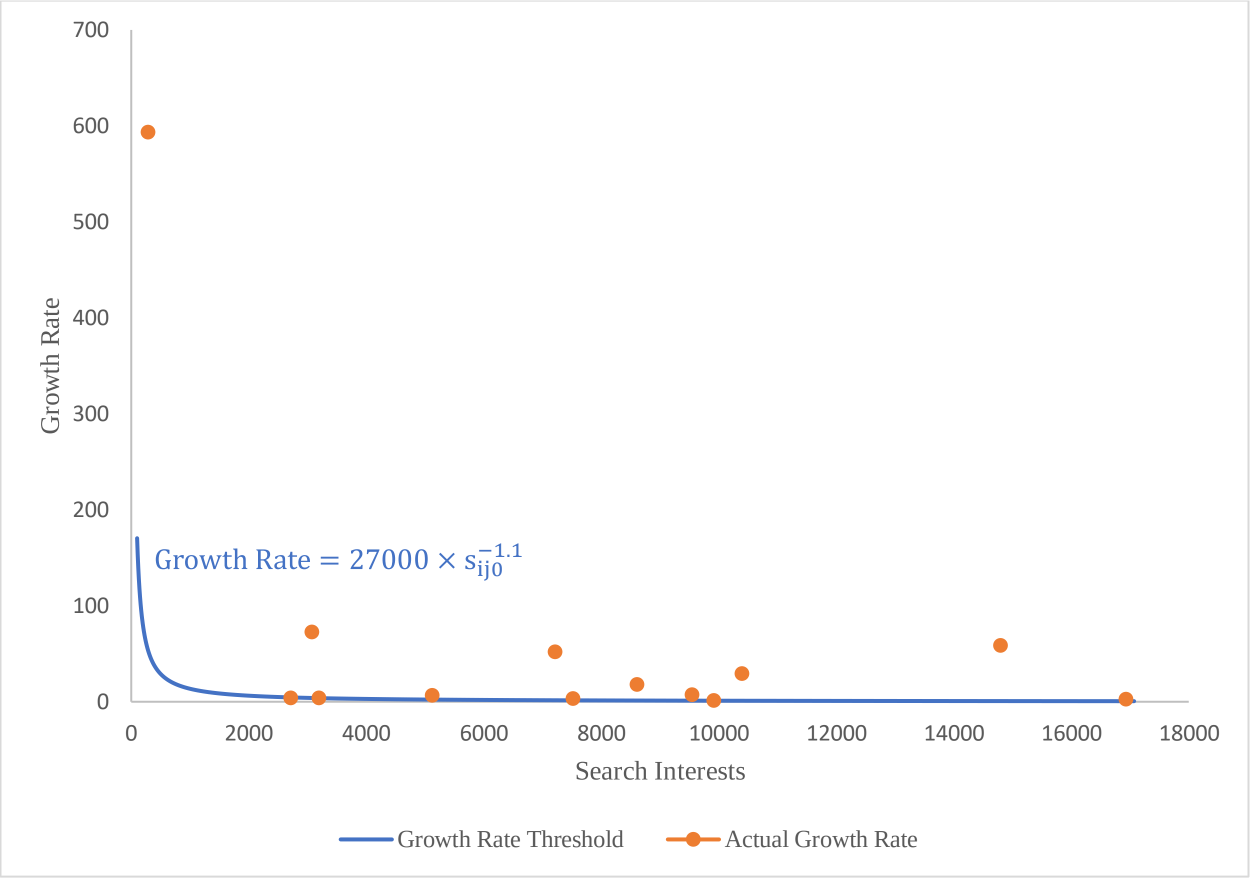


Figure 1.3: Search Interests for Toys in the U.S.

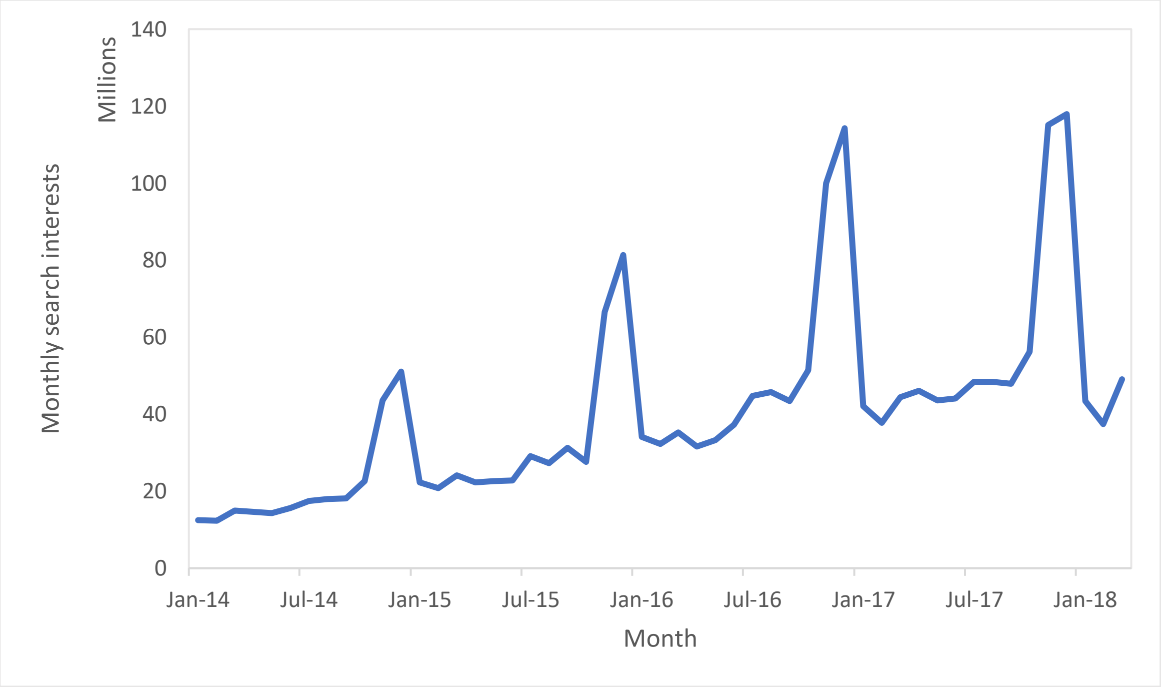
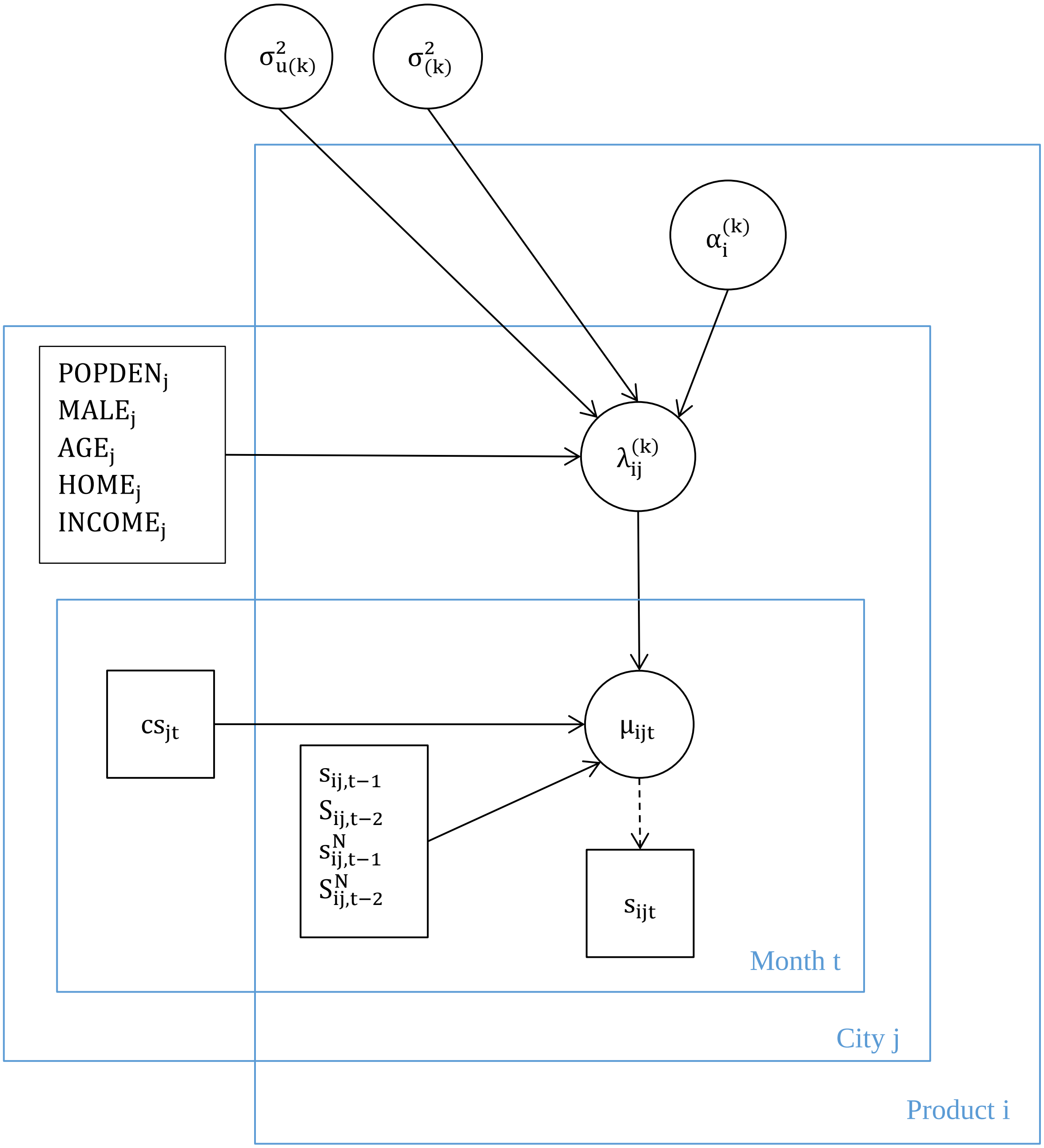


