Business forecasting with online buzz

Oliver Schaer

Submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the Department of Management Science, Lancaster University

August 2019

Marketing Analytics & Forecasting
Lancaster University Management School

Declaration

This thesis is my own work and it has not been submitted in support of an application for another higher degree or qualification elsewhere.

Oliver Schaer

Abstract

In our fast-paced business world with changing consumer preferences, the demand for product life-cycles have shortened and become more volatile. This is the case for both own and competitors' products. To cope with those challenges, practitioners and researchers focus their attention on augmenting forecasting models by incorporating additional information such as online buzz. Online buzz reflects discussions around corporate and user-generated online content, such as social media interactions, forum discussions and online search. A majority of studies report forecast accuracy gains from its inclusion. However, on closer examination, many of these studies exhibit design weaknesses including lack of adequate benchmarking or rigorous evaluation, questioning their reported estimates. This thesis focuses on three research questions: (i) What is the predictive value of post-release buzz and usability for operational decision making? (ii) Can pre-release buzz help to estimate the market potential for sequential released new products? (iii) How suitable is pre-release buzz to predict competitors' new product success?

Our findings are mixed, by demonstrating both its potential, but also limitations of its usefulness. We conclude that online buzz is beneficial during the pre-release case, but not for the post-release phase. In the latter case, online buzz has little if any predictive signal, when considering realistic business lead times, explained by the frequently instant buying decisions and mixed pre- and post-purchase signals. On the other hand, pre-release buzz, which has a clear anticipatory characteristic, enhances new product forecasts. Our proposed framework demonstrates that pre-release buzz is beneficial also in predicting market potential. Moreover, it can be used to construct reliable forecasts for competitors' new products, providing market critical competitive intelligence. The thesis concludes with managerial implications and identifies future research directions stemming from this work and the critical reflection of the often hyped online buzz.

Preface

During my first year of study, Ivan Svetunkov and I co-founded a fictional Journal of Steps for which we measured all essential paths within Lancaster Management School and NATCOR venues to find the shortest path to the next ice cream store or supervisor office for ice cream break. This thesis was a long journey. Starting with the fact that the original research proposal was about developing a new model for forecasting the diffusion of online videos. As with most of our forecasts - what we planned is not how we end up. Instead, I welcome you on a research journey on the topic of business demand forecasting with online buzz. Many small steps were required to make this submission (finally!) happen.

The advantage of a long period spent on a PhD is that it perfected my procrastination skills, such as insulating my flat against the winter coldness or building up Spotify playlists with more than 5600 songs. Potentially due to my Greek supervisors influence, Spotifys AI system recommended me a different version of "Zorba the Greek" every Monday, and ultimately this leads me to believe that there is a desperate need for more research in this area. I also browsed YouTube for research purposes and channels such as Hydraulic Press, Wintergatan, Jelle's Marble Runs or Sous-Vide Everything provided many hours of learning. The remaining time I most probably spent at the local tennis club to whose members I am very grateful for a great time on the court.

If I am asked to pick a highlight of the PhD, then certainly the presentations at conferences and workshops. I would not be able to visit these exciting places without the incredible support from my two supervisors, Nikolaos Kourentzes and Robert Fildes. It is very hard to come up with words that can describe my gratitude for them. They are amazing with their knowledge, willingness to help and are both extremely humble given their academic impact. #whereIsMyPaper and British humour kept me going in times when progress was invisible for months. Assessed on the Net Promoter Score, I would certainly give them both a 10 out of 10 and a wholehearted recommendation.

I would like to thank the Department of Management Science for providing me with a scholarship and the Friend's programme and International Institute for Forecasters for their

conference travel grant. The members of the Centre for Marketing Analytics and Forecasting for giving me the opportunity in engaging more closely with their research and practitioner activities. I would like to express thankfulness to the administrative staff namely Gay Bentinck, Jackie Hughes, Jackie Clifton and Lindsay Newby. Also many thanks to Teresa Aldren for her great support in organising CMAF workshops and courses. I am also grateful to have received many helpful comments on my work and held discussions about academia along the journey from René Algesheimer, Christophe van den Bulte, Florian Dost, Sarah Gelper, Stephan Kolassa, Nikos Korfiatis, Nicos Pavlidis, Fotios Petropoulos, and Juan R. Trapero.

A very important part of the success is thanks to my parents, who contributed a considerable amount of their savings into my education. I do not know how closely they follow my research, but without their love and continuous support this thesis would not have been possible. I would also like to thank my brother for letting me use his tennis racket. As in the past when he gave me his camera, this gift turned into a huge obsession and I can only wonder what comes next.

I received great support from Switzerland where people (rather surprisingly) stayed interested in what I am doing and kept pestering me when the thesis might be finished. In particular, my godmother, Maja Schmid, who always thought of me and sent nice messages together with Swiss chocolate treats. I enjoyed many great Skype sessions with Cyril Mugglin discussing my PhD life and in turn listened to his adventures in industry. Daniel Schmid always provided great photos from the Swiss mountains, news from the football club and many childhood memories. I also received many e-mails with exciting stories about the daily problems in Basel from Lucia Messmer and creative inspiration from Samuele Tirendi.

This project would not have finished without the tremendous help from very good friends I met here in the UK. I start with Ivan Svetunkov, my PhD mentor and an immense help in smoothly mastering the journey of steps and building up the Russian Spotify playlist. Anna Sroginis will be especially remembered for the many great cupcakes and macaroons I got treated with. Another source for Russian delights came from Alisa Yusupova in the form of Marmalandia. There were also some unsuccessful attempts by Yves Sagaert to convince me of the high quality of Belgium chocolate. A special thank goes to Timo Kunz for providing me with the best possible assistance in starting my PhD life, and some of his furniture has just been passed onto the next PhD generation. Simon Spavound for being able to use his brain so many times for problems I was struggling with. He was also my Babel fish when it came to “very British problems”, both in language and the rough life here in the north of England. Amy Benstead who always brought the latest sustainable fashion (her research topic) into our office and gave me a cute little cactus which, sadly, always grew faster than my research. My friends from NATCOR Elvan Gökalp, Andrea Mor and Öykü Naz Attila, whom all invited me to visit their hometown, broadened my cultural knowledge and kicked-off the Turkish, Italian and Dream catcher Spotify playlists.

The last paragraph I would like to devote to my dear friend Christoph Häcki, who unfortunately is not able to read the final version of this thesis. Besides our 20-year friendship, he had a large influence on where I am now. His incredible support during almost my entire time at high-school and university enabled me to overcome many of my limitations. He also introduced me to the programming language R and data analytics during my bachelors degree, which provide a solid foundation for this thesis. You are sadly missed!

Lancaster, in August 2019
Oliver

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Technology is enabling consumers to draw on comprehensive information for their buying decisions. This increase in transparency intensifies competition between products (Li et al., 2011). Instruments such as online reviews allow past customers to critically reflect upon their purchases and directly affect any future ones (Li and Hitt, 2008; Dellarocas, 2003). On top of that, the online discourse is actively fuelled by companies with online promotional activities and content creation on social media, sparking online conversations (Kuksov et al., 2013; Wang et al., 2012). Social media networks such as Facebook, Twitter or Instagram are regarded as effective facilitators of electronic word-of-mouth and can substantially influence interest in products (Seiler et al., 2017; Babic Rosario et al., 2016). Newspapers around the globe report touristic locations for which Instagram helped to reach popularity that exceeded local infrastructure capacity (Rhyn, 2019; Coldwell, 2018) or similarly tremendously increase demand for fashion items (Conlon, 2018).

In the literature, signals that emerge from user-generated content, online search or social media interactions are often referred to as buzz (e.g. Deusser et al., 2018; Xiong and Bharadwaj, 2014). This buzz can be triggered by both companies and consumers (Hewett et al., 2016). It is important to distinguish between the pre-release and post-release buzz of a product, i.e the conversation and searches that take place before the product is launched and vice-versa. While the pre-release is solely anticipatory based, the post-release buzz is a mix between product interest (anticipation) and experience-sharing, for example through electronic word-of-mouth or product recommendations (Houston et al., 2018). Houston et al. identified three types of behaviour that build the pre-release buzz: (i) communication, for example, posting something on social media, (ii) online search that can be reflected in search traffic and (iii) participatory behaviour, such as to like or comment on posts on social media.

Despite being a communication instrument, both scholars and practitioners put their attention on how this rich data source can improve corporate decision-making (see for example Boone et al., 2019; Erevelles et al., 2016, for reviews). Buzz has the advantage that, compared to traditional market intelligence methods, such as surveys or focus groups, it is available in a timely manner and comes with comparably low collection costs. The preeminent question is to what extent buzz can improve predictive ability. This is particularly relevant in times of increased demand volatility due to changes in consumer preferences and competition. Improving forecasts can directly impact the operational costs such as stock-keeping, but also provides managerial insights to adjust, for example, marketing expenditure so as to influence future consumption. This thesis aims to contribute to this fast-growing body of literature on incorporating pre- and post-release buzz into business demand predictive models.

Application for the post-release forecasting with buzz include, for example, product sales (e.g., Cui et al., 2018; Lau et al., 2018; Geva et al., 2017), tourist arrivals (e.g., Önder et al., 2019; Choi and Varian, 2009) and stock market performance (e.g., Tirunillai and Tellis, 2012; Da et al., 2011). However, some studies report limited gains for predictive ability (e.g., Ferrara and Simoni, 2019; Ruohonen and Hyrynsalmi, 2017; Li, 2016; Limnios and You, 2016). A further complication in assessing the value of such inputs for operational decision making, comes from the typically weak forecast evaluation design of many studies with inadequate performance metrics and hold-out samples (Kalampokis et al., 2013). Moreover, a majority of studies only consider short forecast horizons, which often do not realistically relate to business needs. Therefore, it is unclear to which extent post-release buzz can be used as a leading indicator for demand. In addition, the previously outlined interplay between anticipation towards product purchase and experience sharing afterwards and its impact on the predictive accuracy is unexplored. Therefore, there is a further research question in exploring whether the benefits of such information are consistent during a product's life-cycle. This research questions will be discussed in Chapter 2. Earlier versions of this chapter have been presented during the *17th IIF Workshop on ICT and Innovation*

Forecasting, *EURO15* and at the *36th & 37th International Symposium on Forecasting*. It is now published in the *International Journal of Forecasting* (ABS 3; Schaer et al., 2019a).

For pre-release buzz, the majority of the literature has investigated entertainment products, such as box office (e.g., Kim et al., 2015; Asur and Huberman, 2010), music albums (Dhar and Chang, 2009) and video games sales (Xiong and Bharadwaj, 2014). New product forecasting is different to the post-release case in that there is typically no historical sales information available (Goodwin et al., 2014). Nonetheless, managers are interested in knowing how successful a new product might become so as to plan for advertising activities, procurement and manufacturing. In practice a common method for determing the market potential of new products is to use judgmental forecasts (Kahn and Chase, 2018). It is well understood that such forecasts are subject to biases, and in particular overoptimism (Belvedere and Goodwin, 2017; Markovitch et al., 2014; Tyebjee, 1987). Alternative approaches such as consumer surveys (e.g., Moon et al., 2016; Chintagunta and Lee, 2012; Eliashberg et al., 2000) or conjoint analysis (e.g., Orbach and Fruchter, 2011; Lee et al., 2006) are costly and only reflect a specific point of consumer preference which is likely to change throughout the pre-launch phase (Meeran et al., 2017). This makes pre-release buzz an ideal candidate to help estimating demand. The majority of the literature has investigated the forecasting potential only for the initial weeks of sales, which from an operational point of view is not sufficient. Therefore, Chapter 3 investigates whether pre-release buzz is helpful in modelling the market potential and investigates the leading properties of pre-release buzz for total end-of-life sales. Chapter 3 is now available as a working paper (Schaer et al., 2019b) with the aim to be submitted to *Journal of Marketing Research*. Earlier versions of this work have been presented at the *18th IIF Workshop on Forecasting New Products and Services*, the *39th ISMS Marketing Science Conference* and the *38th International Symposium on Forecasting*.

Publicly available data from internet platforms contain information relevant both to own products and competitors' (Teo and Choo, 2001). Insights in competitors' actions are of strategic importance to organisations (Calof and Wright, 2008), in particular for new products, where the trend for shorter life-cycles leads to increased competition on product launch (Calantone et al., 2010). However, the literature on pre-release buzz, except the study presented in Chapter 3, predicted new product sales from datasets that include products from multiple brands assuming availability of information that may not always be feasible in practice (e.g. Divakaran et al., 2017; Onishi and Manchanda, 2012) and studies on competition are not in a predictive context (e.g., Gopinath et al., 2013). While using the richer datasets provides better estimates, it does not reflect the real-world environment where sales information is often only available for own brands and stock keeping units. This questions the realism and feasibility of the reported value of past recommendations. To address this, it is essential to distinguish between internal and competitor products. While internal data typically provide full access to sales history and marketing activities, external competitor insights are constrained to only what is publicly available or possible to purchase at substantial cost and for the modelling purpose at potentially too aggregate format. Assuming some degree of homogeneity among competing products, a key question is whether competitors' pre-release buzz can be used together with internal sales data to infer competitors' success. In Chapter 4, we extend the application of pre-release buzz to estimate sales of competing products. The research idea and methodology has been presented at the *40th & 41st ISMS Marketing Science Conference*, the *60th Anniversary Conference of the Operational Society* and *39th International Symposium on Forecasting*. It is also accepted for presentation at *INFORMS 2019*. We aim to submit this paper to the *Production and Operations Management*.

The remainder of this thesis is organised as follows; Chapter 2 will investigate how online buzz can be used to provide short-term forecasts in the context of operational decision making. In Chapter 3 we present an application to estimate the market potential of sequential products with pre-release buzz. We subsequently extend the idea of pre-release forecasts to the case of competitors' product in Chapter 4. Finally, Chapter 5 discusses the contributions to the literature and concludes with managerial implications and suggests exciting future avenues of research.

Chapter 2

Demand forecasting with user-generated online information

Recently, there has been substantial research on augmenting aggregate forecasts with individual consumer data from internet platforms, such as search traffic or social network shares. Although the majority of studies report increased accuracy, many exhibit design weaknesses including lack of adequate benchmarks or rigorous evaluation. Furthermore, their usefulness over the product life-cycle has not been investigated, which may change, as initially, consumers may search for pre-purchase information, but later for after-sales support. In this study, we first review the relevant literature and then attempt to support the key findings using two forecasting case studies. Our findings are in stark contrast to the literature, and we find that established univariate forecasting benchmarks, such as exponential smoothing, consistently perform better than when online information is included. Our research underlines the need for thorough forecast evaluation and argues that online platform data may be of limited use for supporting operational decisions.

2.1 Introduction

Nowadays it is becoming increasingly easy for organisations to obtain individual consumer behaviour data from potential and actual customers by using internet platforms, such as Google or Twitter. Consumers seek online information on branded and non-branded content (Heinonen, 2011). Companies actively support purchase decisions by distributing branded content through internet channels, which generates further interactions (Kuksov et al., 2013; Wang et al., 2012). Research has argued that information such as search traffic popularity, or numbers of shares on social networks can lead to improved forecast accuracy (e.g. Cui et al., 2018; Geva et al., 2017; Goel et al., 2010). While online shares reflect an electronic word-of-mouth process (Seiler et al., 2017; Babic Rosario et al., 2016), the popularity of a search keyword can be regarded as a proxy for consumer interest in a product (Du and Kamakura, 2012; Stephen and Galak, 2012), but also reflect the success of advertising activities (Srinivasan et al., 2016; Hu et al., 2014).

There are numerous time series modelling papers that incorporate information from the internet; for instance, in econometric now-casting such inputs can be useful to overcome publication lags of governmental economic indicators or market surveys (e.g. Vosen and Schmidt, 2011; Choi and Varian, 2009). Other example include predicting stock volatility (e.g. Bollen et al., 2011); influenza outbreaks (e.g. Ginsberg et al., 2009); tourist arrivals (e.g. Hand and Judge, 2012); car sales (e.g. Fantazzini and Toktamysova, 2015; Du et al., 2015); and retail sales (e.g. Boone et al., 2018; See-To and Ngai, 2016).

One important aspect is that most business decisions, such as allocating resources, inventory decisions or planning marketing expenditures, are based on forecasts and in turn imply some forecast lead time, which is relevant for the decision planning horizon. This makes the usefulness of online information for demand forecasting more contentious. Past research has supported both its usefulness (e.g. Lau et al., 2018; Brynjolfsson et al., 2016; Schneider and Gupta, 2016) and its limitations (e.g. Ruohonen and Hyrynsalmi, 2017; Li, 2016; Limnios and You, 2016). A further complication in assessing the value of such inputs for operational decision making comes from the typically weak forecast evaluation setup that is used and the short forecast horizons, which often do not relate realistically to business needs. Kalampokis et al. (2013) in their review of forecasting with social media data, report that more than one-third of studies do not test the claimed predictive abilities, using hold-out-sample or adequate predictive measures. Their review does not consider research that includes information originating from other than social media networks, for example, search traffic information; and omits any dedicated discussion on the forecasting approaches used.

The aim of this paper is to (i) provide a holistic review of the existing literature on forecasting with internet-based consumer behavioural data for a range of applications; (ii) discuss the limitations and challenges of using such data for predictive purposes and (iii) explore whether the usefulness of such information remains consistent during a product's life-cycle. To exemplify this, consider a consumer who may research a product online prior to purchasing. The search is a leading indicator. Post-purchase the same consumer may search online for support information that does not lead to additional purchases. Therefore, it is reasonable to expect that the usefulness of online information changes over the life-cycle of a product. To support our critical review of the literature, first, we replicate one experiment by Choi and Varian (2012) and second, we model sales of video games and the consumption of viral video advertisements using social network shares.

Although the literature is overwhelmingly positive as to the benefits of search traffic and social media derived variables, we argue otherwise given the evaluation and experimental design of almost all studies. We question the realism of the forecasting setup (for instance the forecast horizon) for a number of papers and also find that several do not include adequate benchmarks. Furthermore, we find no support where the usefulness of the variable changes over the life-cycle of a product from our empirical experiment.

The paper is organised as follows; Section 2.2 provides a review of the literature that uses explanatory variables from internet platforms for forecasting. Section 2.3 highlights the challenges in handling online information. We then present in Section 2.4, two case studies to validate the findings of the literature. Section 2.5 discusses the usability of internet platform information and Section 2.6 presents the conclusions.

2.2 Forecasting with online user generated data

We present the literature in four subsections which are summarised in Table 2.1. The first horizontal grouping summarises Section 2.2.1, which surveys data sources. The columns reflect groups of forecast applications, which are detailed in Section 2.2.2. The second horizontal grouping classifies the forecast models used and is discussed further in Section 2.2.3. The last grouping lists forecasting principles, to which adherence is reviewed in Section 2.2.4. Overall, 95% of the surveyed studies conclude in favour of using user-generated information for forecasting. A detailed table for each area of application is provided in the appendix of this chapter.

We limit our literature review to studies that assess the forecasting performance of time series models on relative short horizons, relevant to operational business forecasting. This precludes areas such as: predicting election outcomes (e.g. Mavragani and Tsagarakis, 2016; Huberty, 2015), product rankings (e.g. Hou et al., 2017; Liu et al., 2016; Goel et al., 2010), pre-launch forecasts (e.g. Kim et al., 2015; Xiong and Bharadwaj, 2014; Dellarocas et al., 2007) and marketing effectiveness (e.g. Kumar et al., 2016; Hu et al., 2014; Du and Kamakura, 2012). Although most of these studies suggest benefits from online user-generated data, their modelling approach as well the forecast target and accuracy measures used, differ substantially and would require a separate discussion that is out of scope for this paper. We do, however, include some of their findings on the handling of such data to support our discussion.

2.2.1 Data sources

Two review papers have been published that cover forecasting with social media networks (Phillips et al., 2017; Kalampokis et al., 2013). This study considers a wider range of internet sources for obtaining user-generated information. These are search traffic, social network sites, blogs and microblogs, forum posts and online product reviews. We do not specifically include studies which obtain data from news streams, such as the GDELT project (e.g., Fast et al., 2018), because such information may not reflect online consumer behaviour. It is worthwhile noting that we were unable to find any research that explores the predictive

Table 2.1: Summary of the literature

	Areas of application (Section 2.2.2)					
	Economic indicators (n = 14)	Financial markets (n = 7)	Public health & environment (n = 10)	Services (n = 16)	Consumer goods (n = 14)	Overall (n = 61)
Data sources (Section 2.2.1)						
Forum and blogs	0%	14%	10%	0%	7%	5%
Reviews	0%	0%	0%	0%	21%	5%
Search traffic	100%	71%	70%	100%	79%	87%
Social networks	7%	29%	30%	6%	14%	15%
Forecast modelling (Section 2.2.3)						
Multistep-ahead	21%	14%	30%	50%	29%	31%
Non-linear models	14%	14%	20%	0%	36%	16%
Nowcast model	64%	0%	60%	25%	36%	39%
Ordinal based[a]	7%	43%	30%	6%	43%	23%
Volume based[b]	100%	71%	70%	100%	71%	85%
Forecast evaluation(Section 2.2.4)						
Adequate benchmarks	21%	0%	0%	31%	21%	18%
Hold-out-sample	93%	86%	100%	94%	93%	93%
Multiple time series	14%	71%	30%	31%	93%	46%
Rolling origin[c]	93%	100%	90%	81%	86%	89%
Statistical testing	43%	71%	0%	31%	36%	34%
Report improvements	86%	100%	100%	100%	93%	95%

[a] sentiment information or product ratings [b] search traffic popularity, shares or mentions
[c] also cross-validation

ability of many popular social media platforms such as Instagram, Snapchat, Pinterest and LinkedIn. The same is true for user-generated videos from platforms like YouTube, even though studies suggest that video blogs can lead to positive purchase intention (Lee and Watkins, 2016). The limited use of these data sources might partly be due to the difficulty to access and identify content.