Figure 1.4: Specification of the Proposed Hierarchical Bayes Model



Notes: k = 1, …, 7

Figure 1.5: Forecasting Holdout Sample in Month t

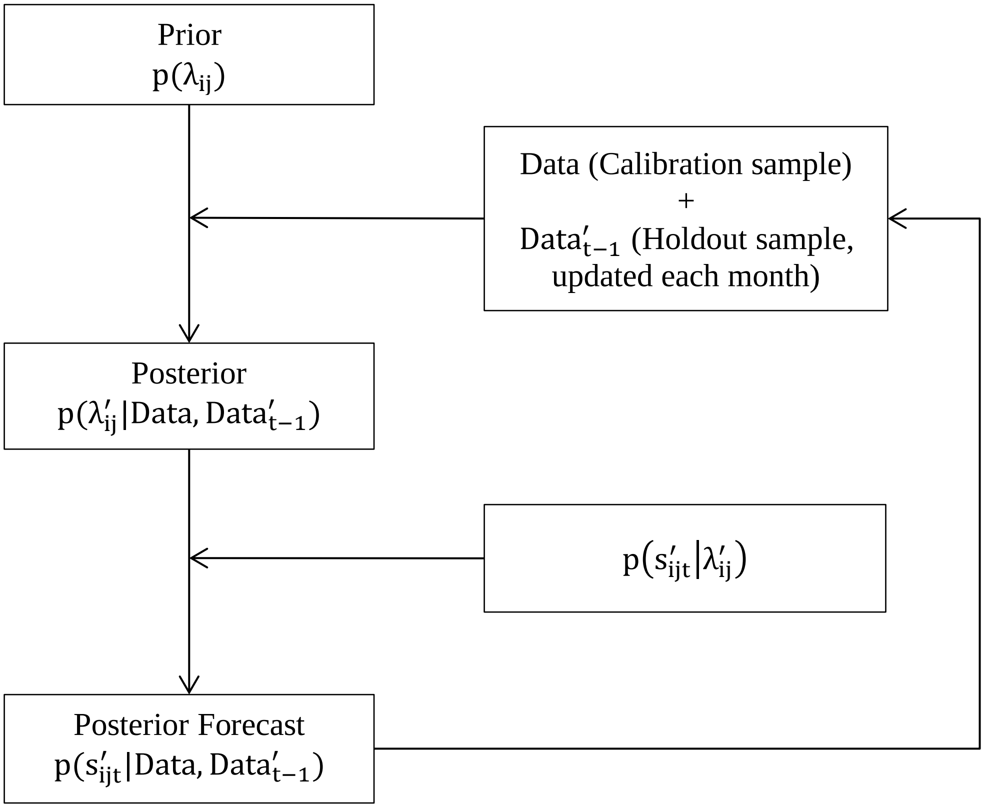

Prior
$p(\lambda_{ij})$
Data (Calibration sample)
+
Data'_{t-1} (Holdout sample, updated each month)
Posterior
$p(\lambda'_{ij}|\mathrm{Data}, \mathrm{Data}'_{t-1})$
$p(s'_{ijt}|\lambda'_{ij})$
Posterior Forecast
$p(s'_{ijt}|\mathrm{Data}, \mathrm{Data}'_{t-1})$

Figure 1.6: Actual vs. Aggregated Posterior Mean Search Interests for Each Product

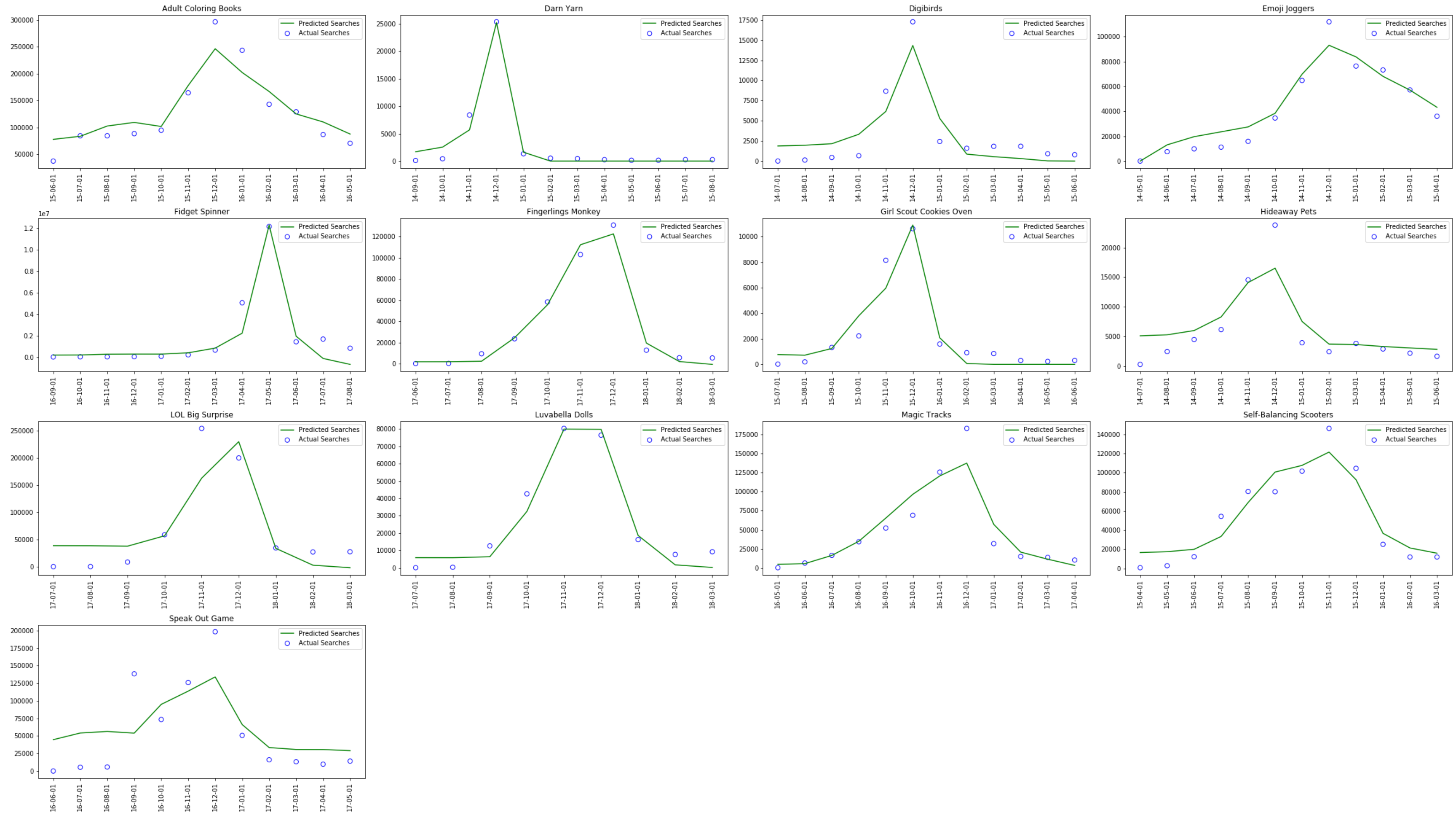
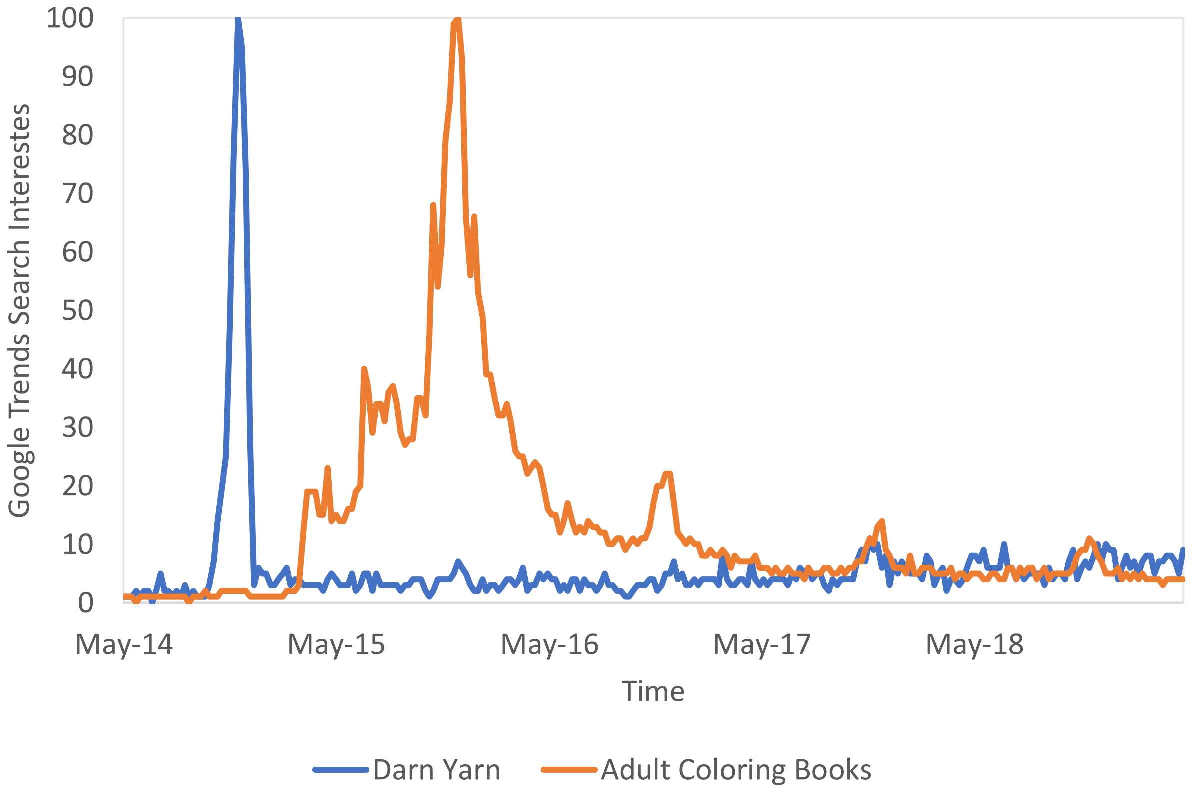


Figure 1.7: Darn Yarn vs. Adult Coloring Books



Notes: Height scales are not comparable. For each keyword (product), the Google Trends normalizes the maximum search interests to 100.

Table 1.1: Summary Statistics of Search Interests for Each RLC Product

Product Name	Takeoff Month	Peak Month	Months	Average	Minimum	Maximum	Std. Dev.
Calibration Sample							
Adult Coloring Books	Jun-2015	Dec-2015	12	16,298	655	233,298	21,960
Darn Yarn	Sep-2014	Dec-2014	12	414	0	23,699	1,519
Digibirds	Jul-2014	Dec-2014	12	378	0	10,773	924
Emoji Joggers	May-2014	Dec-2014	12	4,960	0	78,868	8,653
Fidget Spinners	Sep-2016	May-2017	12	225,854	15	5,890,098	585,557
Fingerlings Monkey	Jun-2017	Dec-2017	10	4,240	0	130,046	9,232
Girl Scout Cookies Oven	Jul-2015	Dec-2015	12	309	0	12,798	860
Hideaway Pets	Jul-2014	Nov-2014	12	753	0	21,659	1,529
LOL Big Surprise	Jul-2017	Nov-2017	9	8,905	0	322,562	22,133
Luvabella Dolls	Jul-2017	Nov-2017	9	3,495	0	110,374	8,269
Magic Tracks	May-2016	Dec-2016	12	5,940	0	166,235	12,021
Self-Balancing Scooters	Apr-2015	Nov-2015	12	7,716	0	167,417	16,999
Speak Out Game	Jun-2016	Dec-2016	12	6,259	0	119,830	10,679
Holdout Sample							
Doc McStuffins Mobile Cart							
J-Animals							

Notes: Google's website updates interrupted my GSI data collection in 2018. The disruption led to fewer than twelve months of data for Fingerlings Monkey, LOL Big Surprise, and Luvabella Dolls.

Table 1.2: Comparison of Model Performance

	Model 1 (Bass Model)	Model 2 (Proposed Model without National Channel)	Model 3 (Proposed Model)
Root Mean Square Error	7,655	5,332	5,287
Mean Absolute Error	1,540	1,077	1,006

Table 1.3: Estimates of Seasonality, Innovation Effect, and Market Size

Product Name	Seasonality	Innovation Effect	Market Size
	$\log(r_1)$	$\log(p_1)$	m_1
Adult Coloring Books	-10.24	-11.86	1.158
Darn Yarn	-8.10	-8.73	0.005
Digibirds	-7.94	-8.23	0.006
Emoji Joggers	-8.98	-10.18	0.115
Fidget Spinners	-12.58	-5.23	3.025
Fingerlings Monkey	-12.16	-6.49	0.051
Girl Scout Cookies Oven	-9.12	-8.72	0.004
Hideaway Pets	-8.00	-10.98	0.017
LOL Big Surprise	-8.62	-8.19	0.090
Luvabella Dolls	-9.86	-5.94	0.033
Magic Tracks	-10.70	-8.76	0.087
Self-Balancing Scooters	-9.01	-9.13	0.099
Speak Out Game	-9.05	-8.89	0.201

Notes: The 95% confidence interval of each posterior estimate in the table does not cover zero.

Table 1.4: Estimates of Imitation Effects

	Local Channel			National Channel			National vs. Local
Product Name	$\log(q^{L}_{1i})$	$\log(q^{L}_{2i})$	$\frac{\exp(\log(q^{L}_{1i}))}{\exp(\log(q^{L}_{2i}))}$	$\log(q^{N}_{1i})$	$\log(q^{N}_{2i})$	$\frac{\exp(\log(q^{N}_{1i}))}{\exp(\log(q^{N}_{2i}))}$	$\frac{\exp(\log(q^{N}_{1i}))}{\exp(\log(q^{L}_{1i}))}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Adult Coloring Books	-7.57	-15.56	2951.30	-11.21	-15.13	50.40	0.03
Darn Yarn	-5.23	-8.36	22.87	-0.10	-7.58	1772.24	169.02
Digibirds	-1.51	-7.61	445.86	-3.50	-7.04	34.47	0.14
Emoji Joggers	-5.36	-6.74	3.97	-6.69	-9.20	12.30	0.26
Fidget Spinners	-9.28	-13.01	41.68	-6.40	-12.55	468.72	17.81
Fingerlings Monkey	-5.76	-7.22	4.31	-2.70	-4.61	6.75	21.33
Girl Scout Cookies Oven	-4.59	-6.96	10.70	-0.14	-6.14	403.43	85.63
Hideaway Pets	-3.95	-9.53	265.07	-6.41	-9.39	19.69	0.09
LOL Big Surprise	-5.46	-9.49	56.26	-3.80	-9.09	198.34	5.26
Luvabella Dolls	-5.60	-8.87	26.31	-2.50	-8.44	379.93	22.20
Magic Tracks	-6.19	-5.26	0.39	-3.74	-6.30	12.94	11.59
Self-Balancing Scooters	-5.49	-8.21	15.18	-5.03	-5.39	1.43	1.58
Speak Out Game	-6.45	-11.06	100.48	-8.41	-11.34	18.73	0.14

Notes: The 95% confidence interval of each posterior estimate in columns 1, 2, 4, and 4 does not cover zero.

Table 1.5: Estimates of Demographic Variables

	Parameters for product i, city j ($\lambda_{ij}^{(k)}$)						
k =	1	2	3	4	5	6	7
	r_{ij}	p_{ij}	q_{1ij}^{L}	q_{2ij}^{L}	q_{1ij}^{N}	q_{2ij}^{N}	m_{ij}
Population density	0.04	**0.23**	-0.04	0.02	0.01	0.01	**-0.15**
(POPDEN$_j$)	(0.03)	**(0.07)**	(0.04)	(0.07)	(0.02)	(0.03)	**(0.03)**
Percentage of male population	-0.09	**0.08**	0.00	**0.14**	-0.02	0.06	-0.04
(MALE$_j$)	(0.07)	**(0.04)**	(0.03)	**(0.04)**	(0.02)	(0.04)	(0.06)
Median age	**-0.13**	**-0.12**	-0.01	0.04	-0.01	**-0.12**	0.06
(AGE$_j$)	**(0.07)**	**(0.04)**	(0.04)	(0.07)	(0.02)	**(0.06)**	(0.05)
Home ownership	**0.12**	0.06	**0.13**	-0.09	0.02	**-0.11**	**-0.23**
(HOME$_j$)	**(0.07)**	(0.03)	**(0.09)**	(0.06)	(0.02)	**(0.04)**	**(0.03)**
Median household income	**0.17**	**0.16**	**0.15**	**-0.32**	0.00	0.02	**0.06**
(INCOME$_j$)	**(0.05)**	**(0.05)**	**(0.06)**	**(0.04)**	(0.02)	(0.03)	**(0.03)**

Notes: Bold fonts indicate the 95% confidence interval of the posterior estimate does not cover zero.

Table 1.6: Forecast MAEs of Holdout Sample Products

A: Doc McStuffins Mobile Cart

	Forecast Month									% decrease in MAE by Model 3
Model	1	2	3	4	5	6	7	8	Average	
Forecast starts at the first month after takeoff										
1	617	4587	510	86	88	46	28	21	748	-96%
2	28	54	108	97	51	26	13	9	48	-32%
3	17	12	68	149	4	4	3	6	33	-
Forecast starts at the second month after takeoff										
1	355	1741	174	76	35	21	18	14	304	-83%
2	148	252	95	30	9	6	4	4	69	-25%
3	43	72	175	24	22	20	23	33	52	-
Forecast starts at the third month after takeoff										
1	243	641	141	36	21	15	12	9	140	-51%
2	183	1373	636	109	32	13	7	4	295	-77%
3	209	302	5	5	3	6	11	12	69	-

B: J-Animals

Model	Forecast Month								Average	% decrease in MAE by Model 3
	1	2	3	4	5	6	7	8		
Forecast starts at the first month after takeoff										
1	405	217	42	61	8	5	4	3	93	-67%
2	59	133	412	395	213	88	46	26	172	-82%
3	9	37	67	70	6	15	18	21	30	-
Forecast starts at the second month after takeoff										
1	142	141	75	11	9	7	7	10	50	32%
2	70	239	862	703	293	119	58	34	297	-78%
3	44	58	50	51	70	78	85	95	66	-
Forecast starts at the third month after takeoff										
1	223	169	25	21	15	13	13	12	61	-13%
2	155	426	259	64	21	9	4	3	118	-55%
3	207	149	11	9	10	12	14	15	53	-

CHAPTER 2. Solving the Cold Start Problem in Online Shopping Search Forecast: An Assessment of Two Approaches

Introduction

The memory of the Internet is fleeting. This rapid memory loss accelerates the life cycle of a product. More and more new products are popping up on the Internet and then quickly disappearing. The characteristics of the Internet pose a huge challenge to our marketing research. That is how to provide early forecast. When forecasting, an early start is valuable, but the data is not available. Therefore, the data is cold. The common time series models are unreliable when the temporal data is insufficient. This is also called a “cold start” forecasting problem.

While the importance of cold start forecasting for new products is evident to businesses, there is very limited research in this area. One of the traditional approaches is mainly to apply diffusion models. These models have some problems with forecasting. The first problem is that traditional diffusion models need to define the life cycle of a product and there are no clear criteria for determining the cut-off point for certain cycles. This leads to inaccurate model estimates.

The classical Bass model (Bass 1969), for example, requires estimating the parameters of the model after takeoff to get relatively accurate numbers. Due to the sparse data prior to takeoff, coupled with a lot of uncertainty, there was no good model to carve out this stage, causing trouble for subsequent product demand forecasts (Golder and Tellis 1997).

The second problem is that sales don't all simply rise gradually after takeoff and eventually become bell-shaped. After one product takes off, the diffusion model predicts a monotonous increase in sales before the growth spike. However, in some cases, a sudden decline in sales might occur after an initial increase. Moore (1991) observed this drop in sales, which he referred to as the split between the early market and the main market. Goldenberg et al. (2002)

referred to this phenomenon as the "saddle phenomenon" and defined it as a pattern in which the initial peak precedes a trough of considerable depth and duration, followed by an increase in sales. This also spells trouble for the forecast.

The third problem is the acceleration of the product cycle. There is substantial evidence that the overall temporal pattern of innovation diffusion accelerated over time (Van den Bulte, Christophe 2004). Moreover, Van den Bulte (2000; 2002) found evidence that such acceleration did exist. He investigated acceleration by studying 31 product categories of consumer electronics and household products using a Bass model with a zero internal impact parameter (p). His findings showed that the average annual acceleration between 1946 and 1980 was about 2%.

In addition to the diffusion models, another approach is to use time series models. However, these time series methods require the use of large amounts of historical data. Forecasting can only be achieved based on repetition over time, also known as a hot start forecasting. In the cold start cases, it is impossible to observe the market for a long enough time.

In general, the increasingly short life cycles require models to provide forecasts earlier. New products that have just come out in the morning hardly have enough information to complete the forecast. Therefore, cold start forecasting is a conundrum. Given these difficulties, the cold start issue has not received enough attention in marketing research.

One of the current solutions is through collaborative filtering. The main idea is to find similar products or customers. The problem with collaborative filtering is that it requires many customers and products. However, there is no standard for the number of customers and products, and the relationship between the number of products and model performance is unclear. Another problem is that collaborative filtering can only provide cross-sectional forecast.

For products with different growth patterns, it has a hard time to provide reliable forecasts over time.

In addition to the cold start issue, this essay also examines the impact of the sample size. From the big data perspective, the bigger the data the better the model performance. However, more products don't necessarily lead to better model performance. A larger sample can bring more noise than useful signals. In the absence of being able to effectively remove the noise in the sample, the argument that a larger sample is better is questionable. On the other hand, in a product recommendation or forecasting task, one may be required to make a response at short notice. For example, for a customer browsing a web page, the forecast should be provided in a short time frame. The introduction of more products may bring a dramatic increase in calculation time. In this situation, it is better to use a reasonable number of products to complete the forecast. Thus, this raises the general research question of studying the impact of sample size.

Moreover, the collaborative filtering primarily uses the correlation coefficient to find the similar products or customers. In addition to the correlation coefficient, this study also systematically incorporates other methods, such as the use of wrapping and embedded methods to find the relevant products.

After identifying the relevant products, two different models are used to make forecasts. In addition, I also analyze the effect of different product selection methods and product sample size. My study has important practical implications for a comprehensive solution to the problem of cold start forecasting.

Finally, I found that bigger data was not necessarily beneficial and harmless to cold start problems in practice. While Big Data encompasses more products, it does not mean that the more products used in the model, the more accurate the forecast results. In the case of insufficient data at the early stage, the performance of models can be improved by including more products. But when the number of products reaches a certain level, increasing the number of products does not continue to improve the performance of the model, but rather impairs the predictive accuracy of the model.

Framework of Solving Cold Start Problem

I provide a framework for solving cold start forecasting, as shown in Figure 2.1. The cold start problem is that the data for the new product is not enough on the temporal dimension. The key idea of addressing this problem is to find similar products with enough temporal data. Here, the similarity between two products is mainly measured by comparing the search growth rates in different locations. Specifically, the search growth rate of a new product will be projected onto multiple cities. Then I can obtain a geographical snapshot of search growth rate at time T. By comparing this geographical snapshot with the snapshots in my training product pool, we can find the most similar products in terms of the spatiotemporal search growth pattern. These products can further provide forecasts for the new product. Therefore, the framework includes two major stages: product selection (Stage 1) and search growth rate forecasting (Stage 2).

[Insert Figure 2.1 about here.]