Search traffic information is the most frequently used source present in 87% of the investigated studies, even accounting for 100% of applications in economics and services. Search engines tend to have a better coverage of the population and topics of past research, such as unemployment, are unlikely to be shared on social networks (D'Amuri and Marcucci, 2017). Most studies use data from Google Trends and fewer from Naver (Jun et al., 2017; Kim and Shin, 2016) or Baidu (Huang et al., 2017; Li et al., 2017) that are popular search engines in South Korea and China, respectively.

Microblogging platforms, such as Twitter (e.g. Bughin, 2015; Skodda and Benthaus, 2015; Rao and Srivastava, 2013) and Weibo (Chen et al., 2017) are the second most popular type of data source. The only study which involves social network sites is by Cui et al. (2018), who use Facebook. Bughin (2015) obtains social media information from SocialMention, a free aggregation service covering various platforms including Reddit. Another source is product reviews that have been collected from sources such as Amazon (Schneider and Gupta, 2016) or CNET (Luo and Zhang, 2013). Furthermore, Google search has been used to obtain forum data (Geva et al., 2017).

2.2.2 Forecasting applications

A typical application in economics is to forecast unemployment rate and claims. While various researcher report positive results (e.g. D'Amuri and Marcucci, 2017; Smith, 2016; Barreira et al., 2013) others struggle to improve accuracy (Li, 2016; Choi and Varian, 2012). Brynjolfsson et al. (2016) report benefits by stressing the importance of keyword selection. Researchers also look into housing market (Limnios and You, 2016; Wu and Brynjolfsson, 2015; Choi and Varian, 2009), private consumption (Vosen and Schmidt, 2011), exchange rates (Bulut, 2017), commodities (Yu et al., 2018; Elshendy et al., 2018) as well as consumer sentiment and gun sales (Scott and Varian, 2015). All but Limnios and You (2016) report improvements.

Financial applications include the predictions of financial market indices (Bollen et al., 2011), their returns (Perlin et al., 2017) or volatility (Dimpfl and Jank, 2016; Hamid and Heiden, 2015). Rao and Srivastava (2013) investigate stock market indexes but also currency exchange rate and gold prices. Other researchers forecast stock returns (Ho et al., 2017; Bijl et al., 2016). All studies report forecast improvements and Perlin et al. (2017) find search traffic to be particularly useful during the financial crisis.

When considering service-oriented applications, a large number of studies find improved forecast accuracy for tourism destinations (Li et al., 2017; Padhi and Pati, 2017; Park et al., 2017; Zeynalov, 2017; Choi and Varian, 2009) and attractions (Huang et al., 2017; Peng et al., 2016). In the case of Bangwayo-Skeete and Skeete (2015) search traffic outperforms the univariate benchmark only for one-third of the examined tourist destinations. Nonetheless, the authors conclude in favour of using this information, as in 77% of the cases accuracy was better or at least as good as the benchmarks. Önder (2017) did not find any clear indication whether search traffic is performing better on city or country level since. Her study reports improvements for both categories, but in several cases the benchmark is not outperformed (Önder and Gunter, 2016, report similar findings). Rivera (2016) reports that the benchmark is outperformed for 12 month ahead forecasts, but not for shorter ones. For hotel room demand forecasting, Pan et al. (2012) improve accuracy, but in a different study Pan and Yang (2016) find no statistically significant difference between using online information or not. Other service-oriented applications include Telecom contract sales (Bughin, 2015) or the number of air passengers (Kim and Shin, 2016).

A majority of studies that focus on consumer goods, forecast on aggregated brand or product category level. For example, Cui et al. (2018) report improvements for fashion sales forecast using sentiment information. Various studies also investigate car sales. Researchers find improvements from either search traffic (Carrière-Swallow and Labbé, 2013; Seebach et al., 2011) or forum posts (Geva et al., 2017). Fantazzini and Toktamysova (2015) report search traffic to be particularly helpful at longer forecast horizons. However, Choi and Varian (2009) report mixed findings and Choi and Varian (2012) as well as Barreira et al. (2013) conclude that there is little support for including search traffic. Also, the search traffic augmented model of Jun et al. (2017) which forecasts global netbook sales fails to outperform the benchmark. They also fail to improve forecast accuracy for Nintendo Wii sales, indicating that forecasting at product level is more challenging. Geva et al. (2017) for instance report increased errors due to additional noise in the data. Nevertheless, several studies report accuracy gains even at product level for speciality food Stock Keeping Units (SKUs) with search traffic (Boone et al., 2018, 2015), but also with sentiment information for electronic products (Lau et al., 2018; Schneider and Gupta, 2016) or fashion products (See-To and Ngai, 2016).

In the area of public health, various studies have been concerned with flu outbreaks. Ginsberg et al. (2009) incorporate highly correlated search terms to predict the flu index which became the Google Flu indicator. Its service was discontinued in 2015, partly because of data reliability concerns (Lazer et al., 2014; Butler, 2013). Despite the critique of Google Flu, studies find the combination of Google Trends and autoregressive terms leads to better results (Lazer et al., 2014; Preis and Moat, 2014). Moreover, with further refinement of keyword selection, there are additional improvements (Brynjolfsson et al., 2016).

Influenza outbreaks are also successfully forecasted with information from Twitter and blogs (Santillana et al., 2015; Won et al., 2013; Lampos and Cristianini, 2012).

Studies that focus on environmental events are sparse and online user-generated information is mainly used for posthoc analysis. Examples of usage are Lampos and Cristianini (2012), who predict the daily rainfall in the UK, and Chen et al. (2017) that predict smog hazards. Both use information from microblogs. It is worth noting that we were unable to identify studies that consider energy demand as an application, even though this is closely related to local weather conditions.

2.2.3 Forecast modelling

The majority of studies use linear regression models (Schneider and Gupta, 2016; Ginsberg et al., 2009), augmented with autoregressive terms (e.g Peng et al., 2016; Barreira et al., 2013; Seebach et al., 2011) and moving average terms (e.g. Li et al., 2017; Padhi and Pati, 2017; Pan and Yang, 2016). Linear vector models are also applied successfully (e.g Dimpfl and Jank, 2016; Fantazzini and Toktamysova, 2015). Other options include Bayesian structural models (Scott and Varian, 2015), dynamic linear models (Rivera, 2016) and Seemingly Unrelated Regression (Ho et al., 2017). Models that use the higher frequency of online information are also considered, e.g. Mixed-Data Sampling (Zeynalov, 2017; Smith, 2016; Bangwayo-Skeete and Skeete, 2015); or Dynamic Factor Models with mixed frequencies (Li, 2016).

Machine learning methods, typically incorporating sentiment information from social networks and review data (including mentions in forums), are also common. These include AdaBoost (Santillana et al., 2015), Support Vector Machines (Yu et al., 2018; Chen et al., 2017; Cui et al., 2018; Schneider and Gupta, 2016; Santillana et al., 2015), Random Forests (Cui et al., 2018) and Neural Networks (Yu et al., 2018; Chen et al., 2017; Lau et al., 2018; Geva et al., 2017; Bollen et al., 2011). These provide evidence of non-linearities in the relationships (e.g. Lau et al., 2018; Geva et al., 2017).

The majority of economic indicators are modelled with nowcasting models that include a contemporaneous internet variable to overcome the publication lag (except D'Amuri and Marcucci, 2017; Limnios and You, 2016; Barreira et al., 2013). Such models are also popular for forecasting influenza outbreaks (Xu et al., 2017; Preis and Moat, 2014; Ginsberg et al., 2009). However, some studies use nowcasting to forecast target variables in an operational context, which raises questions as to their usefulness due to the required lead times: for example, visitor arrivals (Huang et al., 2017; Zeynalov, 2017; Choi and Varian, 2009), Telecom sales (Bughin, 2015), car sales (Scott and Varian, 2015; Carrière-Swallow and Labbé, 2013; Choi and Varian, 2009) and fashion sales (See-To and Ngai, 2016). There are also studies which are not framed as nowcasting, but include contemporaneous inputs (Chen et al., 2017; Jun et al., 2017; Önder, 2017; Schneider and Gupta, 2016; Boone et al., 2015).

A key aspect of the model building is the specification of the user-generated information variables. The data is incorporated directly in the case of search traffic and count data, such as search popularity, the number of mentions or shares. Note that Google Trends and Naver provide peak scaled indexes, where different keywords compare relatively to each other (see Jun et al., 2017). Baidu, on the other hand, provides absolute search values (see Vaughan and Chen, 2015). Other inputs are ordinal such as product ratings (Schneider and Gupta, 2016) or sentiment information. See-To and Ngai (2016) incorporate sentiment information in the form of the absolute number of positive and negative reviews per period, whereas others use the ratio of positive and negative mentions per period (e.g. Geva et al., 2017; Skodda and Benthaus, 2015). If content-based information is available, it is also important to take into account the rating for the comment itself, i.e. by weighting up-votes for helpful reviews (Schneider and Gupta, 2016). Although several studies find sentiment to provide additional benefits over volume based information (Lau et al., 2018; Geva et al., 2017; Bughin, 2015), such gains are still debatable, given the additional complexity. Kübler et al. (2017) indicate that the required method, as well as choice of metrics, depend on brand strength and industry segment. We discuss keyword selection and sentiment measure in more detail in Sections 2.3.3 and 2.3.4.

2.2.4 Forecast evaluation

The forecasting literature has established several forecasting principles that make the interpretation and comparison of forecasts more transparent. These include the need for adequate benchmarks (Armstrong, 2006; Armstrong and Collopy, 1992) and hold-out sample evaluation with rolling origins (Tashman, 2000). The selected error metrics should be conditional on the forecasting objective and a number of alternatives should usually be included as the result maybe contradictory (Davydenko and Fildes, 2013; Fildes and Ord, 2002). For example using relative error metrics when the objective is to compare models. Koning et al. (2005) also stress the importance of statistically testing the performance of competing models. The surveyed literature adherence to these practices in a mixed manner, as Table 2.1 depicts.

There are issues about the clarity of the experimental setup. For example, Yu et al. (2018), Araz et al. (2014) and Won et al. (2013) provide very little details about the model specifications of user-generated variables. Various studies are also unclear on the set up of the evaluation sample (Chen et al., 2017; Jun et al., 2017; Önder, 2017; Choi and Varian, 2009). Although, most studies use hold-out samples with rolling origins some of the studies evaluate them on very few observations (Elshendy et al., 2018; Araz et al., 2014). Other studies do not report extensive forecast results, which makes it difficult to identify any performance improvements (Bangwayo-Skeete and Skeete, 2015; Carrière-Swallow and Labbé, 2013).

If claims are to be made about the generalisability of the results, multiple time series should be used. As the appendix attests, less than half of the investigated literature report results for multiple series, and some of the remaining studies do not investigate more than two series (e.g. Rao and Srivastava, 2013; Seebach et al., 2011). However, in some cases carrying out experiments for multiple time series is not possible for applications that focus on highly aggregated variables. Nonetheless, these could, for example, be split into regions to provide more robust results such as in (Bulut, 2017; Önder and Gunter, 2016; Bangwayo-Skeete and Skeete, 2015; Lampos and Cristianini, 2012). One-third of the surveyed literature also includes statistical testing of the forecast results that typically strengthen their findings. However, in the case of Bulut (2017) they lead to contradictory results since none of the search traffic augmented models outperforms the random walk on the MSPE, but the test find them to be significantly better. The contradiction maybe explained by the distribution of the forecast errors. This led the authors to still draw a positive conclusion on the usefulness of search traffic data.

A further issue is that the conditionality of forecasts is unclear. For example, it is unclear whether Hand and Judge (2012) use a 4 observation long test set in a rolling origin manner or whether the horizon is set to four. The study by D'Amuri and Marcucci (2017) provides 12-month ahead forecast, but the maximum lag length of search traffic is four, requiring unseen future information. Several studies also lack clarity as to whether the future search traffic volume is considered as known or not in the test set (Önder, 2017; Önder and Gunter, 2016; Bangwayo-Skeete and Skeete, 2015; Won et al., 2013). Li et al. (2017) report significant improvement for 4-weeks-ahead forecasts, using 5 lags of search traffic, but are unclear if the values of the shorter lags were considered known or not. Barreira et al. (2013) indicate that the 36 month out-of-sample forecast uses future values. Less than one-third of the surveyed studies considered multistep-ahead forecasts. It is questionable how relevant one-step-ahead forecasts are in a business context that for example require stock keeping (Boone et al., 2018; Lau et al., 2018; Geva et al., 2017; See-To and Ngai, 2016; Seebach et al., 2011).

A further critique of the existing literature is that studies often fail to provide a thorough comparison with adequate benchmarks. For example the studies by Kim and Shin (2016); Won et al. (2013); Lampos and Cristianini (2012); Ginsberg et al. (2009) and partly Choi and Varian (2009) report no benchmarks at all. Lau et al. (2018); Xu et al. (2017); Peng et al. (2016); Skodda and Benthaus (2015) and Hand and Judge (2012) only compare forecast performance amongst models that include online information. Most papers use at least one benchmark that is the univariate equivalent of the proposed model using the additional internet variables. However, established, and common in practice models, such as exponential smoothing or the random walk, are often absent. If such benchmarks outperform both the univariate and the enhanced models, then there is little value in them. Therefore, the apparent lack of a thorough (or even valid in some cases) forecast accuracy evaluation diminishes the value of the reported improvements.

To exemplify this, Cui et al. (2018) report gains over the company forecast, but there is too little information on how the company forecast is produced or whether it was any good at all. This critique echoes the arguments by Li (2016), Önder (2017) and Fantazzini and Toktamysova (2015), who all report cases where the random walk outperforms models that use search traffic information for some evaluation periods. Jun et al. (2017) and Rivera

(2016) similarly find that the simple Holt-Winters method performs better than forecasts that used additional internet information. In a study by Lazer et al. (2014) the Google Flu index model is outperformed by a univariate model. A further downside of not using established benchmarks is that it makes any meta-analysis of performance very difficult. Including the random walk would help to draw overall conclusions.

To further illustrate the importance of including a wide variety of benchmark models we replicate one of the experiments conducted by Choi and Varian (2012) and extend its range of contenders. In addition to the proposed seasonal autoregressive model, we further include the random walk (RW), as well as the Simple Exponential Smoothing model (SES) and the Holt-Winter model (HW). Table 2.2 provides the Mean Absolute Percentage Errors (MAPE). The result suggests that the Holt-Winters model performs best in both evaluation periods. Furthermore, none of the models differ significantly at a 5%-level when evaluated with the Friedman and Nemenyi tests (Demšar, 2006).

Table 2.2: MAPE for motor vehicles and parts (Choi and Varian, 2012)

	AR	*ARX*	RW	SES	HW
06/2005 - 07/2011	*6.34%*	*5.67%*	6.88%	6.70%	**4.37%**
12/2007 - 06/2009	*8.87%*	*6.97%*	5.87%	5.75%	**4.84%**

leadtime = 1, *italic* signifies original models

In the introduction, we posed the question whether such predictive information remains relevant over the life-cycle of a product or service. There is some evidence from the marketing literature that reports the impact of social network variables changing over time, due to changes in the level of customer engagement (Kumar et al., 2016). Smith (2016) finds changing coefficients of Google Trends indicators, some switching from positive to negative, over the life-cycle. It is unclear whether this indicates a spurious or changing relationship. Experiments which have included a rolling window evaluation with re-estimation do not provide insights on the changes of the coefficients and in particular do not discuss the life-cycle aspect (e.g. Cui et al., 2018; Geva et al., 2017; Bughin, 2015).

2.2.5 Summary

To summarise the literature review we note that a majority of investigated papers report positive findings for all types of user-generated data sources. The most frequently applied models are linear in the form of an ARX model, both in nowcasting and forecasting. However, the conclusions of these studies must be tempered by their many limitations, in particular the absence of adequate benchmarks, lack of model transparency as well as what information the forecasts are conditional on. To be useful in operational planning decisions they also need to be focussed on a meaningful forecast horizon. Given these weaknesses in the forecast evaluation framework, we therefore cannot conclude as to which applications are likely to benefit from user-generated information.

2.3 Handling user generated online information

2.3.1 Data consistency and reproducibility

Reproducing the results of forecasting experiments is a major concern for research (Boylan et al., 2015). Lazer et al. (2014) question how stable and reliable are measurement sources such as Google Trends over time. For instance, changes in the search algorithm employed by Google can disrupt the performance of predictive models. Such changes are dependent on decisions by the search engine provider that might be based on commercial interests. Changes in search algorithms not only require model re-calibration, but also hinder scientific replication. Recently, Google restricted the maximum window length for weekly data to 5 years. Hence, to obtain weekly data from 2004, stitching and re-scaling are required (Johansson, 2014, provides a tutorial with one way of combining). This increases the risk of obtaining different values for the search traffic.[1]

Furthermore, Google Trends index depends on samples which are re-drawn from day-to-day (Varian, 2017). According to Barreira et al. (2013) this sampling instability explains some of the inconsistencies in the results of their now-casting exercise. Carrière-Swallow and

[1]We were not able to find any official changelog of Google Trends but the issue is discussed in forums; for example https://www.en.advertisercommunity.com/t5/Water-Cooler/Inconsistent-trends/m-p/1175227 or https://productforums.google.com/forum/#!topic/websearch/HVYS9OnEjOo;context-place=topicsearchin/websearch/category$3Amac

Labbé (2013) report all queries within 24-hours to be identical, but across a 50-day sample, the same query sample exhibit a standard deviation of more than 15%. Although, D'Amuri and Marcucci (2017) report that the cross-correlation between series of different days is never below 0.99, they take the average of 24 downloads over 12 days from two different IP's for forecasting unemployment rate. Li (2016) replicates one of the experiments by Choi and Varian (2012) but achieves different out-of-sample forecasts between the original data and the newly obtained sample, highlighting issues of sample instability from Google Trends that makes the replication of experiments more difficult. Li (2016) suggests that taking multiple samples is a good solution, but it is unknown how many samples are needed to approximate the "true" sample.

The research of Lazer et al. (2014) also points out that other platforms have similar issues. For example, a study by Ho et al. (2017) reports that they were unable to report the number of messages prior to 2011 due to changes to the Yahoo!Finance website. Ruths and Pfeffer (2014) raise concern that social media platforms can enforce changes in data streaming and filtering. For instance, the additional "like"-buttons Facebook introduced to express emotions have an unknown effect on data continuity. Although for practitioners reproducibility is a minor concern, the reliability of the models and the need for continuous monitoring of the specifications is of importance.

2.3.2 Data bias

One of the disadvantages of user-generated information is potential selection sample bias. This bias exists on all platforms and affects search traffic, product reviews as well as social network platforms (Brynjolfsson et al., 2016; Ruths and Pfeffer, 2014). This is because the platforms are not accepted equally in all countries, and furthermore, not the entire population is using the platform equally often. For example, Brynjolfsson et al. (2016) mentions the case that elderly people might not use online technologies to search for products and services. This makes the right choice of platform crucial in order to align with the forecast target.

Bias not only appears in the representation of the population, but also in terms of content type. On social network platforms, such as Facebook, users tend to share a positive image (Barash et al., 2010), and research suggests that negative feelings are more likely to be expressed on forums (Leung, 2013). Moreover, not all customers write reviews and the reflected opinion might not represent the overall opinion of customers. Dellarocas et al. (2007) report customers with strong positive or negative opinion are more likely to post. Moreover, reviews from early adopters have been found to be systematically positively skewed due to potential self-selection bias and the fact that early buyers may have different preferences and requirements than late buyers. Therefore, ratings generally tend to decline over the product lifetime (Godes and Silva, 2012; Li and Hitt, 2008) which impacts sales (Moe and Trusov, 2011).

The often reported J-shaped distribution of online ratings (e.g. Schneider and Gupta, 2016) can have many sources including fraud, selection bias or herding effects (Aral, 2014). Fraud might be due to manipulation by companies and their competitors. Mayzlin et al. (2014) find evidence of fake hotels reviews on Tripadvisor with negative reviews by competitors, but also positive ones created by the owners. Lee et al. (2018) shows that in the movie industry Twitter sentiment is often positively manipulated in the pre-launch phase and drops after release when actual viewers comment. Such manipulation may not only impact sales, but also affect the willingness to post and, therefore, change the final product perception (Moe and Schweidel, 2012). Positively manipulated reviews lead on average to 25% increased final ratings, suggesting an asymmetric herding bias (Muchnik et al., 2013).

Although, these biases are well studied, very little is done to address them in forecast models. Nonetheless, cleansing data post-hoc, might eliminate important signals, since a manipulated negative review that is still online will potentially affect sales and future reviews. Even if it were removed, it is hard to track how it has affected other remaining reviews.

2.3.3 Keyword selection

One of the major complications of using search traffic information is to select keywords (Goel et al., 2010). That keyword selection matter is demonstrated in the research by Brynjolfsson et al. (2016) discussed before. Geva et al. (2017) describes keyword selection to be a trade-off between *accuracy* and *coverage*. Studies that tried to incorporate a very high coverage (Scott and Varian, 2015; Ginsberg et al., 2009) base their selection to identifying keywords with the highest correlation from very large datasets (using Google Correlate one can find correlated search queries to any given time series). While this method effectively filters amongst million of possible queries, it remains prone to return spurious correlated time series (Lazer et al., 2014) and requires a well-designed forecast evaluation to prevent over-fitting. It also introduces major variable selection challenges due to the number of multi-collinear inputs.