Stage 1: Product Selection

The first stage is the process by which a subset of products is selected from the training product pool based on the spatiotemporal search pattern of a new product. A good subset of training products has a huge impact on the performance of the models in the second stage. Considering that there are hundreds of thousands of products available, product selection plays an important role in a reliable forecast. In general, the product selection methods can be divided into three categories (Chandrashekar and Sahin 2014):

Filter Methods

The filter methods are generally a group of methods for selecting the products based on the correlation coefficient between the new products and training products. The method to calculate the correlation coefficient here includes not only the Pearson's correlation, but also other methods, such as Linear Discriminant Analysis (LDA), Analysis of Variance (ANOVA), Chi-Square, and so on. These methods are designed for different data types, which are shown in Table 2.1. The search volume in my research is continuous for both new products and training products. Therefore, I use the metric of Pearson's Correlation in the following analysis.

[Insert Table 2.1 about here.]

Wrapper Methods

The wrapper methods consider the product selection as a search problem. These methods employ a random forest algorithm to select the best subset from the training products. Since the random forest algorithm will train a predictive model on all possible product subsets, the wrapper methods are often computationally expensive. The most common methods include forward feature selection, backward feature elimination, and recursive feature elimination.

Forward Selection: the forward selection method starts by evaluating all products individually and selecting one training product out of the library that can result in the best performance. In the following step, the forward selection method explores all potential combinations of the selected product from the previous step with the remaining products in the library. This approach keeps the pair of products that have the best performance and add more products one by one until the stopping criterion is met.

Backward Elimination: the backward feature elimination is opposite to the forward selection method and starts with all products in the data. In the second step, the backward feature method removes the least important product, which contributes less to the model's performance than the others. One product will be removed from the model each time until the stopping criterion is met.

Recursive Selection: the recursive elimination method tests the model on all possible products combinations. The target of this method is to select the best subset of products that can produce the best performance. Let's say I have N training products; the recursive elimination tests the model on all possible combinations of these N products until the algorithm finds the subset of products with the best performance.

Embedded Methods

The embedded method conducts the product selection during the model training process. Like the wrapped method, the embedded method also takes into account the interaction of products but with much less computation. During the selection process, the embedded algorithm completes the regression analysis at the same time.

One of the most notable features of embedded methods is the use of regularization, which is to add a penalty term to the regression model. The penalty term is multiplied to each coefficient in the model. This method has the advantage of effectively avoiding overfitting and improving the model robustness to the noise. The most common example of embedded method is lasso regression.

Lasso regression is also called L1 regularization. It involves shrinking the coefficients of some variables in the regression to 0. Thus, there variables are not included in the subsequent forecasting process. In my case, if a product has a coefficient of 0, then this product is taken out of the training sample.

Stage 2: Search Growth Rate Forecasting

The selected products in the first stage are then taken to the second stage. They are used to forecast the future search growth rate of the new products. The forecasting process in the second stage is further divided into two approaches.

Approach 1 (Dash Arrows in Figure 2.1)

The forecasting method in approach 1 is primarily by building a bridge between the selected training products and the new product. These selected training products are largely the group of products that are most like the new product in terms of the spatiotemporal search pattern. The underlying assumption here is that the products matched in this period are similar in growth rate pattern in the next period.

As shown by the dashed arrows in Figure 2.1, the forecasting stage includes both the training and test processes. The training process builds a linear model between the selected training products and new products (Left Dash Arrow in Stage 2). During the training process, I can estimate how the selected training products in period T-1 explain the search growth rate of the new product in period T. In the test process, the search growth rate of the selected training products in period T is further used to produce the forecasts for the new product in the Period T+1 (Right Dash Arrow in Stage 2).

Approach 2 (Solid Arrows in Figure 2.1)

Approach 2 is primarily a study of the search trends in the product itself. As mentioned above, the selected training products and the new products have a high degree of similarity in terms of search behavior. If the selected training products all share similar growth patterns, then it is likely that the new product follows the same growth pattern.

As shown by the solid arrows in Figure 2.1, the training process examines how the search growth rate of selected products changes from period T-1 to period T. Then the test process applies the same change to the new product being forecasted. The predicted search growth rate of the new product in period T+1 is a function of its lagged search growth rate.

Specifically, the second approach builds a hierarchical model and assumes each selected existing product has a growth rate and these growth rates would formulate a growth rate distribution. It is assumed that the growth rate of the new product is also a random draw from this distribution. By using this growth rate distribution, I can provide the forecast for the new product.

Data and Model

This study looked at Google's data from 2014 to 2017. A total of 6,100 of the most popular products, such as the Apple iPhone 7, PlayStation 4 and more, were included. The search volume was calculated on an annual basis. For each product, I had 4 years of annual search volume. Search volume for the product came from 763 U.S. cities with a population size greater than 50,000.

A computer algorithm was used to randomly select 100 products from the complete product pool as a holdout product sample. These holdout products were seen as the new products. The target was to forecast the search growth rate of these holdout products in 2017. It was a true holdout sample because all information in 2017 was considered not available. The remaining products from the complete product pool served as the training samples. For the sake of convenience, products in the training sample were called training products and products in the holdout sample become holdout products.

Given 4 years of annual search volume, I further calculated 3 growth rates for each product. The equation I used to calculate the growth rate is described below.

Growth rate of shopping-related searches

My equation for calculating the growth rate of shopping-related search volume is different from the traditional growth rate equation. This difference is manifested in two main ways.

The first difference is the denominator of my equation for calculating growth rates. In calculating the growth rate of search volume in the second year relative to the search volume in the first year, the denominator uses the search volume in the second year. The growth rate is calculated as follows.

$$\mathrm{G_{ijt}} = \frac{\mathrm{SV_{ijt}} - \mathrm{SV_{ijt-1}}}{\mathrm{SV_{ijt}}} \tag{2.1}$$

In equation 2.1, $\mathrm{G_{ijt}}$ is the growth rate of product i in city j from year t-1 to year t. $\mathrm{SV_{ijt}}$ is the search volume of product i in city j from year t-1 to year t.

The second difference is that my search growth rate is truncated. There are boundaries to growth rates. The smallest growth rate is -10 and the largest is 1. The right boundary of the growth rate is determined by the growth rate equation itself. When the search volume in the first year is 0 or non-existent and the search volume in the second year is greater than 0, the search volume growth rate is 1, which reaches to the right-hand boundary. When the first year's search volume is positive and the second year's search volume is 0, I set the value of the growth rate to -10, meaning that the second year's search volume is about 9% of the first year's search volume. At the same time, I truncate the value of the growth rate less than -10 to -10. The assumption is that when the search volume in the second year is less than 9% of the first year, it is basically equal to no search volume in the second year.

The main reason I calculate search volume in this way is that I need to set an upper bound if the growth rate is calculated using the traditional equation. Otherwise, when search volume is 0 in the first year and positive in the second year, the growth rate is infinity. I must subjectively set it to a large positive value, but this positive value should be large enough to cover most cases in my sample. Moreover, this number also highly correlates with baseline search volume. For example, a product has a search volume of 1 in the first year and 101 in the second year. According to the traditional search volume growth rate equation, the growth rate is 10000%. The growth rate cap should be larger than 100. To make the case even worse, many of the new products in the sample have annual growth rates greater than 100. If this upper limit is too large, it will further affect the calculation of the posterior correlation coefficients (see the product selection section for details). These large growth rates could be the influencers in the calculation of correlation coefficients. On the other hand, I can't set this value too small. For those products that have been adopted faster, a small cap on growth rates would abandon these important growth signals.

As a result, I decided to use growth rate equation in 2.1. An upper bound is automatically set to 1, which addresses the previous concerns. In my equation, a lower limit is needed. I set the lower bound at -10, which means that the second year's search volume is about 9% of the first year's search volume. For most of the products in my sample, this value is small enough that the distribution of growth rates is not severely left-skewed.

Based on my growth rate equation, the growth rates of 6,100 products in 736 U.S. cities are calculated for 2015, 2016, and 2017. Summary statistics on growth rates are presented in Table 2.2.

[Insert Table 2.2 about here.]

In Table 2.2, I note that shopping-related search volumes have declined year-on-year since 2014. On the one hand, total search volume has declined slightly from 2016 and 2017. On the other hand, the share of goods with a growth rate of -10 is increasing over time. In total, with the rapid growth of the Internet entering a bottleneck period, it is more difficult for total search volume to grow as quickly as it once did. On the other hand, the Internet has also accelerated the life cycle of new products. More and more products have shorter and shorter life cycles. As a result, the number of products "disappearing" from the Internet each year is increasing.

In the growth rate calculations, there are four cases that require our extra attention.

Case 1: No change. The product is not available, or the search volume is 0 in the first year and remains the same in the following year.

Case 2: Emergence. The product is not available, or the first year's search volume is 0 and the following year's search volume is positive.

Case 3: Disappearing. The product experiences positive search volume in the first year. However, the following year has 0 search volume, or less than 9% of the search volume in the previous year.

Case 4: Normal growth. Search volume is positive in both the first and second year.

The growth rate values are shown in Table 2.3 Panel A. In each year, the proportions of these 4 cases are different. These proportions are detailed from Panel B to Panel D in Table 2.3.

[Insert Table 2.3 about here.]

Consistent with the previous discussion, the proportion of Case 3s increased year on year, from 1.9% in 2015 to 2.5% in 2016. And it reached 3.3% in 2017. In addition, the proportion of cases 2 was also decreasing over time, from 7.5% to 1.5%. This indicated that the number of new products that had been searched completely from 0 was getting smaller and smaller.

Most importantly, I analyzed the quality of my data by looking at the proportions of missing values in growth rate calculations. All the missing values were included in Cases 1 through 3. Case 4 only contained the growth rates in the absence of missing values. The proportion of cases 4 increased from 84.4% in 2015 to 91.3% in 2017. It could be inferred that the number of searches missing from the data for a given year was around 10% or less.

Therefore, in my study, the cold start problem was that how to forecast future growth rates when one only knew the first-year growth rate of a product?

Product Selection Models

As discussed in the previous section, the first stage of my framework for addressing cold start problem is product selection. The methods of product selection are mainly divided into three categories: filter method, wrapped method, and embedded method. Due to space limitations, I have not exhausted all methods in each category. Instead, the most representative methods were chosen from each category. Next, I describe the method selected in each category and the equation used for each method.

Filter Method –Correlation

The first measure in filter method is called weighted Pearson correlation, which reflects the linear relationship of two products. The weight I use is the population size in each city. I first select a holdout product and calculate its correlation coefficient with all training products. I further rank these correlation coefficients. The larger the correlation coefficient, the more similar the search pattern of the two products. By ranking these coefficients, the purpose of product selection is achieved. The correlation coefficient is calculated as follows.

$$\mathrm{r_{XY}} = \frac{\sigma_{\mathrm{XY}}}{\sqrt{\sigma_{\mathrm{X}}\sigma_{\mathrm{Y}}}} \tag{2.2}$$

Where r_{XY} is the weighted Pearson correlation coefficient between a training product X and a holdout product Y. σ_X is the weighted variance for training product X and σ_Y is the weighted variance for training product Y. σ_{XY} is the weighted covariance between two products. The equations for σ_X, σ_y, and σ_{XY} are as follows.

$$\sigma_X = \frac{\sum_j w_j\left(x_{j,2015} - m_X\right)^2}{\sum_j w_j}, \sigma_Y = \frac{\sum_j w_j\left(y_{j,2016} - m_Y\right)^2}{\sum_j w_j} \tag{2.3}$$

$$\sigma_{XY} = \frac{\sum_j w_j\,(x_{j,2015} - m_X)(y_{j,2016} - m_Y)}{\sum_j w_j} \tag{2.4}$$

Where w_j is the population size in city j. $x_{j,2015}$ is the search growth rate of 2015 for training product X in city j. $y_{j,2016}$ is the search growth rate of 2016 for holdout product Y in city j. m_X and m_Y are the weighted average search volume for the two products. The equation 2.5 shows how I calculate m_X and m_y.

$$m_x = \frac{\sum_j w_j x_{j,2015}}{\sum_j w_j}, m_y = \frac{\sum_j w_j y_{j,2016}}{\sum_j w_j} \tag{2.5}$$

The weighted Pearson correlation coefficients range from -1 to 1. The positive 1 means that the two products have the same growth rate in all cities. The negative 1 indicates that the two products have exactly opposite growth rates in all cities. If there is no relationship between the growth rates of the two products, the correlation coefficient is 0. I know that a large positive correlation coefficient can help us in the following forecasting. Moreover, two products with a large negative correlation coefficient can also provide important information in the forecasting process. Therefore, I use the absolute value of the weighted Pearson correlation coefficient to rank the products and select the products with the largest absolute values from the training sample.

Filter Method – K-Nearest Neighbors

The second measure in filter method is called weighted K-Nearest Neighbors (KNN), which is very similar to the correlation coefficient method. In the weighted KNN method, I calculate the weighted Euclidean distance between two products. From there, the obtained Euclidean distances are further sorted to find those products with the smallest distances. The Euclidean distance is calculated as follows.

$$d_{XY} = \sqrt{\frac{\sum_j w_j \left(x_{j,2015} - y_{j,2016}\right)^2}{\sum_j w_j}} \tag{2.6}$$

Where d_{XY} is the weighted Euclidean distance between a training product X and a holdout product Y.

Wrapper Method – Backward Elimination

The wrapper method I use in this study is called backward elimination, which is to select a subset of training products by evaluating their importance in a linear regression. I start with a model that uses all products and then take out the products from the training sample to achieve the best model. These remaining products are my selected training products. The equation of backward elimination regression is as follows.

$$y_{j,2016} = \alpha + \sum_{i=1}^{N} \beta_i x_{ij,2015} + \varepsilon_j \tag{2.7}$$

Embedded Method – Lasso Regression

Lasso method is a also regression model based on the idea of reducing training variable set. By constructing a penalty function, it can shrink the coefficients of variables and make some regression coefficients become 0, thus achieving the purpose of variable selection. Lasso's model is shown in equation 2.8, in which I use L1 regularization, that is, the sum of the absolute values of the product coefficients.

$$\min_{\alpha,\beta_i}\left\{\sum_{j=1}^{J} w_j\left(y_{j,2016}-\alpha-\sum_{i=1}^{N}\beta_i x_{ij,2015}\right)^2\right\}\text{subject to }\sum_{i=1}^{N}|\beta_i|\le t \tag{2.8}$$

where t is a prespecified parameter that determines the amount of regularization.

Forecasting Models

Approach 1

First, I set up the training model between holdout products and selected training products.

$$y_{j,2016}=\alpha+\sum_{i=1}^{M}\beta_i x'_{ij,2015}+\varepsilon_j \tag{2.9}$$

Where I selected M products from the training products in stage 1 and $x'_{ij,2015}$ is the growth rate for the selected training product i in city j from 2014 to 2015.

Secondly, I use the $x'_{ij,2016}$ and estimated coefficients from the training model to obtain the predicted growth rates.

$$\widehat{y_{j,2017}}=\hat{\alpha}+\sum_{i=1}^{M}\hat{\beta}_i x'_{ij,2016} \tag{2.10}$$

where $\widehat{y_{j,2017}}$ is the predicted growth rate from 2016 to 2017 for the test product Y in city j.

Approach 2

First, the training process is to model the trend of the growth rate for the training products. The dependent variable is the growth rate of training products in 2016. The independent variable is their growth rates in 2015.

$$\begin{gathered}x'_{ij,2016}=\alpha_i+\beta_i x'_{ij,2015}+\gamma_i(x'_{ij,2015})^2+\varepsilon_{ij}\\ \alpha_i\sim(\alpha_0,\sigma_\alpha^2)\\ \beta_i\sim(\beta_0,\sigma_\beta^2)\\ \gamma_i\sim(\gamma_0,\sigma_\gamma^2)\end{gathered} \tag{2.11}$$

where I assume the parameters of α_i, β_i, and γ_i follow three normal distributions $N(\alpha_0, \sigma_\alpha^2)$, $N(\beta_0, \sigma_\beta^2)$, and $N(\gamma_0, \sigma_\gamma^2)$ respectively.

After I estimate the model 2.11, the estimated mean of the normal distribution is used to forecast the search growth rate for the holdout products.

$$\widehat{y_{j,2017}} = \widehat{\alpha_0} + \widehat{\beta_0} y_{j,2016} + \widehat{\gamma_0}(y_{j,2016})^2 \qquad (2.12)$$

Benchmark Model

Since I only have one-year growth rate for the holdout products, in the benchmark model, the predicted growth rate is simply the same growth rate that I have. The benchmark model is a straight projection. The predicted growth rate has nothing to do with the training products. The equation of the benchmark model is presented in equation 2.13.

$$\widehat{y_{j,2017}} = y_{j,2016} \qquad (2.13)$$

Results

In this study, 100 products were randomly selected as holdout products. The other 6,000 products served as training products. The number of training samples gradually increased from 50 products to 6000 products. The training sample was also randomly selected from the existing product pool. The subsequent product selection and forecasting procedures were only based on these 50 products.

The reason I did this was to examine whether the size of the training sample would affect the performance of my model. Evaluation of model performance was based on calculating the value of the mean absolute error (MAE) for each product in the holdout sample. Since my holdout sample contains 100 products, I had 100 MAEs for each model. I further used the simple average of these 100 MAEs (MMAE) as the criterion for evaluting the overall performance of my models. The smaller the MMAE value was, the better the overall performance of the model. For the benchmark model, the MMAE value was 1.019.

In my framework of addressing the cold start problem, a total of four product selection methods and two forecasting approaches were used. It was a 4 by 2 design. Thus, I evaluated 8 sets of models. In the next section, I reported the performance of each set of models in run.

Model 1: Filter Method (Correlation) + Forecasting Approach 1

The performance of Model 1 was presented in Table 2.4. The performance was related to the training sample size and the number of products selected in stage 1. Under the same training sample size, the performance of the model got worse as the number of products selected increased. On the other hand, I fixed the number of selected products and looked at the model performance under different sample size. When the number of products selected exceeded 4, the larger the training sample size, the worse the forecast. Moreover, all the sub-models were not better than the benchmark model. The model only performed better than the benchmark model when the training sample contained less than 500 products and the number of selected products was no more than 3. As more and more training products were included in the model, it was clear that the model's performance went off track. This was mainly because in regression models, too many variables were used, making the model prone to overfitting and introducing problems of collinearity. Model 1 achieved the best performance when the training sample size contained 50 products and only selected the most correlated product into the forecasting model. The corresponding MMAE was 0.8812, which improved the benchmark model by about 13.52%

[Insert Table 2.4 about here.]

Model 2: Filter Method (KNN) + Forecasting Approach 1

With the same training sample size, selecting more products into the model deteriorated the forecast sharply. This finding was the same as Model 1. When I fixed the number of selected products, the performance of forecast presented a U-shape over sample sizes. Initially, increasing the training sample size was helpful in improving the overall forecasts. However, as more and more products were available in the training sample, the noise from these products outweighed their value to the model. After more than 4 products were selected in stage 1, the model's forecast became unreliable, about which I would not go into detail. In Figure 2.2, I plotted the MMAE curves where the number of selected products was less than 4. These MMAE curves all showed the U-shape patterns. The MMAE of Model 2 was reported in Table 2.5. The optimal MMAE was 0.855, which improved by 16.15% over the benchmark model.

[Insert Figure 2.2 and Table 2.5 about here.]

Model 3: Wrapped Method (Backward Elimination) + Forecasting Approach 1

The wrapped method selected the best subgroup of products without the need to set a fixed number of selected products as in the filter method. Thus, as shown in Table 2.6, the evaluation of model 2 was only related to the size of the training sample size. When the training sample size was big enough, the wrapped method failed to pick a better subset of products. As a result, the MMAE values were more or less the same. To make the case even worse, the wrapped method did not solve the problems of overfitting, resulting in the forecast that were all inferior to the benchmark model.

[Insert Table 2.6 about here.]

Model 4: Embedded Method (Lasso Regression) + Forecasting Approach 1

In the last model, I employed the Lasso regression to complete the product selection and forecasting procedures. The model 4 improved on the previous models, mainly by adding a penalty factor to the coefficients in the regression.

The Lasso regression method automatically selected the optimal product subset. Only the sample size had an impact on the predicted outcomes. Table 2.6 showed the relationship between

MMAE and sample size. The overall MMAE curve presented a decreasing trend. Initially, the MMAE decreased rapidly as the training sample size increased. After training sample size reached to 700, the MMAE gradually plateaued. Finally, the MMAE had a minimum value of 0.729 at training sample size of 6000, improving the benchmark model by about 29.15%. The improvement was much higher than the previous models. Moreover, the predictive accuracy of the Lasso model fluctuated very little. The worst one also improved by 22.42% over the benchmark model.

Model 5: Filter Method (Correlation) + Forecasting Approach 2

In Model 5, I used the second forecasting approach. The impact of the model performance was still from both the training sample size and the number of products selected. Contrary to Model 1, with the same training sample, the more products selected, the better the model performed. I focused on the relationship between predictive accuracy and training sample size in the case of large training samples. As shown in Figure 2.3, I plotted the model's MMAE curve with training sample size when 50, and 80 products were selected in stage 1. In general, the MMAE curve had a U-shape initially and then gradually increased with the training sample size. While it might contribute to the forecasts when the sample size was small, once the training sample size reached a certain level, the harm from the large sample outweighed its contribution to the model. According to the MMAE output in Table 2.7, when the training sample size was 300 and 80 products were selected, the optimal MMAE was 0.831. The MMAE improved by 18.42% compared to the benchmark model.

[Insert Figure 2.3 and Table 2.7 about here.]

Model 6: Filter Method (KNN) + Forecasting Approach 2

The characteristics of the forecasting results exhibited in model 6 were highly like those in model 5. On the one hand, the MMAE curve gradually decreased as more products were selected, reaching the lowest MMAE after the number of selected products was 50 and then increasing slightly. On the other hand, as shown in the Figure 2.4, the MMAE curve also showed a U-shape with the sample size. For models using forecasting approach 2, model 6 performed better than the other models. With a sample size of 800 and 30 products were selected for forecasting, the minimum MMAE obtained was 0.720, which was 29.37% better than the benchmark model.

[Insert Figure 2.4 about here.]

Model 7: Wrapped Method (Backward Elimination) + Forecasting Approach 2

The MMAE curve for Model 7 also largely reflected the fact that the size of the training sample was not as large as it should be. As shown in the Figure 2.5, when the sample size was small, the impact on the forecast was more pronounced by adding more products. But once the sample size was large enough, say 1000 products, the additional products did not make the model perform better.

[Insert Figure 2.5 about here.]