A large part of the investigated literature uses a judgemental selection based on only a few keywords, such as product or brand name (e.g. D'Amuri and Marcucci, 2017; Geva et al., 2017; Seebach et al., 2011) or words like "dow" for Dow Jones Index (e.g. Dimpfl and Jank, 2016; Hamid and Heiden, 2015). Other studies use more descriptive keywords for example "Gifts for colleagues" to predict a wine and cheese SKU (Boone et al., 2018) or "Vacation" to reflect economic income (Bulut, 2017). While this approach allows a better interpretation of variables selected it might miss out important information. To broaden the numbers of keywords Li et al. (2017) and Peng et al. (2016) use a seeding technique. They initially define a range of keywords that was then used in a second step to gather recommended keywords by the search engine. Perlin et al. (2017) count the frequency from a large list of financial specific words in academic books to derive from a list of 15 selected words. Similarly, Padhi and Pati (2017) identify 63 keywords from different literature sources and interviews with destination clients. Researchers also tried to identify specific keywords to obtain pre-purchase searches only. Von Graevenitz et al. (2016) for instance use scrappage subsidies searches as pre-purchase indicators of new car purchases. Hu et al. (2014) use composite search queries that excluded unrelated search keywords for new car sales such as "repair". Siliverstovs and Wochner (2018) use Google Knowledge Graph that covers linguistic and semantic related keywords to a topic. For example, it can combine search queries for a place covering different languages.

Another approach is to use automatic generated categories that search engines provide. These categories cover several related keywords for areas like travel destinations or industry sectors (e.g. Von Graevenitz et al., 2016; Bughin, 2015; Fantazzini and Toktamysova, 2015; Scott and Varian, 2015; Wu and Brynjolfsson, 2015; Vosen and Schmidt, 2011; Choi and

Varian, 2012). Brynjolfsson et al. (2016) criticise such categories being opaque and might include irrelevant keywords that could harm the predictive ability. Instead, they suggest a crowd-sourcing approach. They asked more than 500 persons to write down five terms that came to their mind when seeing a particular word. Not only did they achieve higher forecast accuracy, but they also found that the forecast accuracy improves steadily when increasing the number of selected variables (up to 20). This result is in contrast to selection via Google Correlate and WordNet lexical database, where forecasting performance decreases when additional variables are added, indicating poor selection in these cases. Although crowd-sourcing via services such as Amazon Turks might be relatively cheap, it can quickly become expensive if keywords for several hundreds of products are required.

Another approach is judgemental pre-selection, which has not been applied to keywords selection yet. Sagaert et al. (2018) report that for selecting macroeconomic leading indicators using experts to pre-select a set of variables leads to forecast accuracy gains over using the full set of variables, with LASSO modelling.

2.3.4 Sentiment analysis

With sentiment analysis one can investigate the opinion towards an entity within a written text, for instance, the attitude people have towards a brand or product. It differs from count or popularity data in that it captures a sentiment orientation (also called valence), classified into positive, neutral or negative (Liu, 2015). Some of our surveyed studies, introduce further levels to describe intensity or strength of the sentiment (e.g. Hou et al., 2017; Skodda and Benthaus, 2015) or capture mood dimensions (Bollen et al., 2011).

The sentiment can either be self-declared (Ho et al., 2017) or derived with additional analysis. Studies use content analysis (e.g. Geva et al., 2017; Cui et al., 2018), measure the text complexity (Elshendy et al., 2018), or count n-grams for messages (Liu et al., 2016; Lampos and Cristianini, 2012).

There is a large variety of methods for classifying sentiment. Typically, the manual approach is very time-consuming (e.g. Liu, 2006) and therefore, text mining algorithms are common. One can derive classification rules by training bespoke sentiment classifiers using machine learning methods or use pre-defined lexicons. The lexicons are typically based on language and slang dictionaries, but can also be built to cover domain-specific knowledge (e.g. Chen et al., 2017; Tirunillai and Tellis, 2012). There are lexicons built on semi-supervised classifiers such as SentiWordNet (Baccianella et al., 2010). These are popular due to their simplicity and reproducibility (Geva et al., 2017; See-To and Ngai, 2016; Rao and Srivastava, 2013). Lau et al. (2018) provide a comparison between different sentiment classification algorithms. They find that for forecasting product demand most lexicons are not granular enough to reflect consumers preferences well and suggest the use of abstract based classifiers, i.e. the sentiment is measured for each aspect (feature) of the product individually such as for the battery or screen.

The survey of Ravi and Ravi (2015) also highlights various limitations of sentiment analysis, one being that current methods still struggle with irony and sarcasm. Together with spelling mistakes, data becomes noisy, and a significant amount of manual intervention and supervision is required. This raises the question of how well-suited reviews are for forecasting tasks, when operational costs are considered. For this reason, Schneider and Gupta (2016) suggest using a bag-of-words model, which counts the frequency of each word, together with dimensionality reduction techniques. This method is computationally fast and able to run almost unsupervised. However, Cui et al. (2018) points out that bag-of-words classifiers are not well suited for short and heterogeneous text such as often seen in social networks comments. We are unaware of any research that compares demand forecasting performance of bag-of-words models against lexicon or machine learning methods.

2.4 Empirical evaluation

2.4.1 Case studies

Based on our review of the literature we argue that it is not possible to assert conclusively about the benefits of search traffic or social network information. More specifically, we are interested in the application to operational forecasting, as there is limited research on this area. We attempt to answer whether online platform information is useful by conducting an empirical evaluation using two distinct case studies. First, we look at forecasting physical video games sales using search traffic information from Google Trends, throughout the product life-cycle. Second, we aim to forecast YouTube views of corporate viral online videos using social network shares.

We have selected these two case studies due to the nature of the target variables. Although direct sales of video games over the internet are increasing, roughly three-fifths are still sold as physical copies (statistic for the US market, Statista, 2017). Accurate demand forecasts are, therefore, vital for the supply chain management. Concerning the second case study, the very nature of viral videos implies that social network shares drive the process (e.g. Abisheva et al., 2014; Broxton et al., 2013; Crane and Sornette, 2008). Corporate videos are used to promote the offered services and products, where together with their virality can be considered as the electronic word of mouth (Babic Rosario et al., 2016), which in turn support sales. Knowing future video views helps marketers to plan and adjust their advertising activities (Liu-Thompkins, 2012).

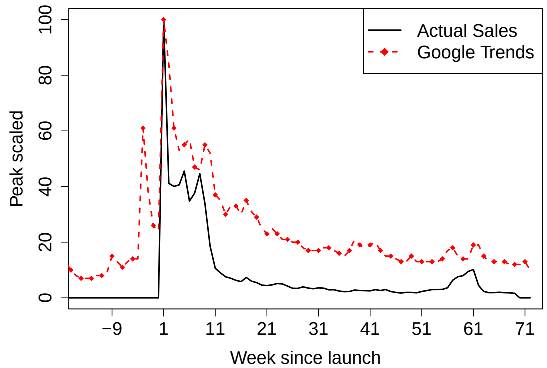
2.4.2 Data

The first dataset consists of 78 global physical video game sales on a weekly frequency. The data was obtained from VGChartz; a company specialised in collecting physical video game sales (http://www.vgchartz.com). The same data provider has been used by various researchers (e.g. Ruohonen and Hyrynsalmi, 2017; Xiong and Bharadwaj, 2014; Goel et al., 2010). The video games considered were launched after November 2005 and belong to different genres, including blockbuster titles such as the Call of Duty or the FIFA football game series. We cover the period of sales up to February 2015 and limit the length of the time series up to the point that 95% percent of the total recorded sales is reached, to filter out high intermittency observed towards the end of the life-cycle. The median length of time series is equal to 160 weeks (minimum 66 and maximum 447 weeks). For each game title, we downloaded the corresponding Google Trends data (www.google.com/trends). For our dataset, we downloaded the Google Trends information on a weekly frequency and used the game title as the search keyword. Where available, we used "Topic search" over "Search term". This option, provided by Google, makes use of Google's Knowledge Graph Search API and combines several keywords associated with the topic for different languages. We find that "Topic search" typically correlates better with our target variable.

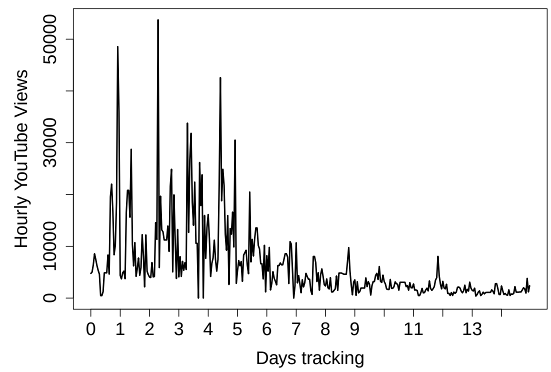
The second dataset contains viral corporate online videos. We collected videos views by building a web crawler that tracked corporate YouTube channels using the Google's YouTube Data API, over the period from March 2015 to April 2016. Each time a new video was published on the YouTube channel of an organisation, the crawler started tracking cumulative views at a 15 minutes interval. In addition to video views, we also collected the cumulative number of shares from Facebook, Twitter, Google+ and LinkedIn using the YouTube video URL as a unique identifier. Note that at times there were outages either at the Google API or our server, introducing missing values. These were imputed using linear interpolation. Furthermore, we noticed that Google adjusts YouTube view counts on an irregular, but quite frequent, basis. This can result in a negative change of the cumulative views, which should not be possible. We believe this is due to algorithms used to avoid artificial or erroneous view counts from bots and synchronisation errors. In order to remove these effects, we treated these as missing values and used linear interpolation to impute them.

We selected the 300 most shared videos on Facebook from our dataset. From the selected videos many exhibit substantial amount of intermittent views towards their mature phase. Similarly to the video games dataset, we have shortened the series when a certain threshold of zero views has been reached. We model the series both at an hourly and an aggregate daily level. In the case of the hourly dataset, this was set to 12 continuous zero observations, while for the daily dataset this was set to 6. Some videos were excluded as they did not have a sufficient number of observations to facilitate a thorough evaluation. The total is further reduced as we only compare time series which contain enough shares in at least two social networks. This allows us to investigate Facebook shares versus further social networks. The final dataset consists of 63 videos with an average 122 days of observations (minimum 72 and maximum 179 days).

We provide two example time series for the two data types in Figure 2.1. The example containing sales and search traffic is scaled for illustration purposes. Note that for clarity the example of the YouTube video is without social shares. For the "GoPro - Best of 2015" video clip we captured 3.6 million views and more than 25 thousand mentions on social networks (www.youtu.be/IyTv_SR2uUo).



(a) Battlefield 4 video game



(b) GoPro YouTube video

Figure 2.1: Sample time series

2.4.3 Experimental setup

Our aim is to assess the predictive usefulness of internet variables, across the life-cycles of the products, and follow the requirements laid out in Section 2.2.4. In order to facilitate this, we employ a rolling window approach. The window has a fixed size of w observations that rolls up to the point $T-h$ where T is the available sample size, and h is the forecast horizon. For each dataset, we consider a number of different window sizes and forecast horizons (Table 2.3). Smaller window sizes resulted in very poor forecasts and were excluded. The rolling window setup allows us to consider the launch phase or the mature phase of the life-cycle of a product separately.

At each forecast origin, we construct forecasts that rely on the additional internet inputs and appropriate univariate benchmarks. For the additional variables to be useful, they have to lead to more accurate out-of-sample forecasts. We assess the performance at each forecast origin using the Average Relative Mean Absolute Error (AvgRelMAE; Davydenko and Fildes, 2013):

$$\mathrm{AvgRelMAE}_{i,h} = \sqrt[n]{\prod_{r=1}^{n} \left(\frac{\mathrm{MAE}_{i,r}}{\mathrm{MAE}_{b,r}} \right)},$$

$$\mathrm{MAE} = \frac{1}{j} \sum_{t=1}^{j} |y_{t+h} - \hat{y}_{t+h}|,$$

where n is the number of times series and j the number of forecast origins for each series. First, the Mean Absolute Error (MAE) across all origins, for a given time series and horizon is calculated for each forecast i. These are then divided by the MAE of the Naïve forecast ($\text{MAE}_{b,r}$) and summarised using a geometric mean to produce the reported AvgRelMAE for each horizon.

This metric has favourable statistical properties and provides an intuitive comparison between forecasts, where an improvement over the benchmark is given by value lower than 1. Subtracting the AvgRelMAE from 1 provides the percentage accuracy gain of a forecast over the benchmark.

Finally, to evaluate whether any differences are due to randomness or not, we employ the non-parametric Friedman test and the post-hoc Nemenyi test (Koning et al., 2005; Demšar, 2006). We use the Friedman and Nemenyi tests as implemented for R (R Core Team, 2016) in the TStools v.2.1.0 package (Kourentzes and Svetunkov, 2016).

Table 2.3: Experimental settings for the different datasets

Dataset	Window sizes (w)	Forecast horizons (h)	Explanatory lags (l)
Video games	$20, 24, \ldots, 52$	$1, 6, 12$	$1, 2, \ldots, 6$
Online videos (daily)	$20, 24, \ldots, 72$	$1, 6, 12$	$1, 2, \ldots, 6$
Online videos (hourly)	$24, 36, \ldots, 120$	$1, 12, 24$	$1, 2, \ldots, 72$

2.4.4 Methods

We use the following regression model:

$$y_t = \alpha_0 + \sum_{i=1}^{m} \alpha_i y_{t\text{-}i} + \sum_{j=1}^{k} \beta_j x_{t\text{-}j} + \varepsilon_t, \tag{2.1}$$

where y_t is the target variable and x_t is the explanatory online information variable; m and k represent the number of autoregressive terms and numbers of lags of the explanatory, respectively, and ε_t is a Gaussian zero-mean error.

As proposed by Hyndman and Khandakar (2008), we make our series stationary by using the KPSS and the OCSB tests to identify level and seasonal-differences, respectively. This also eliminates any spurious connections between y_t and x_t.

The challenge in (2.1) is the specification of m and k. Furthermore, one can consider sparse specification, as not all lags may be informative (Hastie et al., 2015). In the aforementioned literature different approaches have been employed to specify the relevant lags. Granger causality is one of them (e.g. Ruohonen and Hyrynsalmi, 2017; Tirunillai and Tellis, 2012). Another popular modelling approach is to use information criteria, such as AIC, to identify the lag-order (for example, as in Hyndman and Khandakar, 2008). However, note that in the presence of explanatory variables the number of potential models becomes prohibitive very quickly. A stepwise approach can be used to manage the problem, however the stepwise search strategy has been criticised for inadequate search of alternatives, due to its greedy search nature (Hastie et al., 2015). The problem is exacerbated further by limited sample size.

Considering the case where all social network information is available, in the extreme case, our model needs to estimate up to 297 parameters using only 24 observations. To solve this problem we rely on lasso regression that provides an effective and efficient search of the model space and achieves sparsity, if needed, even when the number of coefficients exceeds the available sample size. Lasso works by penalising the model fit with the absolute of the sum of the coefficients, scaled by a shrinkage factor. This forces the coefficient of uninformative variables to zero. For details of lasso, as well as a discussion of alternative selection schemes see Hastie et al. (2015). We fit the lasso regression using R and the package glmnet v.2.0-5 (Friedman et al., 2016) with its default settings.

Hereafter, we refer to these forecasts as ARX for the video games dataset and ARX (FB) or ARX (All) if only shares in Facebook or more platforms are considered for the video dataset.

We allow up to 6 autoregressive terms. For the hourly dataset we include additionally up to 3 seasonal autoregressive terms. Furthermore, the model is augmented by up to l lags of the explanatory variables (Table 2.3). To simulate a true forecasting situation we restrict the included lags to always be of order at least equal to the forecast horizon or longer, as the in-between values would not be available. For example to forecast 3-steps ahead only lags of order 3 or more are considered, as shorter lags would imply knowledge of the future values of the explanatory variable.

To further complete our experiment we also discuss the case where we allow contemporaneous explanatory variables in our model, producing now-casting results. Although this is of limited operational benefit, it allows us to relate our experiment with the nowcasting literature that has used such variables.

We compare our ARX forecasts against various benchmarks from different model families. The first, represents the univariate autoregressive model that uses the same specification method to ARX. Second, we include an ARIMA model, the orders which are identified using AIC corrected for sample size (AICc), based on the model selection procedure by Hyndman and Khandakar (2008). Furthermore, we use exponential smoothing (ETS), the form of which is automatically selected by using AICc (Hyndman et al., 2008). Finally, we include a Random Walk (Naïve) forecast. In cases where the hourly online video time series is seasonal, we further add a seasonal Naïve as a benchmark. The benchmarks are implemented using the forecast v.7.2 package for R (Hyndman, 2016).

2.4.5 Results

Overall results

Table 2.4 presents the results for window sizes $w = \{20, 24, 52\}$ and for forecasting horizons $h = \{1, 3, 6\}$ weeks, across the complete life-cycle for both video game and online video dataset. Results for other tested windows between 24 and 52 weeks are very similar and therefore omitted. The striking result is that the Naïve is consistently the best or at least as good (with no significant statistical differences) as its competitors, followed closely by ETS and ARIMA.

In most cases, the worst performing model is the simple AR that is outperformed by the ARX, on average by about 2.8% on the video game and 2% for online video dataset. We find, no evidence that this difference is significant. On the one hand, this supports findings from the literature that search traffic can improve forecasts, but on the other hand, it also verifies our criticism of weak experimental design. When benchmarked against more appropriate univariate alternatives, here all ETS, ARIMA and Naïve, we cannot support that conclusion. Closer examination of the individual time series reveals that in the presence

Table 2.4: Overall AvgRelMAE forecasting performance across all origins

	$w = 20$			$w = 24$			$w = 52$		
	$h = 1$	$h = 3$	$h = 6$	$h = 1$	$h = 3$	$h = 6$	$h = 1$	$h = 3$	$h = 6$
				Video games: $n = 78$					
ARX	1.188	1.184	1.252	1.171	1.174	1.243	1.005	1.007	**0.985**
AR	1.147	1.211	1.339	1.148	1.210	1.279	1.001	1.023	0.999
ARIMA	1.069	1.075	1.098	1.072	1.084	1.106	1.094	1.095	1.097
ETS	1.066	1.072	1.090	1.072	1.076	1.092	1.038	1.045	1.040
Naïve	**1.000**†	**1.000**†	**1.000**†	**1.000**†	**1.000**†	**1.000**†	**1.000**	**1.000**†	1.000
				Online videos (daily frequency): $n = 63$					
ARX (All)	1.185	1.200	1.295	1.165	1.238	1.217	1.246	1.275	1.314
ARX (FB)	1.184	1.193	1.290	1.174	1.181	1.204	1.289	1.275	1.317
AR	1.187	1.203	1.251	1.216	1.223	1.276	1.305	1.303	1.340
ARIMA	1.108	1.078	1.078	1.096	1.066	1.064	1.121	1.100	1.115
ETS	1.082	1.074	1.075	1.104	1.081	1.086	1.290	1.233	1.201
Naïve	**1.000**†	**1.000**†	**1.000**†	**1.000**†	**1.000**†	**1.000**†	**1.000**†	**1.000***	**1.000***

† Different at 5%-significance level to ARX and all other benchmark models.
* Different at 5%-significance level to ARX model

of adequate benchmarks there is no case where ARX ranks first across all benchmarks, but it is easy to identify a single benchmark that would typically be worse than ARX. The need for thorough benchmarking has been fundamental in forecasting research (Armstrong and Collopy, 1992) and contrasting our results with the mostly positive impression from the literature helps to highlight how important that is.

High frequency and nowcasting

Table 2.5 provides the results for the hourly time series. Although the forecast horizons are now too short to support many operational decisions, looking at higher frequency data allows us to explore whether intra-day lags may be more informative. In this scenario, although the Naïve is no longer best, overall we do not observe benefits from including the additional variables. In fact, AR is in all cases more accurate than either ARX (All) or ARX (FB). For longer window sizes ($w = 120$) ARX (FB) outperforms the Naïve, but is in turn outperformed by other univariate benchmarks.

Table 2.6 presents the now-casting results. For convenience, we provide the one-step-ahead forecast errors as well. The difference in the specification between the two is that the latter permits contemporaneous inputs of the variables. The results suggest that there are improvements, yet still the Naïve is more accurate for both datasets.

Table 2.5: Overall AvgRelMAE forecasting results online videos (hourly) $n = 63$

	$w = 24$			$w = 72$			$w = 120$		
	$h = 1$	$h = 12$	$h = 24$	$h = 1$	$h = 12$	$h = 24$	$h = 1$	$h = 12$	$h = 24$
ARX (All)	1.083	1.114	1.231	1.080	1.066	1.114	1.021	1.011	1.029
ARX (FB)	1.057	1.113	1.222	1.018	1.018	1.052	0.981	0.996	1.018
AR	1.002	1.058	1.198	0.951	0.953	0.972	0.941	0.935	0.939
ETS	**0.953***	**0.970***	0.993	0.943	**0.901***	0.913	0.942	**0.853***	**0.856***
ARIMA	0.984	1.000	1.032	**0.942**	0.902	**0.894***	**0.935**	0.881	0.871
Naïve	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
sNaïve	1.407	1.041	**0.975***	1.368	1.004	0.953	1.363	0.992	0.945

* Different at 5%-significance level to ARX model.

Table 2.6: Now-casting versus one step-ahead AvgRelMAE forecasting performance

	Forecasting			Now-casting		
	$w = 20$	$w = 24$	$w = 52$	$w = 20$	$w = 24$	$w = 52$
	Video games					
ARX	1.188	1.171	1.005	**1.145**	**1.124**	**1.004**
	Online videos (daily frequency)					
ARX (All)	1.185	1.165	**1.246**	1.175	1.193	1.286
ARX (FB)	1.184	1.174	1.289	**1.164**	**1.162**	1.278

Performance across life-cycle stages

Since the search traffic seems not to add much value over the entire life-cycle, we investigate different life-cycle stages. Table 2.7 presents the results for the scenario of three-week-ahead forecasts with a window size of 20, for different weeks since launch. Recalling the typical nature of the demand pattern shown in Figure 2.1a, one would expect that search traffic information would be particularly useful towards the beginning of the life-cycle, where there are lots of spikes. However, as we can see from the results, ARX performs poorly for the first few origins and only towards the end of life it starts to outperform the simpler AR model. In this sense, the forecasting performance of the search traffic model is much worse during the first year of sales than any univariate model. We found this behaviour to be consistent with other window sizes and forecasting horizons.