Model 8: Embedded Method (Lasso Regression) + Forecasting Approach 2

Finally, Model 8 had the worst prediction of all models that used in forecasting approach 2. The overall model performance, while fluctuating little, was much worse than the benchmark model.

Concluding Remarks

This essay not only presents a framework for solving the cold start problem, but also examines the performance of the framework at different sample sizes, because the idea to solve the cold start problem is to find products that are sharing similar in spatiotemporal patterns from the training products. Also, the ability to find the most suitable products depends on the number of products in the training sample. If the number of products in the training sample is too small, it will be difficult to find the right products. Therefore, for most models, as the number of products in the training sample increases, the model's forecasting accuracy becomes better.

In the context of big data, increasing the number of products in the training sample seems endless. The intuitive feeling is that the larger the training sample, the more the model can benefit from it. But at the same time, the larger training sample poses a lot of problems. For example, it increases the time spent on searching for similar products. There is no fixed ratio between marginal search costs and marginal performance improvements. In some practical scenarios, it is not cost-effective to increase the training sample when the search time cost is too high. And to make matters worse, large training samples can have negative impact on model performance, such as introducing more noise and overfitting problems.

Therefore, I compare the model in two dimensions: horizontal and vertical. A horizontal comparison is to examine the performance of the same model at different training sample sizes. The vertical comparison is the comparison looks at the performance of different models under the same sample size.

First of all, after much empirical analysis, the LASSO method obtained the best performance among the models using forecasting methods 1. This is mainly because all the selected products in forecasting approach 1 are used as independent variables in the predictive model. The more products selected, the greater the likelihood that the model will be overfitting. And there are other problems that can arise such as collinearity. Thus, the model in Method 1 has a high sensitivity to the number of selected products. The smaller the number of products selected, the better the forecast.

Forecasting approach 2 employs a random effects model. Since there is only one independent variable, it is more stable than forecasting approach 1. The selected products are used to estimate a growth rate distribution, so the selected products also need to reach a certain number. For example, when 30 or 50 products are selected, most models reach their best forecasts. Among them, KNN's method was used to select products, and Model 6 came out on top among all models. Comparing the Lasso model in Prediction Method 1, both models performed well, and the accuracy of the forecasts did not differ significantly. Both models improved the running score model by as much as 30%.

Secondly, let's look at the results of horizontal comparison. I found that larger training samples were not always better: a gradual increase in training sample size first improved, and then in turn decreased the model performance. I attribute this finding to the fact that the incremental predictive value of additional training data decreases as the sample size increases, while the proportion of noise remains. As a result, the methods commonly used for variable/feature selection fail to remove the additional noise, resulting in overfitting, which reduces predictive performance. This finding cautions us that even in the age of "big data," all else being equal, training data is not necessarily as big as it should be when predicting demand growth for new products.

Figure 2.1: The Framework of Solving Cold Start Problem

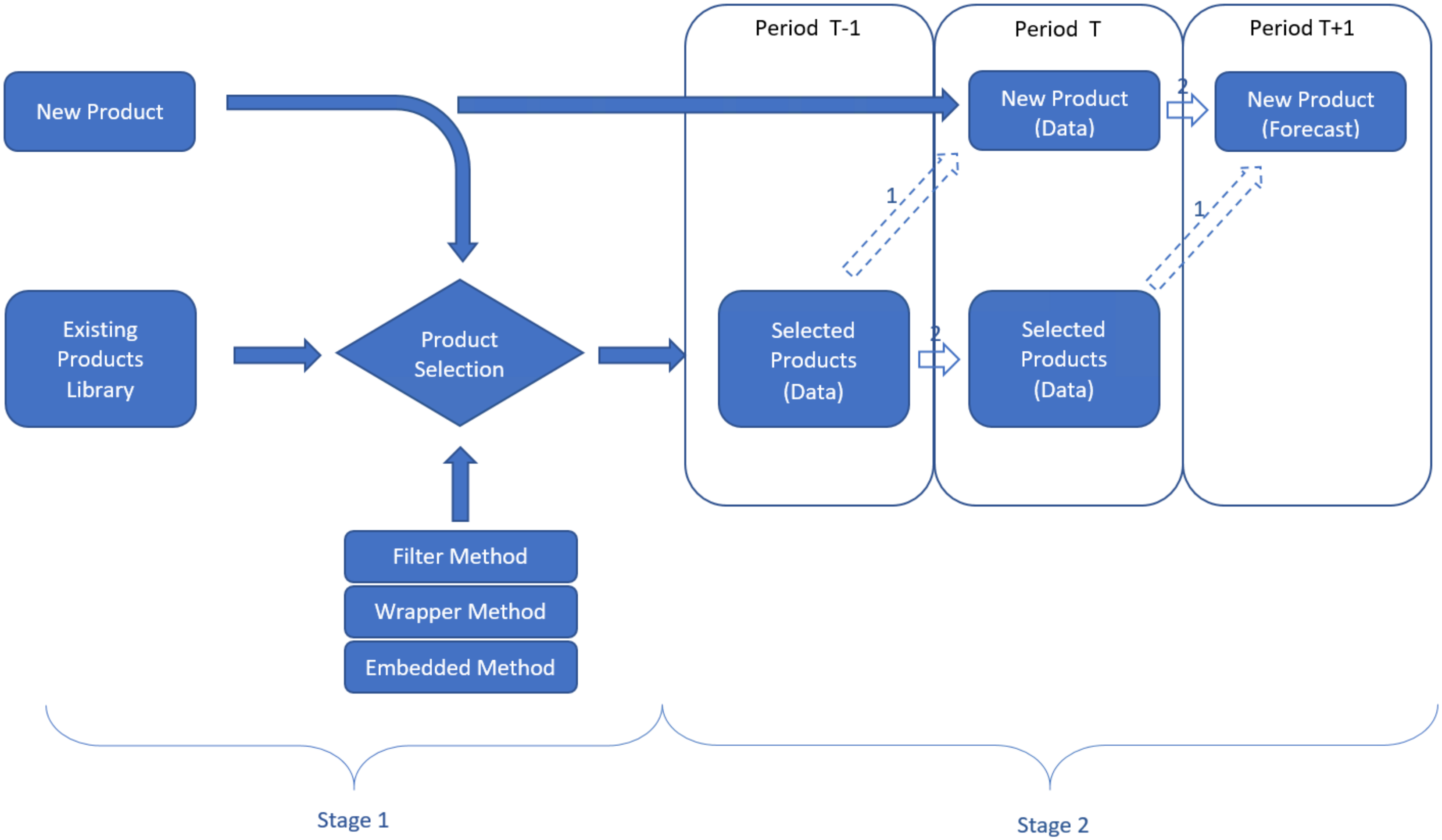
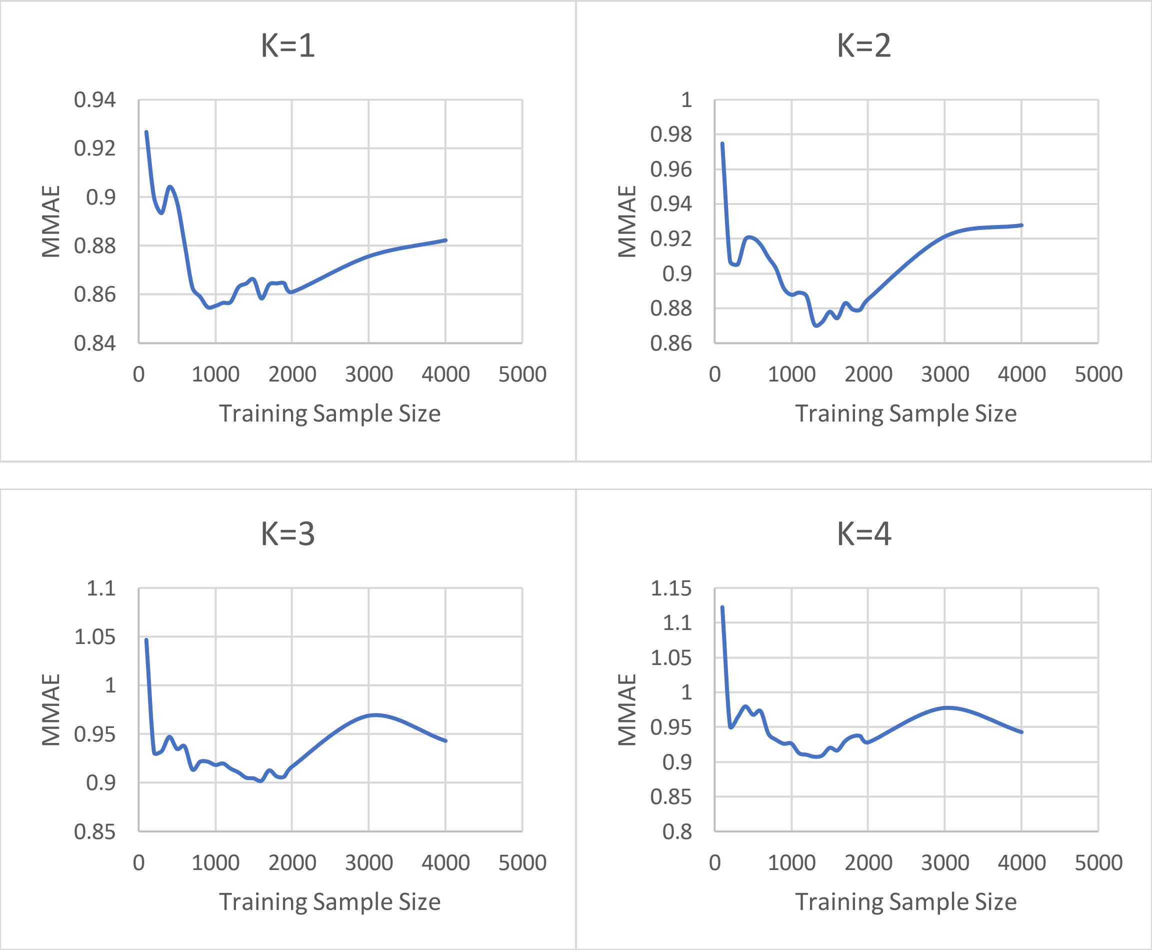
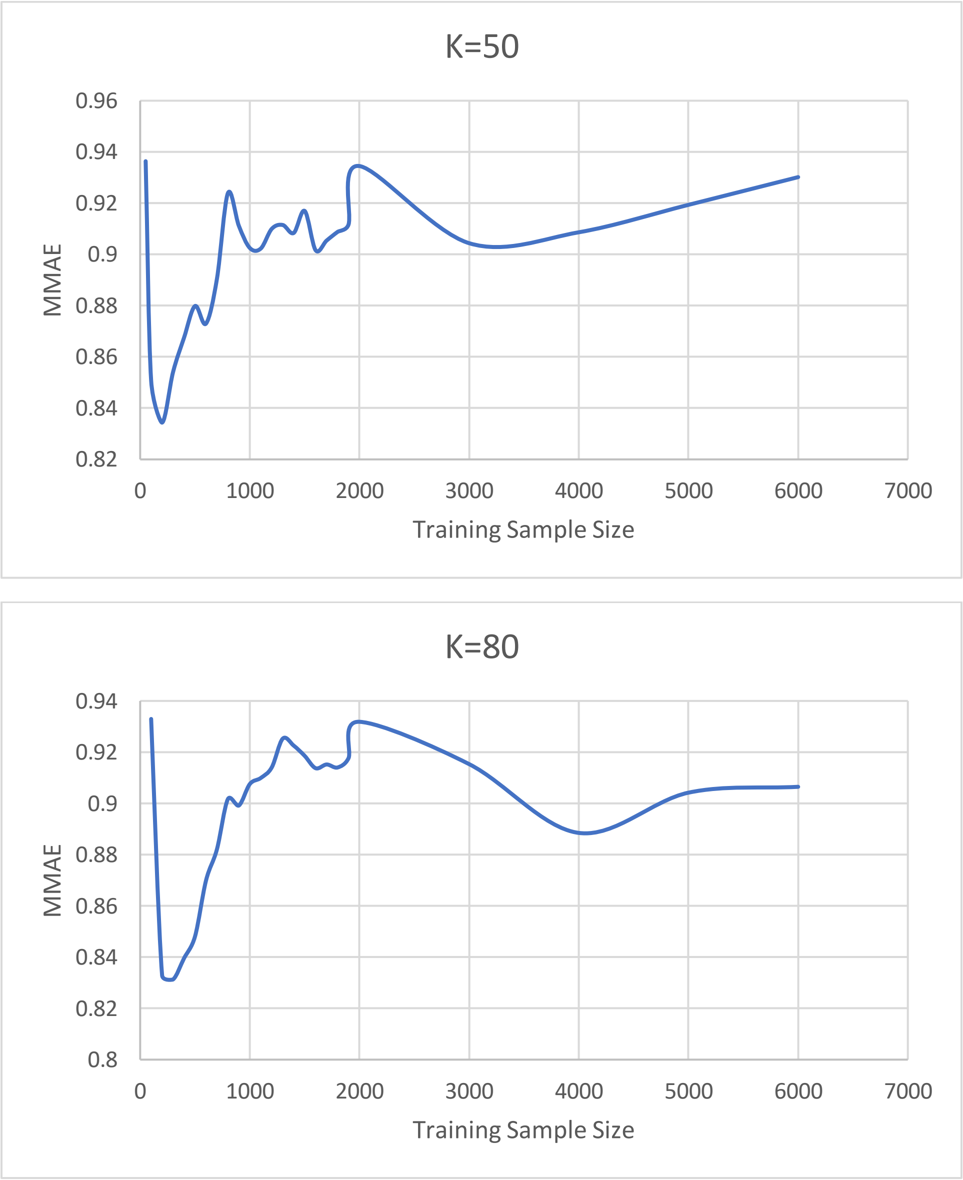


Figure 2.2: The MMAE Curve of Model 2



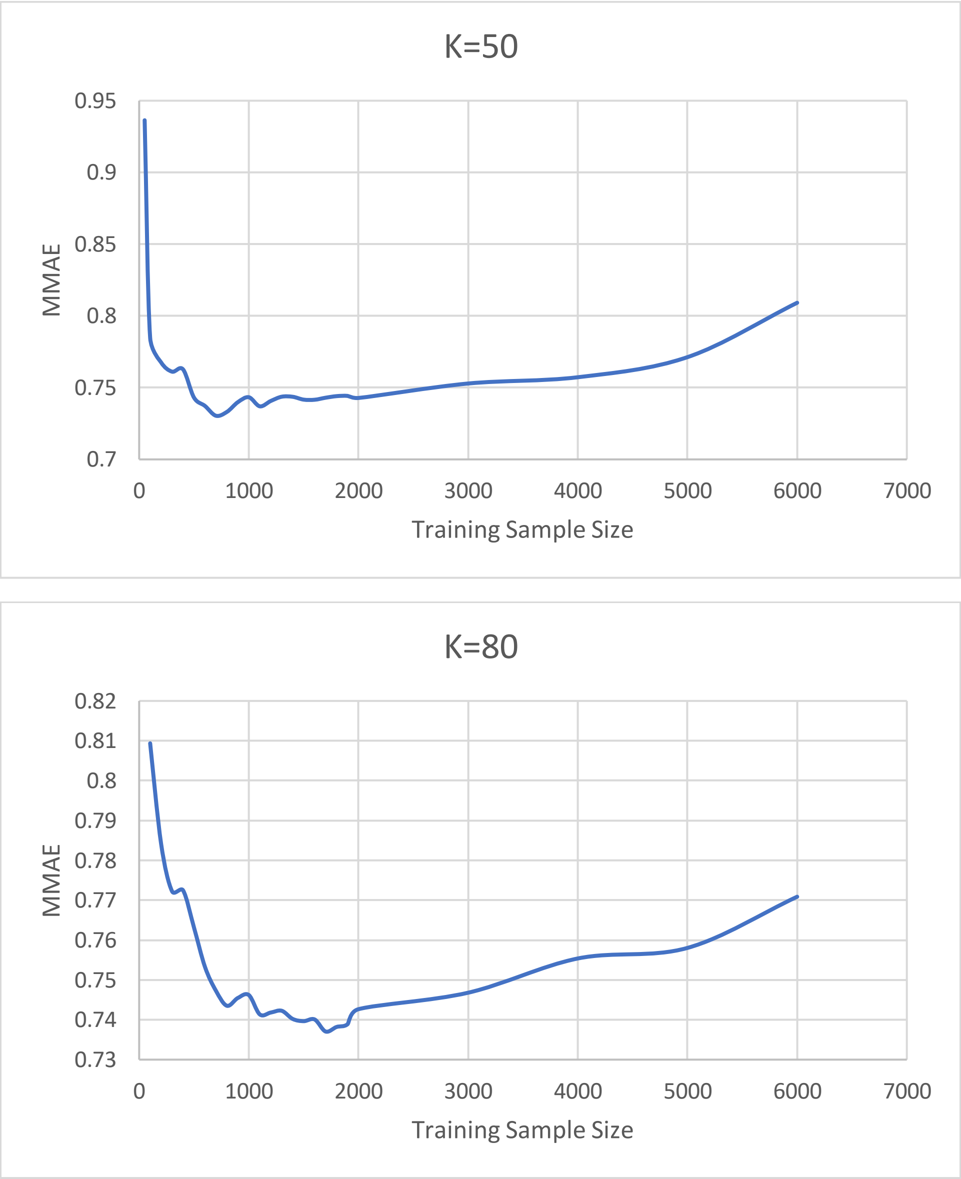
Notes: K is the number of selected products in the stage 1.

Figure 2.3: The MMAE Curve of Model 5



Notes: K is the number of selected products in the stage 1.

Figure 2.4: The MMAE Curve of Model 6



Notes: K is the number of selected products in the stage 1.

Figure 2.5 The MMAE Curve of Model 7

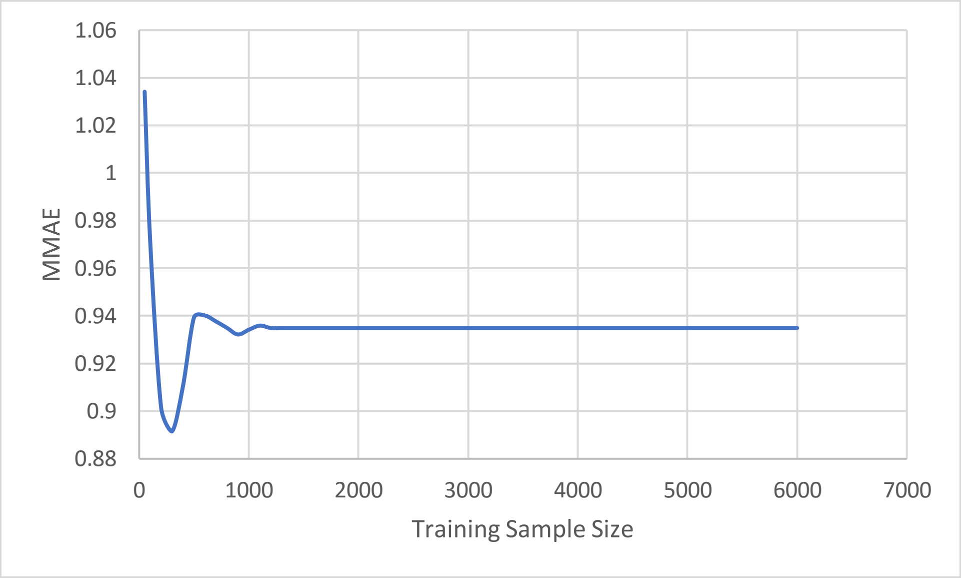


Table 2.1. Variable Type and Corresponding Metrics in Filter Method

Training Product \ Holdout Product	Continuous	Categorical
Continuous	Pearson’s Correlation	LAD
Categorical	ANOVA	Chi-Square

Table 2.2. The Summary Statistics of Search Growth Rate in Each Year

	Growth rate in 2015	Growth rate in 2016	Growth rate in 2017
Count	4489600	4489600	4489600
Mean	-0.07	-0.31	-0.61
Standard Deviation	1.67	1.90	2.07
Minimum	-10.00	-10.00	-10.00
25th Percentile	0.00	-0.15	-0.42
50th Percentile	0.25	0.18	0.00
75th Percentile	0.53	0.41	0.21
Maximum	1.00	1.00	1.00

Table 2.3. The Calculation of Search Growth Rate

Panel A: The Equations of Search Growth Rate in Four Different Cases

2015-2017		$\mathrm{SV_{ijt-1}}$		
		NA	0.00	>0
SV_{ijt}	NA	0	0	1
	0	0	0	1
	>0	-10	-10	$\mathrm{GR_{ijt}} = \frac{\mathrm{SV_{ijt}} - \mathrm{SV_{ijt-1}}}{\mathrm{SV_{ijt}}}$

Panel B: The Proportion of Each Case in 2015

2015		$\mathrm{SV_{ijt-1}}$	
		NA 0.00	>0
SV_{ijt}	NA 0	6.14% (Case 1)	7.52% (Case 2)
	>0	1.90% (Case 3)	84.43% (Case 4)

Panel C: The Proportion of Each Case in 2016

2016		$\mathrm{SV_{ijt-1}}$	
		NA 0.00	>0
SV_{ijt}	NA 0	3.56% (Case 1)	3.96% (Case 2)
	>0	2.51% (Case 3)	89.97% (Case 4)

Panel D: The Proportion of Each Case in 2017

2017		$\mathrm{SV_{ijt-1}}$	
		NA 0.00	>0
SV_{ijt}	NA 0	3.86% (Case 1)	3.32% (Case 2)
	>0	3.32% (Case 3)	91.26% (Case 4)

Table 2.4: The MMAE of Model 1

Training Sample Size	Number of Selected Products						
	1	2	3	4	5	10	20
50	0.881	0.926	1.066	1.120	1.122	1.244	1.478
100	0.934	0.936	0.985	1.165	1.135	1.211	1.396
200	0.910	0.924	0.940	1.037	1.055	1.022	1.182
500	1.203	1.258	1.785	2.321	2.397	3.161	8.357
1000	1.157	1.210	1.243	1.343	5.991	>10	>10
1500	1.009	1.265	1.359	1.359	2.377	>10	>10
2000	1.083	1.454	1.333	1.313	1.577	8.986	>10
3000	1.0341	1.2973	>10	>10	>10	>10	>10
4000	0.9069	0.9819	1.1796	1.4820	>10	>10	>10
5000	0.9103	0.9668	1.0599	1.1777	1.2385	>10	>10
6000	0.9463	1.1181	1.0212	1.2813	1.5567	>10	>10