Table 2.8 provides the forecasting results for the online videos across life-cycle. We classify the life-cycle phases according to Rogers (2003), with the splits accounting for the percentage of total views. The results show that social network information is not able to deliver additional forecasting performance in any of the life-cycle phases and the forecasting

Table 2.7: Forecasting AvgRelMAE performance for video games over life-cycle

	Weeks after launch			
	28-35	36-43	44-51	52-EOL
ARX	1.665	1.308	1.200	1.110
AR	1.560	1.292	1.184	1.161
ETS	1.046	1.119	1.046	1.071
ARIMA	1.244	1.190	1.081	1.049
Naïve	**1.000***	**1.000**†	**1.000**†	**1.000**†

$w = 20$, $h = 3$
† Different at 5%-significance level to ARX and all other benchmark models. * Different at 5%-significance level to ARX model.

performance is quite consistent during the entire life-cycle, apart from the innovator phase.

Table 2.8: AvgRelMAE Forecasting performance for online videos over life-cycle

	Life-cycle phases (%-of life-time views)				
	Innovators (2.5%)	Early adaptors (13.5%)	Early majority (34%)	Late majority (34%)	Laggards (16%)
ARX (All)	1.043	1.093	1.121	1.082	1.109
ARX (FB)	1.021	1.096	1.125	1.084	1.124
AR	1.035	1.051	1.052	1.023	1.086
ETS	1.058	1.012	**0.940***	**0.954***	**0.940***
ARIMA	**0.981**	1.010	1.001	0.985	1.008
Naïve	1.000	**1.000**	1.000	1.000	1.000
sNaïve	1.154	1.016	1.083	1.023	1.005

$w = 24$, $h = 12$
* Different at 5%-significance level to ARX model.

2.5 Discussion

2.5.1 Reasons for poor performance

The reader may ask why did the models with the explanatory variables from online platforms perform so poorly compared to the benchmarks? Or why did the Naïve perform that well? Both datasets contain noisy time series with demand spikes, due to renewed interest by the consumers/viewers. Such time series are notoriously difficult to predict without causal information, which can explain the competitiveness of the Naïve forecast against the other univariate forecasts. This paper set out to evaluate the usefulness of online platform variables for this purpose. We found that in many cases the ARX forecasts outperformed some of the benchmarks, but in no cases, all of them, and overall the impression was that the inclusion of these variables helped only marginally, if at all.

Although our empirical evaluation has its limitations, and we do not claim that the results generally hold for other applications and datasets, it should encourage researchers and practitioners to think critically about the predictive capabilities of such data.

As discussed in Section 2.2, a large number of publications were not strictly in a predictive setup, or when such was used, the forecast horizons were too short to support operational decision making. Requiring forecasts for longer horizons implies an expectation that any causality between internet search traffic or social network shares and sales will hold. However, it is not uncommon that internet searches and buying decisions are made instantly or with a very short lag. Such impulsive buying decisions do not allow the manager to take any reactive operational decisions. Our experiments support this interpretation and also agree with the findings by Ruohonen and Hyrynsalmi (2017) who raise similar concerns. It is unfortunate that the surveyed literature has mostly neglected this; reporting forecasting results that do not match realistic applications does not add new insights into the usefulness of such information.

As we have highlighted in Section 2.2.4, an unhelpful characteristic of many of the published papers has been their weak experimental design, in particular concerning their choice of benchmarks. In our two case studies, we found many cases were the ARX model outperformed single benchmarks, typically its univariate equivalent, but when tested against a set of well known and reliable univariate models it was never the best performing. We have stressed the need for thorough evaluation (Armstrong, 2006; Armstrong and Collopy, 1992). However many publications, in this relatively new modelling research, come from various disciplines that do not strongly adhere to these principles. Therefore, it is important to retain a critical view of the usefulness of such information against well established and tested forecasting models. This is particularly relevant for practitioners, who would need to invest in developing new systems.

2.5.2 Challenges in practice

There are many potential pitfalls when collecting data from internet sources which we discussed in Section 2.3. For instance, we underestimated the effort needed for data cleaning.

Our data obtained from social networks and YouTube contained many arbitrary spikes and changes in volume. We assume that these numbers vary because of potential click bait validation and synchronisation between servers.

User generated content has been praised for its availability at high frequency, i.e. hourly or even minutes (Tirunillai and Tellis, 2012). However, at a high sampling frequency, the collected values may become unreliable, which may also explain to some extent the weak forecasting in our results. While the fast data-stream allows for very granular sampling rate, increased volatility, multiple-seasonalities and intermittency are introduced.

One further complication in practice might be how timely the data becomes available. Most studies, including the one at hand, collect the data ex-post which makes it relatively easy to find matchings keywords. However, given a relatively new product, such signals might not be easy to identify, as search volume or reviews need to build up first. There is a lack of research as to when such signals appear strong enough and when they decline towards the end of the product life-cycle.

As discussed in Section 2.3.3, a further complication can be the selection of keywords. In our case, we used the video game title, which turned out to be highly correlated to sales. In practice, not all products or services will have such a distinct search keyword, and the signal can become distorted by unrelated search events to the product in question. Another issue is that the desired keyword may have too little search volume (Barreira et al., 2013). This limitation becomes more severe when looking at a disaggregate level. Cui et al. (2018) and Seebach et al. (2011) both suggest using hierarchical disaggregation methods for generating SKU level forecasts. However, we are unaware of any study that evaluates the forecasting performance of categorical and geographical disaggregation methods with internet platform data.

2.6 Conclusions

In this paper, we investigated whether search traffic and social network shares are helpful in improving demand forecasting. We first looked at the existing literature and identified limitations regarding their experimental design, both from a statistical and practical point of view. Although the majority of publications argued favourably as to the value of such data, our recommendation for researchers and practitioners is to take a more critical stance in using it for forecasting.

From a forecasting point of view, we did not find substantial differences regarding predictive power in different phases of the life-cycle. However, it is beyond the scope of this study to explore the usefulness of this information prior to launch. There is active research in this area with promising findings (e.g. Kim et al., 2015; Xiong and Bharadwaj, 2014; Kulkarni et al., 2012). It may still be very useful in different forecast settings, such as nowcasting or by providing insights into consumer behaviour. However, we underline the need for adequate benchmarking and thorough forecast evaluation. All benchmarks used in this study are well researched and understood forecasting models, which nowadays are trivial to deploy and automate in a practical setting. At least, these should be outperformed before the inclusion of additional explanatory variables would be warranted. Researchers and practitioners should also be aware of the data complexity, potential biases and dependency from the platform providers.

Naturally, our evaluation has limitations, but it supports aspects of our critical stance towards the literature. One could argue that our comparison is unfair since ARIMA or ETS could also be augmented with additional variables. Although this is a limitation of our design, specifying ARIMA or ETS with automated explanatory variable selection is challenging and neither approach lend to readily select variables with a lasso. We also looked exclusively at linear models and preferred modelling approaches that could be automated and scaled up, reflecting the needs of the practice. Although we did not find any evidence of non-linearity by exploring the datasets in our case studies, this will not be true for every application. For example, Cui et al. (2018) postulate that non-linear models are the most effective to include social media information. Our work leaves space for experimenting with more exotic linear or non-linear models.

A Detailed literature tables

The following Tables provide detailed insights on the literature surveyed for each of the forecasting applications used in Section 2. For each reviewed study we list the *target variable*, data *source*, identified *best model* and *benchmarks* used. Studies with distinct datasets have been split into multiple lines. The column *type* names the internet platform and the column *measure* indicates the nature of the data, e.g. volume or sentiment information. We also report the data *frequency*, number of *series*, *maximum lag* the user generated information variable is incorporated and *forecast horizon*. The last column reports the relative *maximum error reduction* achieved, compared to the best performing benchmark without any online information.

More specifically, Table A1 lists studies which forecast economic indicators including home sales or unemployment benefits claims. Table A2 includes all surveyed papers which look into forecasting financial markets, i.e. stock returns, index volatility etc. A list of studies concerned with predicting public health and environmental target variables like influenza outbreak forecast or environmental hazards is provided in Table A3. Table A4 then shows identified studies which forecast variables for services such as visitor arrivals and cinema spectators. Last but not least, Table A5 highlight research concerned with forecasting consumer goods. The first part in this table is on brand level forecasting and the second part for studies which consider product level. All forecasting model abbreviations used are explained in Table A6.

Table A1: Forecasting economic indicators

Paper	Target Variable	Type*	Measure§	Best model	Benchmark model(s)	Freq.†	num. series	max. lag	max. horizon	max. % err. red.
Choi and Varian (2009)[a,c]	Home sales USA	GTD	Pop.	**ARX**	AR	m	1	0	1	12% MAE
	Retail sales USA	GTD	Pop.	**sARX**	sAR	m	1	0	1	18% MAE
Vosen and Schmidt (2011)[c,d]	Private consumption USA	GTD	Pop.	**ARX**	AR, ARX	m	1	3	1	76% RMSE
Choi and Varian (2012)[a,c]	Motor vehicles and parts USA	GTD	Pop.	**sARX**	sAR	m	1	0	1	10.5% MAE
	Unemployment benefits claims USA	GTD	Pop.	AR	**ARX**	w	1	0	1	AR better
	Consumer confidence Index AUS	GTD	Pop.	**ARX**	AR	m	1	0	1	9.3% MAE
Barreira et al. (2013)[a]	Unemployment rate (var. countries)	GTD	Pop.	**ARX**	AR	m	4	2	36	13.4% RMSE
Scott and Varian (2015)[a,b,c,d]	Consumer sentiment USA	GTD	Pop.	**BSTS**	AR, LR	m	1	0	1	13.5% MAPE
	Gun sales USA	GTD	Pop.	**BSTS**	BSTS	m	1	0	1	55.9% MAE
Wu and Brynjolfsson (2015)[a,c,d]	Home sales USA	GTD	Pop.	**ARX**	ARX	q	1	1	1	2.9% MAE
	Home price index USA	GTD	Pop.	**ARX**	ARX	q	1	1	1	7.1% MAE
Brynjolfsson et al. (2016)[a,c]	Unemployment benefits claims USA	GTD	Pop.	**ARX**	AR	w	1	0	1	3.9% MAE
Li (2016)[a,c]	Unemployment benefits claims USA	GTD	Pop.	**DFM**	AR, DFM, RW	w	1	2	2	0.04% RMSE
Limnios and You (2016)[c]	Home prices USA	GTD	Pop.	DW	**ARX**, ARX, **DW**	m	1	6	1	DW better
	Price-rent ratio USA	GTD	Pop.	IAC	**ARX**, ARX, **IAC**	m	1	6	1	IAC better
Smith (2016)[a,c,d]	Unemployment rate GBR	GTD	Pop.	**AR-MIDAS**	RW w. drift, Survey	m/w	1	1	1	11% RMSE
D'Amuri and Marcucci (2017)[c,d]	Unemployment rate USA	GTD	Pop.	**ARX**	AR, ARX, SETAR, LSTAR, AAR	m	1	4	12	39% RMSE
Bulut (2017)[a,c,d]	Exchange rate movements	GTD	Pop.	RW	**LRX**, LRX	m	11	0	1	RW better
Elshendy et al. (2018)[c]	Crude oil price	GTD; TWR; GDELT; WIA	Pop.; Vol.; Val.[2]	**ARIMAX**	ARIMA	d	1	3	1	88.9% MAPE
Yu et al. (2018)[c]	Oil consumption	GTD	Pop.	**SVR**	LR, SVR,ELM,NN	m	1	6	1	1.4% MAPE

Boldface = Model containing online information; [a] Nowcasting model; [b] Only insample evaluation; [c] Rolling origin evaluation; [d] Statistical model testing; * GTD = Google Trends, TWR = Twitter, WIA = Wikipedia; § Rat. = Rating, Pop. = Popularity, Vol. = Volume, Val. = Valence ([1] = Lexicon, [2] = Machine learning, [3] = n-grams, [4] = Self declared); † d = daily, w = weekly, m = monthly.

Table A2: Forecasting financial markets

Paper	Target Variable	Type*	Measure§	Best model	Benchmark model(s)	Freq.†	num. series	max. lag	max. horizon	max. % err. red.
Bollen et al. (2011)[c,d]	DJIA Index	TWR	Val.[1]	**NN**	NN	d	1	3	1	7.7% MAPE
Rao and Srivastava (2013)[c]	US Oil Funds Index	GTD; TWR	Pop.; Val.[1]	**ARX**	AR	w	4	4	1	4.1% MAPE
	DJIA Index	GTD; TWR	Pop.; Val.[1]	**ARX**	AR	w	4	4	1	26.8% MAPE
	NASDAQ-100 Index	GTD; TWR	Pop.; Val.[1]	**ARX**	AR	w	4	4	1	2.6% MAPE
	Gold Price	GTD; TWR	Pop.; Val.[1]	**ARX**	AR	w	2	4	1	1.99% MAPE
	EURO	GTD; TWR	Pop.; Val.[1]	**ARX**	AR	w	2	4	1	39.8% MAPE
Hamid and Heiden (2015)[c,d]	Index volatility	GTD	Pop.	**ESC**	AR, HAR, ARFIMA	w	1	1	1	6% MSE
Dimpfl and Jank (2016)[c,d]	Index volatility	GTD	Pop.	**VHAR**, **VAR**	AR, HAR	d	1	1	14	3.5% MSE
Bijl et al. (2016)[b,c,d]	Stock returns	GTD	Pop.	**LR**	LR	d	431	5	1	32% R^2
Ho et al. (2017)[c,d]	Stock returns	FOM	Val.[4]	**SUR**	SUR	d	45	1	1	36.4% MAE
Perlin et al. (2017)[c]	Index returns	GTD	Pop.	**VAR**	ARMA-GARCH, BH	m	4	5	1	1014% Return

Boldface = Model containing online information; [a] Nowcasting model; [b] Only insample evaluation; [c] Rolling origin evaluation; [d] Statistical model testing; * BAU = Baidu, FOM = Forum, GTD = Google Trends, TWR = Twitter; § Pop. = Popularity, Vol. = Volume, Val. = Valence ([1] = Lexicon, [2] = Machine learning, [3] = n-grams, [4] = Self declared); † d = daily, w = weekly, m = monthly, q = quarterly.

Table A3: Forecasting public health and environment

Paper	Target Variable	Type*	Measure§	Best model	Benchmark model(s)	Freq.†	num. series	max. lag	max. horizon	max. % err. red.
Ginsberg et al. (2009)[a,c]	Influenza outbreak USA	GTD	Pop.	**LR**	-	w	1	0	1	-
Lampos and Cristianini (2012)[a,c]	Influenza outbreak GBR	TWR	Val.[3]	**LR**	-	w	3	0	1	-
	Daily Rainfall GBR	TWR	Val.[3]	**LR**	-	d	5	0	1	-
Won et al. (2013)	Suicide events in Korea	BLG	Val.[1]	**ARX**	-	3-day bins	1	1	121	-
Araz et al. (2014)[c]	Influenza outbreak Nebraska	GTD	Pop.	**LR**	sARIMA, HW	w	1	1	1	72.8 % RMSE
Lazer et al. (2014)[a,c]	Influenza outbreak USA	GTD	Pop.	**ARX**	AR	w	10	3	1	25.4% MAE
Preis and Moat (2014)[a,c]	Influenza outbreak USA	GTD	Pop.	**ARX**	AR	w	1	0	1	21.3% MAE
Santillana et al. (2015)[c]	Influenza outbreak USA	GTD; TWR	Pop.; Vol.	**ABT**	**LR**, **SVM**, AR	w	1	4	2	38.7% MAPE
Brynjolfsson et al. (2016)[a,c]	Influenza outbreak USA	GTD	Pop.	**LR**	AR	w	1	3	1	9.9% MAE
Chen et al. (2017)[c]	Smog health hazard China	WEO	Val.[1]	**NN**	NN, SVM, RF	d	8	0	1	26.0% RMSE
Xu et al. (2017)[a,c]	Influenza outbreak Hong Kong	GTD	Pop.	**BMA**	**ARIMAX**, **GLM**, **NN**	w	1	0	2	-

Boldface = Model containing online information; [a] Nowcasting model; [b] Only insample evaluation; [c] Rolling origin evaluation; [d] Statistical model testing; * BLG = Blog, GTD = Google Trends, TWR = Twitter, WEO = Weibo; § Pop. = Popularity, Vol. = Volume, Val. = Valence ([1] = Lexicon, [2] = Machine learning, [3] = n-grams, [4] = Self declared); † d = daily, w = weekly.

Table A4: Forecasting services

Paper	Target Variable	Type*	Measure§	Best model	Benchmark model(s)	Freq.†	num. series	max. lag	max. horizon	max. % err. red.
Choi and Varian (2009)[a,b]	Visitor arrivals to HK	GTD	Pop.	**ARX**	-	m	9	0	-	-
Pan et al. (2012)[c]	Hotel room demand	GTD	Pop.	**ARX**	AR, ARIMA, RW	w	1	1	1	27% MAPE
Hand and Judge (2012)	Cinema admissions	GTD	Pop.	**sARX**	**sARX**	m	1	0	-	-
Bangwayo-Skeete and Skeete (2015)[c,d]	Tourist arrivals in the Caribbean	GTD	Pop.	**AR-MIDAS**	AR, sARIMA	m/w	5	-	12	-
Bughin (2015)[a,c]	Telecom sales in BEL	GTD; TWR	Pop.; Val.	**ARX**	AR	m	9	2	1	21% RMSE
Kim and Shin (2016)[c]	Air passengers	NAR	Pop.	**LR**	-	m	1	8	8	-
Önder and Gunter (2016)[c,d]	Tourist arrivals to Vienna	GTD	Pop.	**ADL**	sAR, HW, sRW	m	7	12	12	56.4% MAE
Pan and Yang (2016)[c,d]	Hotel occupancy	GTD	Pop.	**ARMAX**	ARMA, MSDR	w	1	2	2	4.73% MAPE
Peng et al. (2016)[c]	Tourist volume Jiuzhai Valley	BAU	Vol.	**ARX**	**ARX**	d	1	3	1	-
Rivera (2016)	Hotel registrations Puerto Rico	GTD	Pop.	**DLM**, HW	sRW, HW, sARIMA	m	1	1	12	9.18% MAPE
Huang et al. (2017)[a,c]	Tourist volume Forbidden City	BAU	Vol.	**ARX**	ARIMA, ARX	d	1	2	1	14.5% RMSE
Li et al. (2017)[c]	Tourist arrivals to Beijing	BAU	Vol.	**ARMAX**	ARMA, **ARMAX**	w	1	5	4	37% MAPE
Önder (2017)[c,d]	Tourist arrivals to Countries	GTD	Pop.	**ADL**	AR, HW, RW	m	6	12	12	3% MAE
	Tourist arrivals to Cities	GTD	Pop.	**ADL**	AR, HW, RW	m	10	12	12	36.7% MAE
Padhi and Pati (2017)[c]	Tourist arrivals to Kerela	GTD	Pop.	**ARIMAX**	ARIMA, **ARX**, **VARX**	m	1	4	4	53.8% MAPE
Park et al. (2017)[c,d]	Tourist arrivals to South Korea	GTD	Pop.	**sARIMAX**	sARIMAX, HW	m	1	6	1	16.7% MAE
Zeynalov (2017)[c]	Tourist arrivals to Prague	GTD	Pop.	**AR-MIDAS**	ARIMA, **ARIMAX**	m/w	1	1	1	19.7% MAE
	Overnight stays in Prague	GTD	Pop.	**AR-MIDAS**	ARIMA, **ARIMAX**	m/w	1	1	1	8.5% MAE

Boldface = Model containing online information; [a] Nowcasting model; [b] Only insample evaluation; [c] Rolling origin evaluation;
[d] Statistical model testing; * BAU = Baidu, GTD = Google Trends, NAR = Naver, TWR = Twitter;
§ Pop. = Popularity, Vol. = Volume, Val. = Valence; † d = daily, w = weekly, m = monthly, q = quarterly.

Table A5: Forecasting consumer goods

Paper	Target Variable	Type*	Measure§	Best model	Benchmark model(s)	Freq.†	num. series	max. lag	max. horizon	max. % err. red.
Brand or company level										
Choi and Varian (2009)[a,c]	Car sales USA	GTD	Pop.	**sARX**	sAR	m	27	0	1	12.5% MAE
Seebach et al. (2011)[c]	Car sales DEU	GTD	Pop.	**ARX**	AR, RW	m	2	6	3	50.3% MAE
Barreira et al. (2013)[a,c,d]	Car sales var. countries	GTD	Pop.	**ARX**	AR	m	4	0	36	8.2% RMSE
Carrière-Swallow and Labbé (2013)[a,c,d]	Car sales CHL	GTD	Pop.	**ARX**	AR, ARMA	m	9	18d	1	-
Skodda and Benthaus (2015)[c]	Car sales DEU	GTD; TWR	Pop.; Val.[1]	**LR**	**LR**	m	2	6	1	-
Fantazzini and Toktamysova (2015)[c,d]	Car sales DEU	GTD	Pop.	**VECM**, **VAR**, **BVAR**	RW w. drift, AR, VARX, VECMX, BVARX, LSTAR, ESTAR, AAR	m	22	12	24	66% MSE
Cui et al. (2018)[c,d]	Online apparel sales	FBK	Vol.; Val.[2]	**RF**	RF, LR, SVM, GB, Company Forecast	d	1	7	7	20.5% MAPE
Jun et al. (2017)	Netbook sales	GTD; NAR	Pop.; Pop.	HW	**Analogy**, sRW, MA	q	1	0	1	HW better
Geva et al. (2017)[c,d]	Car sales USA	GTD; FOM	Pop.; Vol.; Val.[1]	**NN**	sARX, NN	m	23	2	1	5.6% MAPE
Product level										
Boone et al. (2015)[a,b]	Speciality-food SKUs sales	GTD	Pop.	**ARX**	ARX	w	2	0	-	5.3% RMSE
Schneider and Gupta (2016)[c]	Tablet sales	ORW	Rat.; Val.[2]	**LR**	**SVM**, LR	w	231	0	1	77.2% MAPE
See-To and Ngai (2016)[a,c]	Fashion sales	ORW	Val.[1]	**ARX**	MA	$\frac{1}{2}$d	2527	0	1	44.2% RMSE
Geva et al. (2017)[c,d]	Car sales USA	GTD; FOM	Pop.; Vol.; Val.[1]	**sARX**	sARX	m	78	2	1	1.9% MAPE
Jun et al. (2017)	Nintendo Wii sales	GTD; NAR	Pop.; Pop.	HW	**Analogy**, sRW, MA	q	3	0	1	HW better
Lau et al. (2018)[c]	E-commerce platform USA	GTD; ORW	Val.[1]	**PELM**	**LR**, **SVM**, **ELM**	w	11428	1	1	-
	E-commerce platform CHN	BAU; ORW	Val.[1]	**PELM**	**LR**, **SVM**, **ELM**	w	8115	1	1	-
Boone et al. (2018)[c]	Speciality-food SKUs sales	GTD	Pop.	**ARX**	ARX	w	5	1	1	7.66%

Boldface = Model containing online information; [a] Nowcasting model; [b] Only insample evaluation; [c] Rolling origin evaluation; [d] Statistical model testing; * BAU = Baidu, FBK = Facebook, FOM = Forum, GTD = Google Trends, NAR = Naver, ORW = Online Reviews, TWR = Twitter;
§ Rat. = Rating, Pop. = Popularity, Vol. = Volume, Val. = Valence ([1] = Lexicon, [2] = Machine learning, [3] = n-grams, [4] = Self declared);
† d = daily, w = weekly, m = monthly, q = quarterly.

Table A6: Abbreviations of forecasting models

Model	Abbreviation
AAR	Additive AR
ABT	AdaBoost
ADL	Autoregressive Distributed Lag
AR	Autoregressive
ARFIMA	AR Fractional Integrated Moving Average
ARIMA	AR Integrated Moving Average
BH	Buy and Hold strategy
BMA	Bayesian Model Averaging
BSTS	Bayesian Structural Time Series
BVAR	Bayesian VAR
DLM	Dynamic Linear Model
DW	Dieci and Westerhoff Model
ELM	Extreme Learning Machine
ESC	Empirical Similarity Concept
IAC	Housing a la Iacoviello
GARCH	Generalised AR Conditional Heteroskedasticity
GB	Gradient Boosting
GLM	Generalised Linear Model
HAR	Heterogeneous AR
HW	Holt Winter
LR	Linear Regression
LSTAR	Logistic Smooth Transition AR
MA	Moving Average
MSDR	Markov Switching Dynamic Regression
NN	Neural Network
PELM	Parallel co-evolutionary ELM
RF	Random Forest
RW	Random Walk
SETAR	Self-Exciting Threshold AR
SUR	Seemingly Unrelated Regression
SVM	Support Vector Machine
VAR	Vector AR
VECM	Vector Error Correction Model
VHAR	Vector Heterogeneous AR

Chapter 3

Estimating the market potential with pre-release buzz

With increased competition and shorter product life-cycles, forecasting the life-cycle sales of new products prior to launch is becoming increasingly important to marketers and demand planners. Our study contributes to the literature on new product adoption using analogies by augmenting information from previous generations with pre-release search traffic. In contrast to existing research, which relies on pre-release buzz information only for the launch phase, we consider life-cycle sales. First, we propose a model of pre-release buzz and market potential, establishing the connection between the two. Then, we validate this relationship with an empirical experiment on sequential video game sales. Our findings support that pre-release buzz contains predictive information up to 17 weeks prior to release and can increase life-cycle sales forecast accuracy up to 20%. The explanatory power of pre-release buzz varies across product generations. This evolution opens up marketing opportunities and highlights why it is important to manage pre-release buzz.

3.1 Introduction

With shorter product life-cycles and increased competition, generating accurate pre-launch forecasts is vital for companies to plan marketing expenditures and production. Today, a large number of new product launches represent incremental product innovation, rather than radical new products (Markham and Lee, 2013). In this case a common forecast strategy is to use analogies of previous or similar products (e.g. Hu et al., 2019; Lenk and Rao, 1990). While the adoption shape is relatively straightforward to obtain, estimating the market potential of the new product is challenging (Trusov et al., 2013). In practice, this is typically done by using expert judgment (e.g. Kim et al., 2013; Bass et al., 2001; Mahajan et al., 1986). However, there is substantial evidence that experts are biased when forecasting the success of a new product (Belvedere and Goodwin, 2017; Markovitch et al., 2014; Tyebjee, 1987).

Alternative approaches to predict the market potential such as consumer surveys (e.g. Moon et al., 2016; Chintagunta and Lee, 2012; Eliashberg et al., 2000) or conjoint analysis (e.g. Orbach and Fruchter, 2011; Lee et al., 2006) can be costly to obtain and only reflect consumer preferences at a specific point in time. Meeran et al. (2017) showed, that throughout the pre-launch phase, consumer preference for short life-cycle products change significantly and can invalidate results from previous market research surveys. A more timely option compared to surveys is to measure online pre-release buzz (PRB).

PRB reflects the aggregate anticipation of consumers towards a new product (Houston et al., 2018). Various studies report improved pre-launch forecast accuracy when incorporating information from sources such as search engines (e.g. Tian et al., 2014; Kulkarni et al., 2012), blogs (e.g. Kim et al., 2015; Dhar and Chang, 2009), microblogs (e.g. Gelper et al., 2015; Asur and Huberman, 2010) and forums (e.g. Wang et al., 2010; Liu, 2006). However, the majority have only investigated the forecasting potential for the initial weeks of sales, which from an operational point of view might not suffice, given that traditional time series models require a reasonable length of sales history. Moreover, the time-to-order often requires longer forecast horizons, and many fashion market products with a short life-cycle have very few, or no, batch re-orders (Mostard et al., 2011).

With the aim of enhancing the information base of analogy based forecasts, in this research, we provide evidence on whether search traffic information from Google Trends, as proxy for PRB, is useful in improving the estimation of the market potential prior to product launch. Search traffic is particularly relevant when viewed as a proxy for consumer interest in a product (Houston et al., 2018; Du and Kamakura, 2012; Stephen and Galak, 2012), but also partially capturing the marketing expenditures (Srinivasan et al., 2016; Hu et al., 2014). The market potential has been forecasted from post-launch advertising activities (e.g. Toubia et al., 2014; van der Lans et al., 2010) or online reviews (Dellarocas et al., 2007) demonstrating their impact on sales, but the literature suggest that pre-launch advertising is more effective than post-release advertising (Burmester et al., 2015). Although, search traffic and advertising expenditures are correlated, the user-generated content has been argued to have a more substantial effect on search activities than advertising (Kim and Hanssens, 2017) and more predictive power (Xiong and Bharadwaj, 2014).

In this paper, we propose an analogy based predictive life-cycle PRB model, to support operational decision making, with which we (i) test the relation between pre-release buzz and the market potential of a new product generation; (ii) analyse how the explanatory power of PRB changes as the product generations mature and (iii) investigate the leading properties of pre-release buzz for total end-of-life sales. We argue that any model that improve predictive power has adequately approximated key elements of the underlying market structure.

Section 3.2 presents the relevant literature on pre-launch forecasting with pre-release buzz. Section 3.3 elaborates how search traffic can be modelled in order to estimate the market potential. Section 3.4 demonstrates empirically the value of the proposed models by predicting physical video game sales prior to launch, using a variety of adoption curves. Our findings suggest that pre-release buzz is a useful source, not only in predicting the initial week of sales but also in explaining better the long-term market potential compared to models that only contain past market potential information. Our empirical case of forecasting pre-launch sales of video games supports the conjecture that pre-release buzz contains predictive information 17 weeks prior to release and can increase life-cycle sales forecast accuracy up to 20%. Our model achieves higher explanatory power for the total market potential than it explains first week sales, underlying the importance of pre-release buzz for the market potential, with implications for marketeers. However, the explanatory power of pre-release buzz varies across product generations. This provides new insights on how the predictive value of pre-release buzz evolves, and this is discussed in Section 3.5 together with practical implications on why it is important to manage PRB and concluding remarks.

3.2 Forecasting with pre-release buzz

The literature investigating pre-launch buzz information is summarised in Table 3.1. Besides the PRB source that models the target variable, the Table also reports the dataset size, model specification, benchmarks used as well as the maximum forecast horizon and investigated lead time. The predominant applications are movie box office sales forecasting and other types of hedonistic products, such as video games and music album sales. Only Mülbacher et al. (2011) has investigated a product outside the entertainment industry (alpine skis), but the sample size did not allow for a thorough forecast evaluation. Several studies do not focus on the predictive side of these inputs and report only in-sample statistics (e.g. Gopinath et al., 2013; Dhar and Chang, 2009; Liu, 2006). Other studies were more predictively oriented but used a weak forecast evaluation design, i.e. missing benchmark models with no clarity on the conditionality of their experiments with regards to PRB (e.g. Tian et al., 2014; Onishi and Manchanda, 2012).

The studies that have investigated forecasting applications in more detail, have reported positive findings. Kim et al. (2015) showed a reduction of over 40% MAPE when using information from online blogs. Xiong and Bharadwaj (2014) forecasted video game sales with search traffic information, blog and forum messages, which led to improvements of more than 35% MAPE. Other studies used more specialised sources of information that would not be available for all applications but proved useful in the specific pre-launch context. For example, Foutz and Jank (2010) used a prediction market to forecast box office sales, which led to a decrease in MAPE of more than 70%. Hann et al. (2011) use information from peer-to-peer network shares as an indicator for predicting album sales, improving MAPE by over 30%. This is in contrast to studies that have investigated short-term forecasts using online information for established products. In this case, the findings are more mixed (see

Table 3.1: Forecasting with pre-launch buzz

Study	Target variable	# series	Model*	Model relation	Buzz source*	Buzz measure	Benchmark◇	Max. lead†	Max. Horizon†
Liu (2006)	Box office sales	40	LR	log – log	FOM	Vol.; Val.	-	1 w	1 w
Dhar and Chang (2009)	Music album sales rank	108	LR	log – log	BLG	Vol.	-	3 w	3 w
Foutz and Jank (2010)	Box office sales	262	LR	linear	VSX	-	PChs, ASpd	40 w	1 w
Asur and Huberman (2010)	Box office sales	24	LR	linear	TWR	Vol.; Val.	HSX	1 d	1 d
Wang et al. (2010)	Box office sales	51	BC	linear	FOM	Vol.	-	1 w	2 w
Hann et al. (2011)	Music album sales	172	FR	linear	P2P	-	PChs	4 w	1 w
Mülbacher et al. (2011)	Ski sales	10	LoR	linear	FOM	Vol.; Val.	-	1 d	1 y
Kulkarni et al. (2012)	Box office sales	61	LR	log –linear	GTD	Vol.	PChs, Aspd	4 w	1 w
Onishi and Manchanda (2012)	Box office sales	1729	SEM	log – log	BLG	Vol.; Val.	-	1 d	1 d
	Cell phone service	90	SEM	log – log	BLG	Vol.; Val.	-	1 m	1 d
Gopinath et al. (2013)	Box office sales	75	LR	log – log	BLG	Vol.; Val.	-	1 d	1 m
Tian et al. (2014)	Box office sales	92	GLM, DT, NN	linear	BAU	Vol.	-	1 d	1 d
Xiong and Bharadwaj (2014)	Video game sales	681	FR	linear	BLG; FOM; GTD	Vol.	PChs, ASpd, Pors	151 d	3 w
Craig et al. (2015)	Box office sales	62	LR	log –linear	FOM	Vol.	PChs	3 w	1 w
Gelper et al. (2015)	Box office sales	106	LR	log – log	TWR	Vol.; Val.	PChs, ASpd	30 d	1 d
Kim et al. (2015)	Box office sales	212	LR, SVR, GPR, k-NN	log –linear	BLG	Vol.	PChs	1 w	1 w
Ding et al. (2017)	Box office sales	64	LR	log – log	FBK	Vol.	PChs	1 w	1 w
Divakaran et al. (2017)	Box office sales	373	PR	log – log	BLG	Vol.	-	1 w	1 w
Kim and Hanssens (2017)	Box office sales	137	LR	log – log	GTD; BLG	Vol.; Vol.	PChs, ASpd	3 w	1 w
This study	Video game sales	255	AY	log – log, linear	GTD	Vol.	RW, AR(1)	26 w	52 w

⋆ AY = Analogy, BC = Bass curve, DT = Decision Tree, FR = Functional Regression,GLM = Generalised Linear Model, GPR = Gaussian Process Regression, K-NN = K-Nearest Neigbhour Regression, LgR = Logistic Regression, LR = Linear Regression, NN = Neural Networks, PR = Partial Regression, SVR = Support Vector Regression;
* BAU = Baidu, BLG = Blog, FBK = Facebook, FOM = Forum, GTD = Google Trends, P2P = Peer-to-Peer Network, TWR = Twitter, VSX = Virtual Stock Exchange;
◇ ASpd = Ad Spend, HSX = Historic Stock Exchange Prices, PChs = Product Characterstics, POrd = Pre-Orders, - = Not tested against benchmark;
† d = daily, w = weekly, m = monthly, y = yearly.

Schaer et al., 2019a, for an overview), since user-generated content apparently becomes less effective as a driver of sales once the product is on the market (Marchand et al., 2017; Onishi and Manchanda, 2012; Liu, 2006).

User-generated content has been incorporated into forecasting models in various ways as shown in Table 3.1. For example, search traffic has been modelled by including it directly to a regression model (e.g. Kim and Hanssens, 2017), by parametrising its shape with principal components (Xiong and Bharadwaj, 2014; Hann et al., 2011; Foutz and Jank, 2010) or hazard models (Kulkarni et al., 2012). Some studies also compared the predictive value of different sources of information but were overall inconclusive in their findings. For example, Xiong and Bharadwaj (2014) reported higher forecast accuracy using inputs from blogs than search, whereas Kim and Hanssens (2017) achieved better forecast accuracy when using only search. Apart from using shape and volume based data, relying on the valence of posts is another frequently applied approach (e.g. Gelper et al., 2015; Asur and Huberman, 2010). In forecasting box office sales Onishi and Manchanda (2012) found sentiment to contain more information than volume, while Liu (2006) reached the opposite conclusion.

With the exception of the work by Mülbacher et al. (2011), these other studies have focused on the opening week sales looking no more than one month ahead. Kulkarni et al. (2012) argued that for movies a week-to-week forecast is not as important as the subsequent weeks tend to follow a similar sales pattern that can be determined from the initial week sales (for example using the approach by Sawhney and Eliashberg, 1996). From their perspective, the most important objective was to estimate the magnitude of the initial week sales, but it remains unclear whether the established relationship between pre-release buzz also holds over a longer period of time. One of the key challenges in life-cycle modelling is to capture the changing pattern of adoption period-by-period. Although this is conditional on the initial sales, this is far from the only relevant factor. Our first hypothesis, therefore, addresses the question of whether pre-release buzz can also predict sales over longer-term horizons;

Hypothesis 1 *The market potential is conditional on pre-release buzz.*

Studies focusing on the interaction between pre-release buzz and sales employ structural equation models (Divakaran et al., 2017; Onishi and Manchanda, 2012). The literature that has focused on generating forecasts, all predict for a specific week, i.e. a single data point, most often through linear regression (e.g. Ding et al., 2017; Gelper et al., 2015; Asur and Huberman, 2010; Liu, 2006). Other studies used non-linear machine learning methods (Kim et al., 2015; Tian et al., 2014) or base their estimation on functional regression (Xiong and Bharadwaj, 2014; Hann et al., 2011; Foutz and Jank, 2010). The latter approach is reported to perform well and for example achieved better results than regressing Bass curve growth model parameters against the initial week of sales (Hann et al., 2011). Wang et al. (2010) used pre-release buzz to improve the estimation of the growth parameters of a diffusion curve, but not the market potential parameter, which arguably is closely connected to the pre-release interest that consumers show for a product. In all surveyed studies the connection of pre-releease buzz and the market potential aspect of a product has been neglected and remains an open question.

Several studies have suggested that buzz appearing closer to the launch date is the most useful (Xiong and Bharadwaj, 2014; Chintagunta and Lee, 2012; Hann et al., 2011; Wang et al., 2010). Kim and Hanssens (2017) pointed out that most of the literature concerned with a box office revenue only forecast very near the day of release, which essentially leads into a now-casting approach possibly limiting its managerial relevance. As a justification for the short-term focus, Chintagunta and Lee (2012) stated that a week before the release is sufficient to adjust the proposed movie marketing campaign and negotiate the rental fees. However, there are few studies that have explored predictive ability with longer lead times, despite the increased relevance for planning purposes: this is of particular interest when the forecast is aimed to provide information for production and supply chain decisions.

An issue related to the forecast horizon is the lead in the pre-release buzz data. Kulkarni et al. (2012) reported that forecasts of box office sales four weeks prior to launch are more accurate than forecasts in the week before launch when using online search. Gelper et al. (2015) developed forecasts with Twitter data up to 30 days prior to launch, and Kim and Hanssens (2017) improved forecast accuracy three weeks prior to release from blog posts and search volume information. The longest lead time in our literature survey considered was 40 weeks. However, in this case, Foutz and Jank (2010) relied on prediction market information, which can be generated earlier than online blog posts. For music album sales, Hann et al. (2011) considered up to four weeks prior launch, but their models underperformed compared to the benchmark based on album characteristics on median percentage errors 4 and 3 weeks before launch. For physical video games sales, Xiong and Bharadwaj (2014) generated forecasts 151 days before product launch. Even at this early stage, their forecasts outperformed their benchmarks that were based on product characteristics and ad spending. Closer to launch, the forecast errors of the pre-release buzz model gradually reduced by 30% MAPE and also outperformed a benchmark which included pre-orders 90 days before release. Their finding suggested that online user-generated information could be useful for predictions substantially in advance and contain more explanatory information than company internal data.

A few studies have investigated the difference between sequels and non-sequels. They reported pre-release buzz of sequels to observe a steeper trend (Xiong and Bharadwaj, 2014; Foutz and Jank, 2010) and a higher initial level of buzz (Craig et al., 2015; Xiong and Bharadwaj, 2014). Liu (2006), on the other hand, has reported that the investigated Word-of-Mouth of non-sequels accounts for 6% higher box office revenue, but this conclusion was based on only 3 sequels were available in that dataset. All those studies incorporated sequels as binary dummy variables (also noted in Marchand, 2016) and did not investigate the predictive ability of pre-release buzz for different product generations. Therefore, we formulate a second hypothesis which considers whether pre-release buzz differentiates between product generations;

Hypothesis 2 *The predictive value of pre-release buzz is different between product generations*

A common approach to forecasting new product demand is to rely on information of previous generations or products with similar characteristics (e.g., Hu et al., 2019; Baardman et al., 2017). In the reviewed literature, only Hann et al. (2011) with movies and Tian et al. (2014) analysing music albums took such information into account. In this research, we propose a different modelling approach that relies on pre-release buzz and information from analogous products. Therefore, we examine whether in predicting future adoption, it is beneficial to supplement past adoption profiles with pre-release buzz, which effectively reflects multiple underlying factors. This approach has the benefit of being very economical in terms of the required fitting sample. Specifically, we argue that the realised market potential is conditional on the pre-release buzz.

3.3 Market potential estimation with pre-release buzz

We regard market potential as the overall life-cycle sales, which is the cumulative adoption. Within the Bass modelling approach this corresponds to parameter m (Bass, 1969). The cumulative adoption (A_t) of a product can then be described as

$$A_t = mF_t\left(\cdot\right), \tag{3.1}$$

where m is the market potential and $F\left(\cdot\right)$ can be any function defining the cumulative shape of adoption across time. We discuss alternative life-cycle models in Section 3.4.2, while in this Section we formulate the connection between overall product sales and pre-launch of generations of products.

In this research, we focus on sequential product launches, as these offer the most natural setting for forecasting by analogies. A large number of companies spend a significant amount of their research budget on incremental product innovation, rather than radical new products (Markham and Lee, 2013), or in the case of movies and video games, follow a brand extension strategy where an existing brand name is attached to the sequel (e.g., Hennig-Thurau et al., 2009; Sood and Drèze, 2006). Irrespective of the specific context there is a strong link between the previous product generation and the new one in their sharing similar product features. Without this distinctive connection, one would have to gather additional information about related products and markets, to identify useful analogies (Goodwin et al., 2014).

Pre-release buzz (PRB) has been encoded in various ways by the literature. In this research we will focus on the volume of search traffic. Typically search traffic is represented as a relative measure and reflects the interest towards a product (Houston et al., 2018; Du and Kamakura, 2012; Stephen and Galak, 2012). We postulate that:

Hypothesis 3 *Pre-release buzz has a multiplicative (proportional) relationship with market potential.*

In the surveyed literature both multiplicative and additive model specification for pre-release buzz exist. However, the modelling motivation seems to be mainly data driven and has not been discussed in depth. While the majority of the research has used log-log models, and therefore a multiplicative connection, Craig et al. (2015) and Kim et al. (2015) only used logarithmic sales, resulting in a log-linear models, whereas Tian et al. (2014), Xiong and Bharadwaj (2014), Hann et al. (2011); Asur and Huberman (2010) and Foutz and Jank (2010) used fully additive models. In this research, we explore these alternative model specifications in more depth.