Table 2.5: The MMAE of Model 2

Training Sample Size	Number of Selected Products						
	1	2	3	4	5	10	20
50	0.896	0.910	0.927	0.951	0.951	0.975	1.015
100	0.927	0.975	1.047	1.122	1.168	1.274	1.372
200	0.900	0.907	0.931	0.951	0.980	1.094	1.135
500	0.898	0.920	0.935	0.968	0.973	1.103	1.190
1000	0.855	0.888	0.918	0.926	0.947	1.042	1.219
1500	0.866	0.878	0.904	0.920	0.940	1.080	1.227
2000	0.861	0.885	0.917	0.928	0.957	1.044	1.232
3000	0.876	0.921	0.969	0.977	0.995	1.001	1.248
4000	0.882	0.928	0.943	0.943	0.975	1.021	1.171
5000	0.875	0.939	0.975	0.977	0.992	1.058	1.132
6000	0.864	0.928	0.972	0.975	0.994	1.054	1.096

Table 2.6: The MMAE of Model 3 and Model 4

Training Sample Size	Model 3	Model 4
50	4.666	0.791
100	3.771	0.769
200	2.848	0.739
500	>10	0.731
1000	8.365	0.735
1500	3.327	0.730
2000	2.326	0.740
3000	1.887	0.733
4000	1.447	0.726
5000	1.319	0.721
6000	1.242	0.722

Table 2.7: The MMAE of Model 5

Training Sample Size	Number of Selected Products									
	1	2	3	4	5	10	20	30	50	80
50	1.044	0.990	0.951	0.931	0.933	0.871	0.852	0.856	0.936	-
100	1.025	0.984	0.957	0.926	0.909	0.883	0.856	0.843	0.851	0.933
200	1.054	0.936	0.935	0.913	0.890	0.888	0.878	0.865	0.834	0.833
500	1.278	1.069	0.986	0.929	0.904	0.861	0.880	0.896	0.880	0.848
1000	1.243	1.147	1.097	1.055	1.003	0.926	0.897	0.900	0.903	0.908
1500	1.293	1.168	1.224	1.148	1.087	1.012	0.927	0.898	0.917	0.919
2000	1.346	1.242	1.114	1.228	1.237	1.077	0.971	0.947	0.934	0.932
3000	1.334	1.244	1.160	1.083	1.052	1.066	0.967	0.927	0.904	0.915
4000	1.168	1.207	1.171	1.110	1.037	1.054	0.966	0.930	0.909	0.888
5000	1.161	1.180	1.137	1.122	1.050	1.037	0.953	0.947	0.919	0.904
6000	1.140	1.094	1.089	1.054	1.048	0.949	0.941	0.923	0.930	0.906

Table 2.8: The MMAE of Model 6

Training Sample Size	Number of Selected Products									
	1	2	3	4	5	10	20	30	50	80
50	1.038	0.860	0.770	0.754	0.747	0.765	0.784	0.785	0.936	-
100	1.217	0.991	0.846	0.815	0.776	0.747	0.753	0.767	0.785	0.809
200	1.156	1.008	1.047	0.916	0.859	0.772	0.762	0.757	0.768	0.784
500	1.058	0.975	0.904	1.053	1.011	0.787	0.729	0.728	0.743	0.763
1000	1.028	0.950	0.918	1.032	1.008	0.793	0.733	0.739	0.743	0.746
1500	0.899	0.881	0.862	0.853	0.986	0.792	0.729	0.739	0.742	0.740
2000	0.882	0.883	0.883	0.833	0.825	0.852	0.737	0.732	0.743	0.743
3000	0.959	0.869	0.834	0.836	0.826	0.842	0.800	0.760	0.753	0.747
4000	1.015	0.879	0.825	0.805	0.814	0.811	0.835	0.778	0.757	0.755
5000	0.985	0.880	0.874	0.839	0.837	0.816	0.942	0.812	0.771	0.758
6000	0.943	0.861	0.855	0.816	0.825	0.805	0.816	0.864	0.809	0.771

Table 2.9: The MMAE of Model 7 and Model 8

Training Sample Size	Model 7	Model 8
50	1.034	2.592
100	0.972	2.218
200	0.902	1.509
500	0.939	3.263
1000	0.934	3.576
1500	0.935	3.970
2000	0.935	3.101
3000	0.935	3.176
4000	0.935	2.932
5000	0.935	2.994
6000	0.935	2.984

REFERENCES

Anderson, Carl R. and Carl P. Zeithaml (1984), “Stage of the product life cycle, business strategy, and business performance,” *Academy of Management Journal*, 27 (1), 5–24.

Bakshy, Eytan, Itamar Rosenn, Cameron Marlow, and Lada Adamic (2012), “The role of social networks in information diffusion,” in *Proceedings of the 21st international conference on World Wide Web*, ACM, 519–528.

Bass, Frank M. (1969), “A new product growth for model consumer durables,” *Management Science*, 15 (5), 215–227.

———, Trichy V. Krishnan, and Dipak C. Jain (1994), “Why the Bass model fits without decision variables,” *Marketing Science*, 13 (3), 203–223.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch (1992), “A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades,” *Journal of Political Economy*, 100 (5), 992–1026.

Bronnenberg, Bart J. and Carl F. Mela (2004), “Market roll-out and retailer adoption for new brands,” *Marketing Science*, 23 (4), 500–518.

Chandrashekar, Girish, Ferat Sahin. (2014), “A survey on feature selection methods,” *Computers & Electrical Engineering* 40(1), 16–28.

Choi, Tsan-ming (2013), *Fast Fashion Systems: Theories and Applications*, Communications in Cybernetics, Systems Science and Engineering, CRC Press.

Clayton, David and John Kaldor (1987), “Empirical Bayes estimates of age-standardized relative risks for use in disease mapping,” *Biometrics*, 671–681.

Du, Rex Yuxing, Ye Hu, and Sina Damangir (2015), “Leveraging Trends in Online Searches for Product Features in Market Response Modeling,” *Journal of Marketing*, 79 (1), 29–43.

——— and Wagner A. Kamakura (2012), “Quantitative Trendspotting,” *Journal of Marketing Research*, 49 (4), 514–36.

Easingwood, Christopher J., Vijay Mahajan, and Eitan Muller (1983), “A nonuniform influence innovation diffusion model of new product acceptance,” *Marketing Science*, 2 (3), 273–295.

Flood, Alison (2015), “Colouring books for adults top Amazon bestseller list,” *The Guardian*.

Garber, Tal, Jacob Goldenberg, Barak Libai, and Eitan Muller (2004), “From density to destiny: Using spatial dimension of sales data for early prediction of new product success,” *Marketing Science*, 23 (3), 419–428.

Garber, Tal, Jacob Goldenberg, Barak Libai, and Eitan Muller (2002), “Riding the saddle: How cross-market communications can create a major slump in sales,” *Journal of Marketing* 66(2), 1–16.

——— (2004), “From density to destiny: Using spatial dimension of sales data for early prediction of new product success,” *Marketing Science*, 23 (3), 419–428.

Golder, Peter N. and Gerard J. Tellis (1997), “Will it ever fly? Modeling the takeoff of really new consumer durables,” *Marketing Science*, 16 (3), 256–270.

Hu, Ye, Rex Yuxing Du, and Sina Damangir (2014), “Decomposing the Impact of Advertising: Augmenting Sales with Online Search Data,” *Journal of Marketing Research*, 51 (3), 300–319.

Johnson, Steven (2011), *Where Good Ideas Come From: The Natural History of Innovation*, New York: Riverhead Books.

Katona, Zsolt, Peter Pal Zubcsek, and Miklos Sarvary (2011), “Network effects and personal influences: The diffusion of an online social network,” *Journal of Marketing Research,* 48 (3), 425–43.

Kurawarwala, Abbas A. and Hirofumi Matsuo (1996), “Forecasting and inventory management of short life-cycle products,” *Operations Research*, 44 (1), 131–150.

Lilien, Gary L. and Arvind Rangaswamy (2004), *Marketing engineering: computer-assisted marketing analysis and planning*, DecisionPro.

Mahajan, Vijay and Robert A. Peterson (1979), “Integrating time and space in technological substitution models,” *Technological forecasting and social change*, 14 (3), 231–241.

Moore, G. A. (1991), “*Crossing the chasm*,” New York: HarperBusiness

Neelamegham, Ramya and Pradeep Chintagunta (1999), “A Bayesian model to forecast new product performance in domestic and international markets,” *Marketing Science*, 18 (2), 115–136.

Nicolaou, Anna (2017), “Fidget spinner craze turns the toy industry on its head,” *Financial Times*, (accessed May 21, 2019), [available at https://www.ft.com/content/5ead667c-3c0a-11e7-821a-6027b8a20f23].

Peers, Yuri, Dennis Fok, and Philip Hans Franses (2012), “Modeling seasonality in new product diffusion,” *Marketing Science*, 31 (2), 351–364.

Radas, Sonja and Steven M. Shugan (1998), “Seasonal marketing and timing new product introductions,” *Journal of Marketing Research*, 35 (3), 296–315.

Redmond, William H. (1994), “Diffusion at sub-national levels: a regional analysis of new product growth,” *Journal of Product Innovation Management: An International Publication of the Product Development & Management Association*, 11 (3), 201–212.

Rogers, Katie (2016), “The ‘Ice Bucket Challenge’ Helped Scientists Discover a New Gene Tied to A.L.S.,” *The New York Times*.

Semuels, Alana (2018), “The Strange Phenomenon of L.O.L. Surprise Dolls,” *The Atlantic*, (accessed May 25, 2019), [available at https://www.theatlantic.com/technology/archive/2018/11/lol-surprise-dolls-and-mystery-toys/576970/].

Shapiro, Carl and Hal R. Varian (1998), *Information rules: a strategic guide to the network economy*, Harvard Business Press.

Sharma, Praveen and S. C. Bhargava (1994), “A non-homogeneous non-uniform influence model of innovation diffusion,” *Technological Forecasting and Social Change*, 46 (3), 279–288.

Spiegelhalter, David J. (1998), “Bayesian graphical modelling: a case-study in monitoring health outcomes,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47 (1), 115–133.

Tellis, Gerard J., Stefan Stremersch, and Eden Yin (2003), “The international takeoff of new products: The role of economics, culture, and country innovativeness,” *Marketing Science*, 22 (2), 188–208.

The Economist (2017), “The lessons of fidget spinners,” *The Economist*, (accessed May 21, 2019), [available at https://www.economist.com/business/2017/09/09/the-lessons-of-fidget-spinners].

Valente, Thomas W. (1996), “Network models of the diffusion of innovations,” *Computational & Mathematical Organization Theory*, 2 (2), 163–164.

Van den Bulte , Christophe (2000), “New product diffusion acceleration: Measurement and analysis,” *Marketing Science* 19(4), 366–380.

——— (2002), “Want to know how diffusion speed varies across countries and product? Try using a Bass model” *PDMA Visions* **26** 12–15.

——— (2004), “Multigeneration innovation diffusion and intergeneration time: A cautionary note,” *Journal of the Academy of Marketing Science* 32(3), 357–360.

Venkatesan, Rajkumar, Trichy V. Krishnan, and Vineet Kumar (2004), “Evolutionary estimation of macro-level diffusion models using genetic algorithms: An alternative to nonlinear least squares,” *Marketing Science,* 23 (3), 451–464.