Table 3.2: Multiplicative market potential estimation with pre-release buzz (PRB)

Model	Equation	Estimation point	Model restriction
M.1	$\hat{m}_j = c_0 \cdot \Delta\mathrm{PRB}_j^{c_1} \cdot m_{j-1}^{c_2} \cdot \varepsilon_j$	$j \geq 5$	Full
M.2	$\hat{m}_j = c_0 \cdot \Delta\mathrm{PRB}_j^{c_1} \cdot \varepsilon_j$	$j \geq 4$	Restricted forms
M.3	$\hat{m}_j = c_0 \cdot \Delta\mathrm{PRB}_j^{c_1} \cdot m_{j-1} \cdot \varepsilon_j$	$j \geq 4$	
M.4	$\hat{m}_j = \Delta\mathrm{PRB}_j^{c_1} \cdot m_{j-1} \cdot \varepsilon_j$	$j \geq 3$	
M.5	$\hat{m}_j = \Delta\mathrm{PRB}_j \cdot m_{j-1} \cdot \varepsilon_j$	$j \geq 2$	
M.6	$\hat{m}_j = \Delta\mathrm{PRB}_j^{\frac{1}{2}} \cdot m_{j-1} \cdot \varepsilon_j$	$j \geq 2$	

Table 3.2 introduces the multiplicative models for the estimation of the market potential of the jth generation that in the full form is an autoregressive indicator model with differenced (Δ)PRB between the two consecutive generations (model 1) as an explanatory indicator with the error assumed to be a log-normal $\varepsilon_j \sim \log\mathcal{N}\left(0, \sigma_j^2\right)$. In the multiplicative form the model reflects our hypothesis of proportional relationship (*H3*). Moreover, the autoregressive process takes into account the carry-over effect between product generations (Situmeang et al., 2014). However, as indicated in the second last column, the full model requires at least 5 product generations, due to the number of parameters that need to be estimated. For this reason, we introduce alternative restricted versions. Model 2 is without the autoregressive term, Model 3 incorporates past market potential with a fixed coefficient of one. Model 4 and 5 are further restricted by dropping the intercept and coefficient es-

timation of PRB, respectively. In these cases, the latter only requires two observations for its estimation. Motivated by the findings of Kim et al. (2015), who find model combination to work well, we introduce Model 6 which has its PRB coefficient set to 0.5, in contrast to Model 5. That is a combination of 50% PRB and 50% random walk, i.e. the last observation m_{j-1}. Although, there are various ways to determine optimal combination weights, equal weights have been found to work well and often outperform more complex weighting schemes (Claeskens et al., 2016; Genre et al., 2013), in particular when the estimation sample is very limited as is the case here, when forecasting generation $j+1$. Furthermore, reducing the PRB coefficient towards zero underlies the principal idea of shrinkage (Hastie et al., 2015; Green and Armstrong, 2015), where we take a more conservative approach dampening the effect of PRB over the simpler random walk model.

Table 3.3 provides the corresponding additive models with errors $\eta_j \sim \mathcal{N}(0, \sigma_j^2)$. We consider both so as to be able to test *H3*, investigating the nature of the PRB relation with market potential.

Table 3.3: Additive market potential estimation with pre-release buzz (PRB)

Model	Equation	Estimation point	Model restriction
A.1	$\hat{m}_j = \alpha_0 + \alpha_1 \Delta\text{PRB}_j + \alpha_2 m_{j-1} + \eta_j$	$j \geq 5$	Full
A.2	$\hat{m}_j = \alpha_0 + \alpha_1 \Delta\text{PRB}_j + \eta_j$	$j \geq 4$	Restricted forms
A.3	$\hat{m}_j = \alpha_0 + \alpha_1 \Delta\text{PRB}_j + m_{j-1} + \eta_j$	$j \geq 4$	
A.4	$\hat{m}_j = \alpha_1 \Delta\text{PRB}_j + m_{j-1} + \eta_j$	$j \geq 3$	
A.5	$\hat{m}_j = \Delta\text{PRB}_j + m_{j-1} + \eta_j$	$j \geq 2$	
A.6	$\hat{m}_j = \frac{1}{2}\Delta\text{PRB}_j + m_{j-1} + \eta_j$	$j \geq 2$	

Our approach is limited in that we cannot forecast the initial generation. Furthermore, more complex market potential models require the estimation of additional coefficients, and therefore rely on additional past generations data.

3.4 Pre-release video game adoption modelling

For our empirical evaluation, we focus on the adoption of video games. The release of new games within a franchise is common amongst publishers and account for 80% of the top 20 games sold in 2017 (ESA, 2018). Furthermore, the gaming community is actively using internet platforms, such as blogs or social media, to interact with upcoming video game releases. Publishers devote a substantial amount of advertising budget into pre-release advertising (Marchand and Hennig-Thurau, 2013) making them an ideal subject for studying pre-release buzz. We use physical video games sales data from VGChartz (http://www.vgchartz.com), a data source also used in other pre-release research studies (e.g. Marchand et al., 2017; Xiong and Bharadwaj, 2014). The dataset contains weekly aggregate sales, across platforms, of 255 games that are part of 57 popular game franchises, including titles such as Call of Duty and EA FIFA Football series. The number of instalments within a franchise range from two to eleven, and on average each franchise has 3.4 game titles.

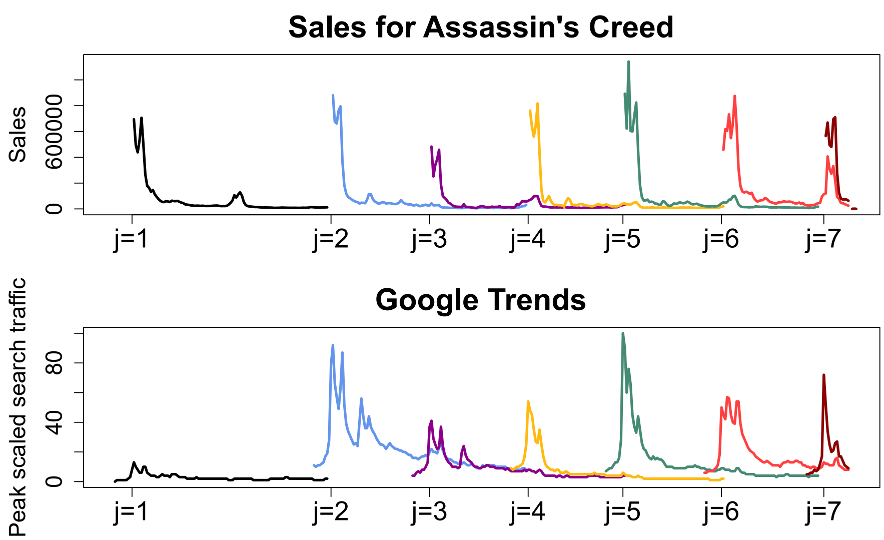


Figure 3.1: Generations of the Assassin's Creed compared to Google Trends

In this research, we use Google Trends search traffic as a proxy for PRB. Figure 3.1 illustrates the relationship between observed sales and PRB, where the correlation between the two is apparent, motivating the models introduced in Tables 3.2 and 3.3. Situmeang et al. (2014) find that there is a significant carry-over effect between video game generations, i.e. a generation with good or bad user experience will impact the upcoming release, further strengthening the stipulated models.

3.4.1 Pre-release buzz data

To represent PRB, we collected search traffic information from Google Trends (www.google.com/trends) for each video game. In using Google Trends, there are two relevant modelling questions. The first regards the number of periods of recorded search volume we consider, the window length w, and how many periods l this information is leading before the launch of the product of interest at time T, the launch date. Consequently, for these windows, we then sum Google Trends (GT) so that:

$$\text{PRB}_j = \sum_{k=1}^{w} \text{GT}_{j,(T-l-k)}. \tag{3.2}$$

In the experimental setting it is possible to either expand the PRB window over time or keep it fixed. While studies with an expanding window size (e.g. Gelper et al., 2015; Xiong and Bharadwaj, 2014) report gradually improved forecast accuracy towards release, Kulkarni et al. (2012) find better performance in smaller window sizes further away from the release date. They argue that one month prior to the release, PRB models the future adoption better, but they do not evaluate smaller window sizes closer to release. Therefore, it is unclear whether forecasts change due to the window size change or the distance from the release date. Some indication comes from the findings by Kim and Hanssens (2017), where the same window size helps to measure the adoption closer to release, but the window consists of one observation only. Therefore, since there is no clear recommendation stemming from the literature, we investigate both dimensions: (i) the window size; and (ii) the lead time.

To select keywords we use the Google Knowledge Graph entity that combines linguistic and semantic related keywords into one search query. In contrast to a single keyword, this leads to more robust search traffic coverage (e.g., Schaer et al., 2019a; Siliverstovs and Wochner, 2018). When there is no Google Knowledge Graph entity available, we use the video game title as the keyword.

Google Trends is peak scaled, i.e. the queried data are always scaled between 0 and 100, thus making different data requests non-comparable. Each query allows retrieving search volume for up to five keywords. Since franchises may have more than five games, we use a neutral scaling keyword, in our case the keyword "marker", to achieve comparable data, without requiring an excessive number of queries. We scale each video game's popularity relative to the reference keyword. This is done by calculating the sum of the scaling keyword's popularity for a specific period and scale it to the sum of video game keyword (see, Schaer et al., 2019a; Kim and Hanssens, 2017, for a discussion on scaling and data reliability of Google Trends).

We run the experiment with a preset rolling window evaluation from 26 until 1 week prior to release and re-run it with different PRB window sizes from 4 to 12 weeks. At a maximum we consider search traffic information up to 38 weeks prior release, an approach similar to Xiong and Bharadwaj (2014) who included pre-release buzz 181 days pre-launch (26 weeks).

3.4.2 Life-cycle curves

For the adoption curve $F(\cdot)$ in Equation (3.1) we use a set of parametric and a non-parametric life-cycle curves. We consider a number of alternative curve specifications to establish the usefulness of PRB beyond a single life-cycle specification.

Parametric curve fitting

The adoption curve illustrated in Figure 3.1, with its steep decline, is typical for some hedonistic products and the literature suggests several potential candidates (see, Clement et al., 2006, for a review). When forecasting with analogies, the Bass curve is the most frequently used (e.g. Lee et al., 2014; Goodwin et al., 2013; Bass et al., 2001). The cumulative adoption from the Bass curve for generation j is

$$\hat{F}(j,t) = \frac{1 - \exp\left(-\left(\hat{p}_j + \hat{q}_j\right)t\right)}{1 + \exp\left(\frac{\hat{q}_j}{\hat{p}_j}\right)\left(-\left(\hat{p}_j + \hat{q}_j\right)t\right)}, \tag{3.3}$$

where p_j and q_j are the coefficients of innovation and imitation for generation j respectively (Bass, 1969). After each generation, we update the shape parameters, to take into account any potential change in buying behaviour over time, i.e. the rate of innovators may be different in the second generation due to the gaming experience in the previous generation.

We consider two variants of the Gompertz curve, which provided the best fit on a mobile game diffusion study (Yi et al., 2019). The cumulative adoption is defined by

$$\hat{F}(j,t) = \exp\left(-\hat{a}_j \exp\left(-\hat{b}_j t\right)\right), \tag{3.4}$$

where a_j is the shape parameter and b_j the coefficient for scaling. The three parameters of the Gamma/Shifted Gompertz (G/SG) (Bemmaor, 1994) offer greater flexibility, and in the literature, there is evidence of good forecasting performance (Bemmaor and Zheng, 2018; Meade and Islam, 2006; Bemmaor and Lee, 2002). The shift is achieved by adding the c_j parameter:

$$\hat{F}(j,t) = \left(1 - \exp(-\hat{b}_j t)\right) \exp\left(-\hat{a}_j t\right)^{\hat{c}_j t}. \tag{3.5}$$

Furthermore, we examined the Weibull curve that has been used for hedonistic products (e.g. Moe and Fader, 2002) but this resulted in very poor performance. Similarly, we also abandoned generational diffusion models that take into account interaction between generations as they did not converge (e.g. Bass and Bass, 2001; Norton and Bass, 1987).

All parametric life-cycle models are estimated using non-linear least squares using the *diffusion* package (Schaer and Kourentzes, 2018) for R (R Core Team, 2018). Each curve is fitted only on data available up until the release of the new game including the required lead time, i.e. the parameters of a curve for generation j are estimated on generation $j-1$ up until period $T-l$. Since the sales history for some games can be several years until the launch of the next instalment in the franchise, we truncated the right tail to facilitate the estimation of the curve parameters, which is common practice for life-cycle models (e.g., Acimovic et al., 2019). The truncation occurs when the growth rate is less than 0.05% of the cumulative sales per week. On average this limits the series to around 40 weeks.

Non-parametric curve fitting

Besides the parametric models, we introduce a non-parametric curve, based on the centred moving average. In contrast to parametric curves, this approach can flexibly approximate any adoption shape. The curve is defined as

$$\hat{F}(j,t) = \frac{1}{2k+1} \sum_{v=-k}^{k} Y_{j-1,t+v}, \tag{3.6}$$

where the order of the centred moving average is $2k+1$. We set k equal to 4 periods and interpolate the first and last $\frac{k}{2}$ periods of observed sales $Y_{j,t}$ from and to zero linearly.

3.4.3 Performance evaluation

There are various ways to asses the performance of the suggested market potential estimation models, one being investigating its in-sample fit. However, such performance assessment is not reliable, as it is prone to overfitting. To test the validity of the proposed PRB models we evaluate their predictive performance. Any model that can predict has adequately approximated key elements of the underlying market structure. We generate forecasts for the market potential m_j, with $j \geq 2$, for each product, as described in Section 3.3. Then we generate cumulative forecasts for each of the life-cycle curves in Section 3.4.2, for 52 weeks. For some cases, the evaluated forecast horizon is shorter, as there are no further observations in the dataset.

Table 3.4: Market potential estimation benchmarks

Model	Equation	Estimation point
B.1	$\hat{m}_j = m_{j-1} + \varepsilon_j$	$j \geq 2$
B.2	$\hat{m}_j = c_0 \cdot m_{j-1}^{c_1} \cdot \varepsilon_j$	$j \geq 4$
B.3	$\hat{m}_j = \alpha_0 + \alpha_1 m_{j-1} + \varepsilon_j$	$j \geq 4$

To assess the gains from the inclusion of PRB in the estimation of the market potential, we consider a number of benchmarks, presented in Table 3.4. A good way to assess predictive performance is to compare against simpler models that do not rely on parameter estimation (e.g. Armstrong, 2006). Such models, are simple to implement, at little computational cost and offer a good reference point for cross-study comparisons. In the case of pre-release forecasting, the Random Walk model corresponds to the observed market potential from the previous product generation (B.1). A second class of benchmarks (B.2 & B.3) assesses the performance against the univariate equivalent of the models introduced in Table 3.2

and 3.3. This directly displays how much benefit the incorporation of PRB has within the same model family. For each model we assess the performance at each generation using the Geometric Mean Relative Absolute Error (GMRAE):

$$\mathrm{GMRAE}_{i,h} = \sqrt[n]{\prod_{i=1}^{n}\left(\frac{\mathrm{AE}_{i,r}}{\mathrm{AE}_{i,b}}\right)},$$
$$\mathrm{AE}_{i,h} = \left|\hat{y}_{i,t+h} - y_{i,t+h}\right|,$$

where r is the candidate model that gets divided by the Absolute Errors of the benchmark b, which in our case is the Random Walk forecast (B.1) in Table 3.4, for each series i and horizon h. The error is summarised with the geometric mean across all n products. GMRAE is an intuitive scale independent error metric. Errors smaller than 1 imply that the evaluated model outperforms the selected benchmark (Ord et al., 2017) by (1-GMRAE)*100%. Furthermore, we track the resulting bias of the predictions, as this provides insights on whether a model is prone to over- or under-forecast. We report bias in the form of the Relative Mean Error (RelME)

$$\mathrm{RelME}_{i,h} = \frac{n+1}{2}\frac{\mathrm{ME}_{i,r}}{|ME_{i,b}|},$$
$$\mathrm{ME} = \frac{1}{n}\sum_{h=1}^{n}\left(\hat{y}_{t+h} - y_{t+h}\right),$$

where the mean error of the candidate model r gets divided by the absolute Mean Error of the Random Walk benchmark (B.1) b. This is done to adjust for the scale of the errors. It follows that a candidate models with a positive value tends to over-forecasts and vice-versa, while normalisation by the benchmark allows for comparison across products. The metric is summarised with the median across all n products, as the geometric mean is unsuitable due to the possible negative values that can appear.

3.4.4 Results

Performance for the first week and end-of-life

Table 3.5 presents the performance for the first week (FW) and end-of-life (EoL) cases for the different game generations and lead time of 7 periods. The numbers of estimated games for each generation is indicated in brackets for each column. We find that a PRB window size of 6 periods consistently performs best across different lead times, which we will discuss in more detail later in this Section. We treat the end of the forecasted horizon of 52 weeks as defining the end-of-life of a product. Any empty cells reflect models that cannot be estimated due to insufficient observations. Models are identified as specified in Table 3.2 and 3.3, with the multiplicative specification marked as M and A for the additive case. The reported values are the combined performance for all life-cycle models. Next to each error, an arrow indicates the average bias direction; pointing down or upwards for under-forecasting or over-forecasting, respectively. The best performing model for each column is highlighted in bold.

Overall, the best performing model for the first week sales is M.5, however for the 2nd generation of games M.6 performs better. The latter also outperforms all other models on the end-of-life scenario. Both models help to better explain the market potential compared to B.1, the best performing benchmark, for the first week as well the end-of-life scenario by 13% and 24.3%, respectively.

We also note that the fully specified model (M.1 & A.1) trails against most of its restricted counterparts. This suggests that more parsimonious model specifications can better estimate the market potential, both for the first week and the end-of-life sales. The sample size needed to estimate the full unrestricted model limits its applicability and performance. This effect is also evident in the benchmarks, where the autoregressive based B.2 is able to outperform B.1 for 5 and more generations but still trails M.5 and M.6 by 3.8% and 10.3% for the first and end-of-life, respectively.

So far we have only considered the performance of the first and last week of sales. Marketers, however, are also interested in the predictive performance over the evolution of the life-cycle. For instance, this supports planning advertising campaigns. Table 3.6

Table 3.5: Generation overall GMRAE performance across all adoption curves ($w = 6, l = 7$).

	Gen. = 2 (57)		Gen = 3 (46)		Gen = 4 (29)		Gen = 5+ (66)		Overall	
Model #	FW	EoL	FW	EoL	FW	EoL	FW	EoL	FW	EoL
M.6	**0.918** ↓	**0.943** ↓	0.923 ↓	**0.617** ↓	0.884 ↓	**0.687** ↓	0.867 ↓	**0.822** ↓	0.898 ↓	**0.757** ↓
A.6	1.005 ↓	1.000 ↓	1.005 ↓	0.977 ↓	1.000 ↓	1.000 ↓	1.003 ↓	1.002 ↓	1.003 ↓	0.995 ↓
M.5	1.073 ↓	1.364 ↓	**0.898** ↓	0.806 ↓	**0.716** ↓	0.734 ↓	**0.832** ↓	0.914 ↓	**0.870** ↓	0.927 ↓
A.5	1.007 ↓	1.000 ↓	1.005 ↓	0.977 ↓	1.000 ↓	1.000 ↓	1.003 ↓	1.002 ↓	1.003 ↓	0.995 ↓
M.4			1.209 ↓	1.346 ↓	1.193 ↓	1.337 ↓	0.953 ↓	0.998 ↓	1.112 ↓	1.215 ↓
A.4			1.058 ↓	0.988 ↓	1.138 ↓	1.179 ↓	0.968 ↓	0.967 ↓	1.053 ↓	1.040 ↓
M.3					1.906 ↑	3.616 ↑	0.955 ↑	1.793 ↑	1.349 ↑	2.546 ↑
A.3					1.269 ↓	1.475 ↑	0.914 ↓	0.919 ↑	1.077 ↓	1.164 ↑
M.2					0.931 ↓	1.383 ↓	1.037 ↓	1.662 ↓	0.982 ↓	1.516 ↓
A.2					1.066 ↓	1.338 ↑	1.088 ↓	1.426 ↑	1.077 ↓	1.381 ↓
M.1							0.975 ↓	1.099 ↑	0.975 ↓	1.099 ↑
A.1							0.976 ↓	1.124 ↑	0.976 ↓	1.124 ↑
B.1	1.000 ↓	1.000 ↓	1.000 ↓	1.000 ↓	1.000 ↓	1.000 ↓	1.000 ↓	1.000 ↓	1.000 ↓	1.000 ↓
B.2					2.804 ↓	4.882 ↑	0.870 ↓	0.925 ↓	1.562 ↓	2.125 ↓
B.3					1.707 ↓	3.201 ↓	0.956 ↓	0.947 ↓	1.278 ↓	1.741 ↓

Bracket indicate numbers of games within each generation bucket; best performance in bold
↑= over-forecasting on average; ↓= under-forecasting on average

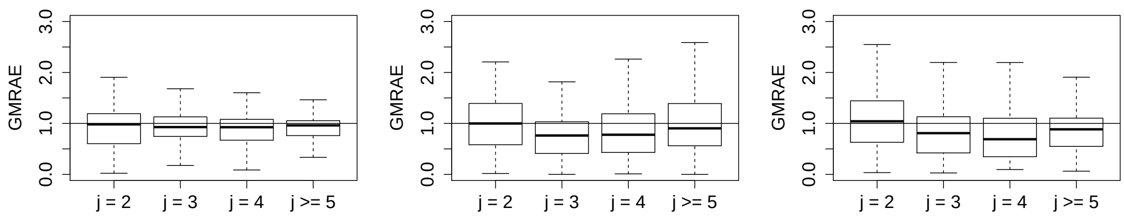
presents the result for model 1 and its most restricted counterparts, model 5 and 6. The overall result in the last column confirms that at longer horizons only the very restricted models are competitive with the benchmark B.1. Looking at the individual life-cycle curves, the *CMA* and the *Gompertz* curves perform best. Note that for the latter comparison we assessed all individual curves against the Bass Random Walk (B.1 Bass).

Table 3.6: GMRAE performance across all generations and forecasting horizons ($w = 6, l = 7$)

Model #	Bass	Gompertz	G/SG	CMA	Overall
M.1	1.259	1.138	1.216	**1.090**	1.150
A.1	1.225	1.112	1.332	**1.021**	1.143
M.5	0.908	0.904	0.927	**0.887**	0.913
A.5	1.004	0.985	1.009	**0.952**	0.995
M.6	0.817	0.797	0.836	**0.772**	**0.811**
A.6	1.004	0.985	1.009	**0.952**	0.995
B.1	1.000	0.991	1.021	**0.958**	1.000
B.2	1.379	1.673	**1.403**	1.408	1.437
B.3	1.452	1.438	1.629	**1.403**	1.452

In order to further confirm the suggested finding that PRB adds predictive value we test whether differences are significant and not due to randomness. As the GMRAE errors are not normally distributed, we apply the non-parametric Friedman and the post-hoc Nemenyi tests (Hollander et al., 2014; Demšar, 2006). We use them to see whether any of the errors from models M.6, A.6 and B.1 are significantly different, by using the tsutils v.0.9.0 package for R (Kourentzes, 2019). We find that for for all scenarios the first week and end-of-life sales all models are significantly different from each other. Therefore, we can confirm *H1* that PRB contains predictive value not only for the initial week but is also for the entire market potential.

In *H2* we argue that the predictive content of pre-release buzz changes across product generations. Since the generations have a different number of sample sizes, the Friedman test is not appropriate, and we replace it with a non-parametric Kruskal-Wallis. It appears that the errors of M.6 are only significantly different across generations for the the end-of-life and all horizon case. In Figure 3.2 the box plots show the error spreads of M.6. We see that, for the first week sales (Figure 3.2a), the variance steadily decreases the more mature the franchise gets and is smaller than the end-of-life scenario (Figure 3.2b) where error variance increases over generations. Therefore, we only accept *H2* that the predictive ability of PRB varies across product generations for longer forecast horizons.



(a) First week sales (b) End-of-life sales (c) All forecast horizons

Figure 3.2: Variance of GMRAE errors across generations for M.6

Finally, we test whether there is a multiplicative or additive relation between PRB and sales. In order to test this connection, we rely on the models that approximate the behaviour of life-cycle best, which in our case are the multiplicative M.6 and its additive equivalent A.6. We subtract additive GMRAE errors from the multiplicative ones for all life-cycles curves. Since the errors are non-normally distributed, we rely on a non-parametric comparison, and consider a negative median to indicate a multiplicative relationship, and vice-versa. To test for significance, we conduct a two-sided Wilcoxon signed-rank test as to whether the differences are non-zero for the first week, end-of-life and all forecast horizons. The test confirms in all scenarios a significant multiplicative relation and we accept *H3*, which suggest that PRB has a multiplicative (proportional) relationship with market potential.

Bias evaluation of PRB

The observed bias, indicated by arrows in Table 3.5, depicts that the best performing models, 5 and 6, are consistently under-forecasting. To provide a better insight of how many time series are effectively under-forecasted, Table 3.7 illustrates the percentage of underestimated time series per generation for M.6, A.6, M.5, A.5 and B.1 (top 5 rows). The bottom rows show the difference between the alternative models. It is evident that there is an overall tendency to underestimate. Note that the bias tendency for B.1, A.6 and A.5 are nearly identical. Except for the second generation, all models are under-forecasting the end-of-life less often than the first week sales, but no clear trend is visible as franchises mature.

Table 3.7: Percentage of under-forecasting across the different generations ($w = 6, l = 7$)

	Gen. = 2		Gen = 3		Gen = 4		Gen = 5+	
Model #	FW	EoL	FW	EoL	FW	EoL	FW	EoL
M.6	75%	69%	67%	60%	74%	55%	70%	58%
A.6	73%	76%	63%	64%	72%	50%	67%	57%
M.5	72%	64%	68%	57%	77%	61%	75%	62%
A.5	73%	76%	63%	64%	72%	50%	67%	57%
B.1	73%	76%	61%	61%	72%	50%	67%	57%
M.6 - A.6	2%	-7%	4%	3%	3%	5%	3%	2%
M.5 - A.5	-1%	-12%	5%	-7%	5%	11%	8%	5%
A.6 - A.5	0%	0%	0%	0%	0%	0%	0%	0%
M.6 - B.1	2%	- 7%	7%	-1%	3%	5%	3%	2%
M.5 - B.1	-1%	-12%	7%	-4%	5%	11%	8%	5%
A.5/6 - B.1	0%	0%	2%	3%	0%	0%	0%	0%

Leading properties and impact of Google Trends window size

The results so far correspond to window size and lead time of 7 weeks. When testing for different Google Trends window sizes (w), we found that having smaller or larger window sizes consistently damaged the forecasting accuracy for all tested lead times. However, we tentatively conclude that the performance of the window size is application dependent and the analyst should evaluate the predictive power of different sizes to identify the appropriate one via out-of-sample test.

The dashed (green) line in Figure 3.3 illustrate the performance of M.6 across different lead times (l) for the first week. PRB has predictive value more than 26 weeks prior to the product release. End-of-life sales (dotted red line) and across all forecast horizons (solid blue line) are both estimated better than first week sales for the last 10 weeks prior to release. Their explanatory power, however, becomes less 17 weeks or earlier to the release date. Note that the sample sizes is reduced to 227 games for the case of 26 weeks lead time, as for some games there was no search traffic signal available. Also, note that the performance for the all horizon scenario steadily improves towards release. For the first-week sales predictions, one of the best performances is observed 10-13 weeks in advance while end-of-life sales have the lowest around 4-6 weeks. For M.5 the results are similar, however the end-of-life sales

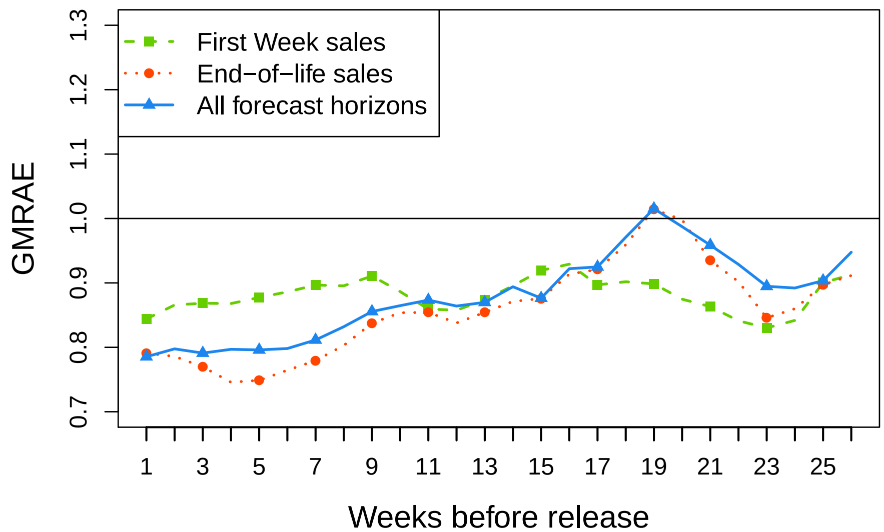


Figure 3.3: Forecast performance of M.6 for all scenarios at different lead times ($w = 6$)

trail B.1 after 10 weeks.

Instead of just using the sum of Google trends the literature suggest to approximate it by using splines (Xiong and Bharadwaj, 2014) or growth curves (Kulkarni et al., 2012; Hann et al., 2011). We investigated a similar approach by predicting Google Trends using exponential smoothing models but found no significant benefit.

3.5 Discussion and concluding remarks

In this study, we set out three hypotheses that we test by performing an empirical out-of-sample evaluation to measure the predictive power of pre-release buzz. We confirm *H1* in that pre-release buzz is not only adding value to predict early week sales but also in explaining the entire market potential of new products. We find that PRB is only heterogeneous in its predictive ability across sequential products explaining end-of-life sales, partly confirming *H2*. Finally our findings suggest that the relationship between PRB and the market potential is multiplicative as postulated in *H3*.

The empirical analysis shows that PRB significantly contributes to explaining first week and end-of-life sales above naïve models. Our results confirm the previous findings from the literature (cf. Section 3.2) with regards to first week sales and provide new insights on PRB containing information for the entire market potential. Pre-release buzz carries a different amount of information for the first week and end-of-life sales. It seems counter-intuitive that long-term estimations consistently perform better than the first week sales when forecasting with PRB closer than 9 weeks prior to release. The literature suggests pre-release advertising and publicity to positively impact long-term sales (e.g. Burmester et al., 2015; Elberse and Anand, 2007). We show that this relationship is also true for PRB. To some extend this is expected, given the close dependency of marketing efforts and pre-release buzz (Kim and Hanssens, 2017), but our findings highlight that the benefits from measuring PRB reaches beyond the first week, a fact which has been overlooked by the literature.

Forecast improvements are achieved using search traffic well in advance of the release date. Notably, the first week of sales forecasts including PRB outperform our benchmarks for all lead times up to 26 weeks, however, for the entire market potential we find that only shorter lead times were helpful. We argue that the closer to release, the greater is the impact on the the long-term sales, as it increases the certainty of consumers. Moreover, advertising campaigns for video games typically start around 9 weeks prior to release (Marchand et al., 2017) and may explain why PRB becoming more informative towards release. The maximum feasible lead-time that is achievable, however, depends on the product type, as consumers invest different amounts of effort to inform themselves about different products (Bhatnagar and Ghose, 2004).

We observed that the first week error variance for M.6 reduces as the franchise matures, but remains changing for the end-of-life and all horizon scenario. This explains why *H2* is only accepted for end-of-life and all forecast horizons. A higher error spread for long-term scenarios is expected as additional factors become relevant, such as word-of-mouth or post-release advertising activities, which affect the overall adoption. The literature argues that the uncertainty of product adoption is reduced over the life-span of a franchise and directly cannibalises advertising effects (Hennig-Thurau and Houston, 2019). This would intuitively imply that pre-release buzz becomes a weaker indicator over generations since consumers are familiar with its content and might reduce their search activity which would indicate fewer sales. However, our findings do not support this interpretation since the forecast error variance decreases for first week sales and heterogenous for long-term over time. One possible reason is that internet searches might become more focused: If consumers know the product, they might still check up on product, for example its release date or price, but it receives fewer unfocussed searches from consumers who may eventually not buy the product.

From a modelling perspective, it is interesting to see that the most restricted alternatives (M.5 & M.6) have the best explanatory performance of new product sales. In contrast to model M.5, which has a PRB coefficient set to 1, setting it to 0.5 in M.6 greatly reduces error variance. Although, the first week estimations overall achieves a slightly lower explanatory power (+2.8%), the model becomes particularly useful for the end-of-life scenario and 2nd generation where accuracy increases more than 10%. In general, the multiplicative formulation is more prone to generating larger outliers, since it is sensitive to shifts in the PRB. The additive formulation is more conservative and exhibits less bias, but adds only limited predictive information. M.6 largely overcomes this issue. This shows that (i) parsimonious PRB models clearly outperform more complex counterparts in the presence of few product generations and (ii) further restricting the coefficients, as in M.6, which is motivated by model combination and shrinkage, increase predictive performance further.

3.5.1 Managerial implications

Expert judgment is the predominant approach in new product forecasting, partly due to the complexity of relevant statistical methods (Kahn and Chase, 2018). Our research addresses this issue by introducing a forecasting PRB adoption model for sequential product generations that is relatively simple to implement and can be readily automated. This is particularly true given that the best performing prediction of market potential are achieved by incorporating percentage change between two generations and the simple moving average model being the best individual life-cycle curve. We advise to use a multiplicative formulation of PRB as shown in model M.6. Moreover, the high percentage of demand under-forecasts offers a potential route for further exploring forecast combinations. In a recent survey, more than 70% of managers believe they over-forecast new product demand (Kahn and Chase, 2018). Although, the literature on combining statistical pre-launch forecasting models and expert judgment is sparse (Perera et al., 2019), a 50-50 equally weighted forecast combination has been shown to work well (Petropoulos et al., 2018; Blattberg and Hoch, 1990).

For video game publishers the appropriate allocation of ad spending is crucial. Our research suggests that stimulating PRB is important for achieving greater product success. The study by Marchand and Hennig-Thurau (2013) suggests that for the duration of our dataset coverage, publishers invested more in post-release campaigns. This is counter-intuitive to how the sales adoption takes place and also conflict with the scholarly opinion, which found pre-release advertising of video games to be marginally more effective than post-release advertising (Hennig-Thurau and Houston, 2019). Both Burmester et al. (2015) and Kim and Hanssens (2017) suggest a greater focus on pre-release campaigns and to create awareness through communication. Furthermore, the timing and design of the pre-release advertisement impacts on the generated buzz (Nguyen and Chaudhuri, 2018; Schroll and Grohs, 2018).

3.5.2 Limitations and further areas of research

Our approach only includes search traffic information from Google Trends to model future market potential. To cover PRB fully, Houston et al. (2018) suggest the inclusion of anticipatory communication, search and participation in experiential activities type of data. However, so far the investigated studies only cover one or two sources (cf. Section 3.2). Therefore, we cannot rule out the possibility of other sources contain more predictive information.

In practice, a limiting factor might be that not every product has clear keywords to obtain pre-release buzz signals (for a review see, Schaer et al., 2019a). We assume, however, that many manufacturers of popular electronic goods such as digital cameras, mobile phones and to some extent computer components, are ideal candidates to apply the suggested approach since those products are often highly anticipated and their communities discus the expected new arrivals actively online. In contrast, more generic or niche products may not have easy to capture pre-release signals. For those type of products crowd-sourced keywords, as suggested by Brynjolfsson et al. (2016), is a possible solution that we will explore further in future work.

A natural extension of our work is to investigate the interaction of PRB, its relationship to marketing spend and the effects on product adoption shape parameters (as some initial findings by Wang et al., 2010, suggest). Moreover, besides the option to combine statistical and judgmental forecasts, as suggested earlier, there is also room to investigate to what extent PRB can be used to support judgmental decision making as evidence from the literature suggest that contextual information can lead to better forecasts than just using historical sales data information (Seifert et al., 2015).

3.5.3 Conclusions

Our study directly contributes to the literature on diffusion modelling. Many studies have omitted the estimation of the market potential by assuming it to be known (e.g. Baardman et al., 2017; Lee et al., 2014; Goodwin et al., 2013). We conclude that the evidence provided by our PRB model supports its importance in predicting market potential of future generations of sequential products. Our proposed method, based on PRB, has the advantage that it is inexpensive to collect and reflects the consumer interest directly.

The suggested PRB model relies on a freely available source that provides organisations with insights on product standing and competition. Since the data is available with high frequency, shifts in consumer interest can be monitored and associated with the sales expectations (Belvedere and Goodwin, 2017). The lead time over which such predictive PRB information is available would allow decision makers to continuously adjust their marketing and operational planning until release. Last but not least, the proposed M.6 model requires only two data (generations) points to obtain valuable insights about the market potential, in comparison to models suggested in the literature.

Chapter 4

Competitive intelligence with pre-release buzz: Predicting competitors' new product sales

In today's competitive market environment it is vital to have insights about competitors' new product launches. Past studies have demonstrated the predictive value of pre-release buzz (PRB) for forecasting new products. Relying on those findings and its public availability, we investigate its usefulness for estimating sales of competing products. We propose a model for predicting the success of competitors' product launches based on our past product sales data and competitor pre-release Google Trends. We find PRB to increase the predictive accuracy by more than 18% compared to models that only use internal available sales data and product characteristics of video game sales. We conclude that this inexpensive source of competitive intelligence can be helpful when managing the marketing-mix and in supporting new product release planning.

4.1 Introduction

In today's competitive market environment, marketers and senior management are concerned with both the success of their products and closely monitoring competitors' actions. The internet has facilitated Competitive Intelligence (CI), offering companies new ways to observe competitors' actions (Teo and Choo, 2001) and is of strategic importance to organisations (Calof and Wright, 2008). At the aggregate high-level, typical key performance indicators at brand level include the estimation of penetration, market share or the share of wallet (Zheng et al., 2012). At a lower-level companies regularly assess the performance of new product and service success (Fehringer et al., 2006).

Before a company can release a new product, a critical task is to estimate its market potential, which allows the adjustment of marketing and operational actions. New product forecasting is a challenging task, in particular if conducted pre-launch (Goodwin et al., 2013; Trusov et al., 2013). Recent studies suggest that pre-release buzz (PRB) information can substantially improve new product forecasting (e.g. Schaer et al., 2019b; Kim et al., 2015; Xiong and Bharadwaj, 2014). PRB represents the aggregated anticipated interest of consumers towards a new product (Houston et al., 2018) and has been collected for example in the forms of search traffic information (e.g. Schaer et al., 2019b; Tian et al., 2014), blogs (e.g. Kim et al., 2015; Dhar and Chang, 2009) or microblogs (e.g. Gelper et al., 2015; Asur and Huberman, 2010).

Companies can easily obtain PRB for competitor products. This makes it attractive to apply new product forecasting in the context of CI, to infer the success of their products. Studies which include PRB information have focused on own products sales only (Schaer et al., 2019b) or they have estimated and predicted new product sales from a dataset that includes products from multiple brands (e.g. Divakaran et al., 2017; Onishi and Manchanda, 2012). While the richer dataset typically provides better estimates, it does not reflect a real-world environment in which sales information is only available for their own brands. This questions the realism and feasibility of the reported value of these estimates. In the analysis that follows we use the word brand as a synonym for companies, publishers or producers. We distinguish between internal and competitor brands. While internal brand data typically provide full access to sales history and marketing activities, external competitor insights are constrained to only what is publicly available. Assuming some degree of homogeneity among competing products, a key question is whether competitor PRB can be used together with internal sales data to infer competitor success.

In order to understand the role of PRB, in this study, we focus on video games that have a short life-cycle and exhibit intense competition. We find that competitor PRB information improves predictions of competitor's new video games market potential by more than 18% compared to models without PRB information. Furthermore, we find that splitting the data by publishers is at least as effective as using product clustering, which supports our homogeneity assumption for the video games market. Our findings enable direct managerial actions, as the additional knowledge about competing products allows marketers to adjust the marketing mix and provides useful insights to plan product releases.

In Section 4.2 we first examine the literature on new product forecasting with PRB information and how this research has addressed competition. Section 4.3 illustrates our forecasting approach to measure the market potential of competitor products with internal sales information and competitor PRB, which we then evaluate in Section 4.4 by forecasting video game sales prior to release using search traffic information from Google Trends. Section 4.5 then discusses the findings and the practical managerial implications with conclusions following in Section 4.6.

4.2 Predicting new products with PRB and the competition aspect

The efficiency of predicting the success of a new product using pre-release buzz information has been well researched, as Table 4.1 illustrates. The majority of studies focused on box office sales. Other studies looked at music album sales (Hann et al., 2011; Dhar and Chang, 2009), video game sales (Schaer et al., 2019b; Xiong and Bharadwaj, 2014) and alpine skis sales (Mülbacher et al., 2011). The authors have used a variety of PRB sources including forum discussions (e.g. Craig et al., 2015; Liu, 2006), blog posts (e.g. Divakaran et al., 2017; Onishi and Manchanda, 2012), posts on Twitter (e.g. Gelper et al., 2015; Asur and

Huberman, 2010) and Facebook (Ding et al., 2017; Kim et al., 2017). Another major PRB source is online search traffic in the form of Google Trends (e.g. Kim and Hanssens, 2017; Kulkarni et al., 2012) or Baidu (Tian et al., 2014). Although not all studies evaluated the predictive performance on a hold-out sample (Table 4.1), the overall conclusion is that PRB provides substantial improvement for pre-release estimation, both with volume and valance based PRB inputs.

In Table 4.1 the column *estimation / hold-out* shows that most studies estimate and evaluate their models across multiple brands. While such a *cross-brand* approach has the advantage of evaluating the effects of PRB on a richer dataset, it is somewhat impractical in terms of operational decision support, as it does not reflect what might be seen in practice, where companies typically only have access to sales data about their own products. However, even with a smaller *intra-brand* sample, PRB can provide significant improvements (Schaer et al., 2019b), but its potential remains unexplored for competitor products. On the other hand, various brand-related variables are not helpful for *intra-brand* samples (Onishi and Manchanda, 2012; Foutz and Jank, 2010; Dhar and Chang, 2009) and their predictive information has been questioned by Foutz and Jank (2010). This possibly explains why only limited studies included it, as the column *company control* depicts.

Although the general scholarly view is that the success of entertainment products depends on the competition and strength of own brand (see Hennig-Thurau and Houston, 2019), there are mixed findings regarding their relevance when PRB is considered. For example, studies that measured competition by the number of competitors' products released during the same period (Divakaran et al., 2017; Xiong and Bharadwaj, 2014; Kulkarni et al., 2012), or the average age of top 20 movies (Liu, 2006), found an insignificant competition effect on sales. Contrary to this finding, Kim et al. (2017) reported substantial gains in forecast accuracy when including a broader set of competition variables such as the number of seats and screens for the top 5 movies from the same genre or distributor. However, none of these studies use any competitor PRB information.

Research suggests that PRB is impacted by advertising expenditure and genre (Xiong and Bharadwaj, 2014) and whether it is a sequential and non-sequential product (Craig et al., 2015). Divakaran et al. (2017) reported significant effects from star power within a

Table 4.1: Forecasting with pre-release buzz

Study	Buzz predictor*	Estimation / Hold-out◇	Company control†	# series	Model⋆	Target variable	Buzz measure°	Horizon‡
Liu (2006)	FOM	x / -	b / c	40	LR	Box office	Vol.; Val.	1 w
Dhar and Chang (2009)	BLG	x / -	- / -	108	LR	Music album sales	Vol.	3 w
Foutz and Jank (2010)	VSX	x / x	b / -	262	FDA	Box office sales	-	1 w
Asur and Huberman (2010)	TWR	x / -	- / -	24	LR	Box office sales	Vol.; Val.	1 d
Wang et al. (2010)	FOM	x / x	- / -	51	DnC	Box office sales	Vol.	2 w
Hann et al. (2011)	P2P	x / x	- / -	172	FDA	Music album sales	-	1 w
Mülbacher et al. (2011)	FOM	x / -	- / -	10	LcR	Ski sales	Vol.; Val.	1 y
Kulkarni et al. (2012)	GTD	x / x	- / c	61	LR	Box office sales	Vol.	1 w
Onishi and Manchanda (2012)	BLG	x / x	b / -	1729	SEM	Box office sales	Vol.; Val.	1 d
	BLG	x / x	b / -	90	SEM	Cell phone service	Vol.; Val.	1 d
Tian et al. (2014)	BAU	x / x	b / c	92	ML	Box office sales	Vol.	1 d
Xiong and Bharadwaj (2014)	BLG; FOM; GTD	x / x	- / c	681	FDA	Video game sales	Vol.	3 w
Craig et al. (2015)	FOM	x / -	- / -	62	LR	Box office sales	Vol.	1 w
Gelper et al. (2015)	TWR	x / x	- / -	106	LR	Box office sales	Vol.; Val.	1 d
Kim et al. (2015)	BLG	x / x	- / -	212	ML	Box office sales	Vol.	1 w
Ding et al. (2017)	FBK	x / -	- / -	64	LR	Box office sales	Vol.	1 w
Divakaran et al. (2017)	BLG	x / x	- / c	373	PR	Box office sales	Vol.	1 w
Kim and Hanssens (2017)	GTD; BLG	x / x	- / -	137	LR	Box office sales	Vol.; Vol.	1 w
Kim et al. (2017)	SNS	x / x	- / c	175	ML	Box office sales	Vol.; Val.	PLC
Schaer et al. (2019b)	GTD	i / i	- / -	255	DnC	Video game sales	Vol.	PLC
This study	GTD	i / c	b / -	240	FDA	Video game sales	Vol.	PLC

* BAU = Baidu, BLG = Blog, FBK = Facebook, FOM = Forum, GTD = Google Trends, P2P = Peer-to-Peer Network, SNS = Social Network Services, TWR = Twitter, VSX = Virtual Stock Exchange; ◇ x = cross-brands, i = intra-brand, c = competitor, - = no hold-out evaluation
† b = brand variable, c = competition variables; ⋆ DnC = Diffusion curves,FDA = Functional Data Analysis, LcR = Logistic Regression, LR = Linear Regression, ML = Machine Learning, SEM = Structural Equation Model; ° Vol. = Volume based PRB measure, Val. = Valence based PRB measure; ‡ d = daily, w = weekly, m = monthly, PLC = Product life-cycle.

new movie. Xiong and Bharadwaj noted that for video games there is a publisher effect, but is only significant early in the pre-release phase and vanishes closer towards release, as more details about the video game emerge. This illustrates that PRB is a wide measure which carries various informative dimensions, which are difficult to directly measure for competitors.

Another stream of CI literature investigated how user-generated content can identify new competitors (Abrahams et al., 2013) and developed benchmarks to asses their own social media performance (e.g. He et al., 2015, 2013), but we are unaware of specific research for the pre-release phase. Gopinath et al. (2013) analysed blog market coverage of movie studios, but not in a predictive context. This motivates our investigation on the extent that PRB contains predictive value for competitor's products so as to enhance the CI insights further.

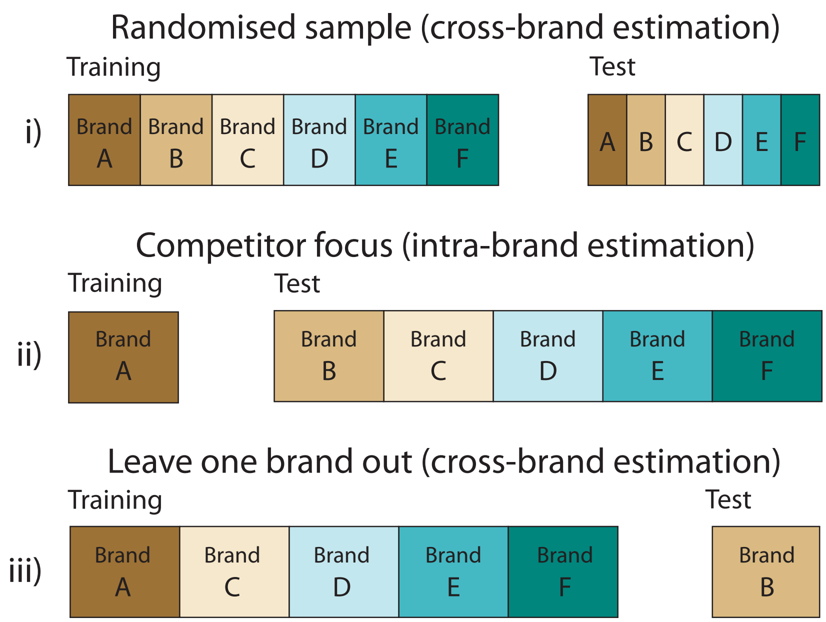


Figure 4.1: Difference between randomised cross-company, intra-company and cross-company competitor forecast

Figure 4.1 illustrates the different ways to split a *cross-brand* dataset. Scenario (i) is the standard approach that is used by most studies that investigated new product forecasting with pre-release buzz: all brands are represented randomly in the training and test sample. This is ineffective for our research, since we are interested in forecasting competitor sales with *intra-brand* information only, we propose to split the dataset according to scenario (ii), and formulate our first research question:

Research Question 1 *Does PRB improve the predictions for competitors' new product, using internal sales data, compared to non-PRB models?*

While this will allow us to investigate the impact of PRB, it does not demonstrate its efficiency compared to *cross-brand* estimation. Therefore, we formulate a second research question that investigates the information loss when restricting the training to *intra-brand* data:

Research Question 2 *Does the incorporation of cross-brand information into PRB models outperform intra-brand PRB competitor models?*

To answer this question we alter the sampling methodology to the case shown in scenario (iii), where sales from multiple brands are available to assess predictions for a single competitor brand, increasing the estimation sample over the previous case. As we noted, in many cases this may be infeasible in practice.

Strictly speaking, restricting the dataset from the *cross-brand* data to just *intra-brand* data is a crude form of clustering. This will only work when the different brands are homogeneous enough, so that the internal products act as analogies for the competition. Such a use of analogies is a common forecasting approach (Goodwin et al., 2014). An alternative way to manual clustering is to use segmentation to split the dataset into more homogeneous sub-samples. Our third research question investigates using brand versus product segments.

Research Question 3 *Are predictions by brands superior to using product segments?*

Last but not least, having the sample reduced to *intra-brand* level substantially reduces the number of available observations, especially for smaller brands. Nonetheless, as discussed earlier, even with small datasets PRB could still provide improvements. The last research question, investigates how estimation sample size impacts predictability.

Research Question 4 *Do brands with a larger sample of intra-brand sales data (from previously launched games) result in PRB models with smaller forecast errors?*

In summary, our first research question investigates the efficacy of the approach for inferring competitors' sales, while the other research questions explore the efficiency and conditions of good performance of the proposed model.

4.3 Estimating the market potential of competitors' new product

To answer the outlined research questions, we introduce the two models outlined in Figure 4.2. The first, is a regression-based model which can be applied to both *intra-brand* and *cross-brand* data, abbreviated subsequently with *iRg* and *xRg*, respectively. The second is a two-staged model, based on segmentation, using clustering and classification. Although in principle we can cluster *intra-brand* data, the available sample size of own launched products might limit its applicability in practice. Therefore, we consider the segmentation model only for the *cross-brand* data, which we use as a benchmark, labelled as *xCl*.

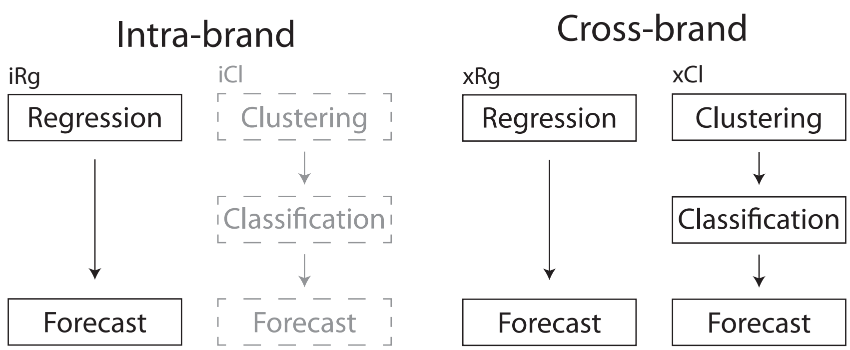


Figure 4.2: Different approaches to obtain competitor forecasts. Clustering for intra-brand data is subject to sufficient sample.

4.3.1 Functional regression

A common way to predict new product sales with PRB is to use functional data analysis (FDA; e.g. Xiong and Bharadwaj, 2014; Hann et al., 2011; Foutz and Jank, 2010). In contrast to classical regression, FDA does not regress the raw inputs directly on the target variable, but instead, for each observation, compresses vectors of values (curves) into scalars (Ramsay and Silverman, 2005). For example, instead of considering all past PRB up to period t as a vector vector of length t, it uses the total PRB volume that is a single number. We regress the cumulative sales $Y_{j,T+h}$ for a new product j at time $T+h$ where h is the desired forecast horizon and T is the product release date. Note that by using cumulative sales the last observation within a product life-cycle automatically reflects the total sales (or market potential within a product adoption context; Bass, 1969). The model is

$$Y_{j,T+h} = f(PRB_{j,t}, ProdChar_j) + \varepsilon_j, \tag{4.1}$$

where $f(\cdot)$ represents a scalar of variables of PRB and product characteristics and $\varepsilon_j = N(0, \sigma_j^2)$. An advantage of this representation is that it allows easy mixing of time-variant information from PRB with time-invariant variables such as product characteristics. This is achieved by reducing vector across time into single vectors describing their characteristics (see Fulcher, 2018, for an general overview on time series dimensionality reduction). The most common way to reduce the time dimension of PRB is either to sum its volume over a certain period (e.g. Gelper et al., 2015; Tian et al., 2014; Wang et al., 2010) or by capturing its adoption dynamics with diffusion model parameters (Kulkarni et al., 2012) and functional principal components (FPC; Xiong and Bharadwaj, 2014; Hann et al., 2011; Foutz and Jank, 2010). Although Xiong and Bharadwaj (2014) and Foutz and Jank (2010) directly compared FPC against volume-based PRB models, they did not include both information types that can be considered complimentary, as they summarise the information in different ways. Since our goal is to maximise predictive power, we use all information types and let the forecasting model determine the influential ones. In addition to the measures described in the literature, we also quantify the velocity and trend, which are defined below.

PRBvolume

For brevity we define the per-period PRB as $g_{j,t}$ and its cumulative as $G_{j,t}$. When measuring the PRB volume, two relevant aspects are the number of periods over which PRB is calculated and the time before release, i.e. the window length w and lead time l to product release date T. We measure the volume as follows:

$$\text{PRBvol}_j = \sum_{k=1}^{w} g_{j,(T-l-k)}. \tag{4.2}$$

Characterising PRB by its volume has the disadvantage that any dynamics are lost and the same volume may follow very different trajectories, which can affect the later product adoption (see Xiong and Bharadwaj, 2014). The subsequent variables try to capture these dynamics in different ways.

PRBvelocity

A simple way to measure the adoption speed is to measure the number of periods it takes to reach a certain percentage of the overall PRB adoption:

$$\text{PRBvelocity}_j = \sum_{n=1}^{N} \mathbb{1}\left\{\frac{g_{j,T-l-w+n}}{G_{j,T-l}} < \tau\right\}, \tag{4.3}$$

where τ is a threshold that can be any number between 0 and 1 and $\mathbb{1}$ is an indicator function that takes the value 1 when its condition is satisfied and zero otherwise.

PRBtrend

Another way to describe the dynamic is to fit a linear trend through to the observed PRB up to the point g_{T-l}. Once this curve is estimated, we can use the slope parameter as the *PRBtrend* measure.

PRBadoption

Another alternative is to parametrise the increasing buzz through diffusion curves (e.g. Lee et al., 2014; Goodwin et al., 2013). The best-known diffusion curve is Bass (Bass, 1969), which describes the word-of-mouth process through innovators and imitators as captured by the coefficients p and q, respectively. First we can fit a Bass curve to each product $G_{j,t}$:

$$\hat{G}_{j,t} = \frac{1 - \exp\left(-\left(\hat{p}_j + \hat{q}_j\right) t\right)}{1 + \exp\left(\frac{\hat{q}_j}{\hat{p}_j}\right)\left(-\left(\hat{p}_j + \hat{q}_j\right) t\right)}. \tag{4.4}$$

Then, we use the estimated $\hat{p}_j$ and $\hat{q}_j$ coefficients as our measure to describe *PRBadoption*:

$$PRBadoption_j = (\hat{p}_j, \hat{q}_j). \tag{4.5}$$

Note that we also experimented with other diffusion curves, but those had a worse fit on PRB.

PRB functional principal components

Functional Principal Components (FPC) analysis has gained popularity for predicting new product adoption (e.g. Fan-Osuala et al., 2018; Sood et al., 2009). Studies that use FPC for forecasting with PRB report better predictive performance against diffusion curves (Hann et al., 2011) and volume-based models (Xiong and Bharadwaj, 2014; Foutz and Jank, 2010). The idea is to characterise and identify unique shapes of all PRB adoption curves through Principal Components Analysis (PCA). It is recommended to first smooth the raw shape by using smoothing splines (Ramsay and Silverman, 2005). This is advised so as to reduce the noise in the recorded PRB.

We decompose the smoothed curve into principal components and use the principal scores as our *PRBfpc* measure, i.e. include for each product a number of individual principal component scores that describe its variation from the mean behaviour. FPC has the disadvantage that all PRB curves are required to be of the same length. Generally, it is easy to have equal length PRB inputs. However, FPC will take dynamics such as slow adoption into account.

Product characteristics

In addition to the PRB information, we consider additional product characteristics related to competitors' product that are available before release. Since we aim to evaluate our research questions on video game sales data, we consider information which are typically available for entertainment products. Those include information such as the *Genre* (labeled as *PCTGenre;* Kim and Hanssens, 2017; Xiong and Bharadwaj, 2014), the *Sequel* number (*PCTsequel*; Craig et al., 2015; Xiong and Bharadwaj, 2014; Hann et al., 2011; Foutz and

Jank, 2010; Liu, 2006) and information about the *Release* time (Kim and Hanssens, 2017; Xiong and Bharadwaj, 2014, *PCTrelease*;). Several studies also included marketing information such as ad spend (e.g. Kim and Hanssens, 2017; Xiong and Bharadwaj, 2014) or competition (e.g. Gopinath et al., 2013; Kim et al., 2017). However, such information will only be available to companies that have access to marketing intelligence databases and potentially at an aggregation level that limits its usefulness for predicting individual product launches. In this research we focus on freely available sources.

The model in equation (4.1) can be written as:

$$\begin{aligned} Y_{j,T+h} = \alpha_0 + \alpha_1 PRBvolume_j + \alpha_2 PRBtrend_j + \sum_{m=1}^{M} \alpha_{m+2} PRBdvelocity_{j,m} + \\ \sum_{d=1}^{D} \alpha_{M+d+2} PRBadoption_{j,d} + \sum_{e=1}^{E} \alpha_{M+D+e+2} PRBfpc_{j,e} + \\ \sum_{z=1}^{Z} \alpha_{M+D+E+z+2} PCTgenre_{j,z} + \alpha_{M+D+E+Z+3} PCTsequel_j + \\ \sum_{v=1}^{V} \alpha_{M+D+E+Z+v+3} PCTrelease_{j,v} + \varepsilon_j, \end{aligned} \tag{4.6}$$

for $j = 1$ to I for *intra-brand* estimation and $j = 1$ to X for *cross-brand*. We will use logarithmic version of all raw inputs (sales, PRB and product characteristics), as suggested by Schaer et al. (2019b).

4.3.2 Functional clustering

A different approach is to use product segmentation to obtain more homogeneous subgroups and use them to forecast instead. We construct segments using clustering, which is a common approach to predict new product sales (recent examples include Hu et al., 2019; Baardman et al., 2017; Basallo-Triana et al., 2017), but we are not aware of applications to PRB data. We use functional clustering where functions are used to reduce the time dimensionality, similarly to FDA (see Goia et al., 2010; Sood et al., 2009, for applications to short time series). We first obtain k-clusters and then train a multi-class classification algorithm to predict the corresponding cluster of a new product. This has the advantage that we can include additional post-release information to increase the richness of the classifier inputs for better performance.

A helpful source to further enrich our segmentation dataset is the inclusion of product reviews (labelled as *PCTreview; see* Chintagunta and Lee, 2012; Zhu and Zhang, 2010; Dellarocas et al., 2007). Another set of information that may be beneficial is to incorporate the dynamic of sales, similarly to how we model PRB information, in the form of sales volume (*SLSsales*), the time it takes until a certain percentage of total sales is reached (*SLSvelocity*) and its adoption curve parameters (*SLSadoption*). However, for modelling video game sales the Gompertz curve is recommended (Schaer et al., 2019b; Yi et al., 2019). We therefore, estimate the adoption of the cumulative sales $Y_{j,t}$ using Gompertz:

$$\hat{Y}_{j,t} = \exp\left(-\hat{a}_j \exp\left(-\hat{b}_j t\right)\right), \tag{4.7}$$

where $\hat{a}_j$ is the estimated shape parameter and $\hat{b}_j$ the estimated coefficient for scaling (Meade and Islam, 2006). These additional inputs are only used for clustering. The classification algorithm uses only data that is available pre-launch, as in (4.6).

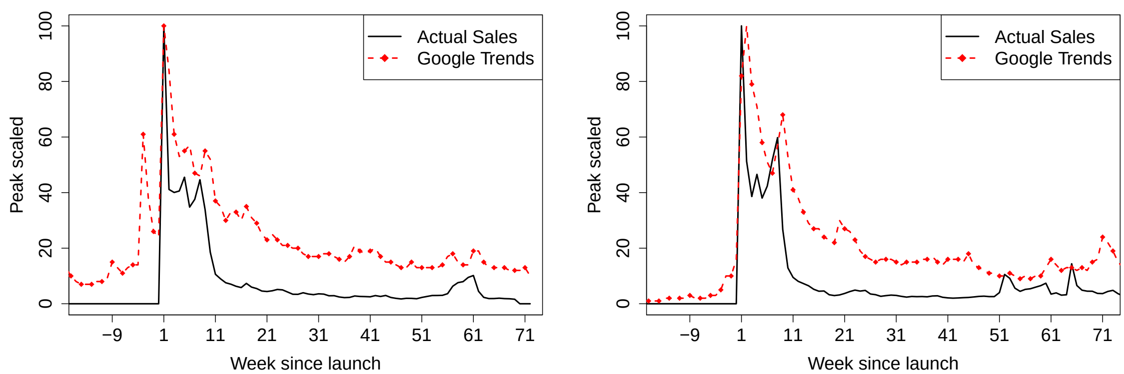
4.4 Predicting the success of new video games releases

We empirically evaluate the outlined research questions, by predicting the sales of video games from different publishers. The video game industry operates in a highly competitive multi-billion dollar market (ESA, 2018). In this market it is common that the gaming community actively engages online on platforms, such as blogs or social media, to discuss new video game before release. This is further fuelled by publishers with advertising activities (Marchand and Hennig-Thurau, 2013). Furthermore, the relative homogeneity of products and short-life cycles is helpful for our research. The literature on PRB, predominately focuses on forecasting opening sales. Schaer et al. (2019b) showed that PRB contains valuable information to predict complete life-cycle sales. This aligns with the objectives of competitive intelligence that typically have a long term strategic focus. Therefore, we evaluate both horizons, opening and total sales achieved by end-of-life *(EoL)*.

4.4.1 Data

Our dataset consists of weekly physical video games sales data from VGChartz (http://www.vgchartz.com; also used, for example, by Marchand et al., 2017; Xiong and Bharadwaj, 2014). We aggregate sales across console platforms for 240 games that represent 23 well-known publishers such as EA, Ubisoft and Capcom. The average number of games per publisher in our dataset is 10.4 (median 4) and ranges from 1 to 61 titles.

In some instances, the sales history spans several years with its tail only observing a few sales. In those circumstances, we truncate the sales time series when the growth rate of the cumulative sales per week is less than 0.05%. Given this assumption, the average end-of-life is typically reached 40 weeks after launch. This is considered as our total sales target.



(a) Battlefield V

(b) Dragon Age

Figure 4.3: Typical video game sales pattern and the pre-release buzz of search traffic

To represent PRB, we collected for each video game search traffic information from Google Trends (www.google.com/trends). Figure 4.3 illustrates the available pre-release buzz signal before release and the subsequent adoption of sales. For each game, we download its topic popularity, using the Google Knowledge Graph entity. This method combines linguistic and semantic related keywords into one search query, which leads to more robust search traffic coverage opposed to single keywords (see discussion by Schaer et al., 2019a; Siliverstovs and Wochner, 2018). If there is no Google Knowledge Graph entity, then we use the video game title as the keyword.

Since Google Trends is peak scaled (values between 0 to 100) it makes individual data queries non-comparable. Although Google allows retrieving search popularity for up to five keywords per request, downloading multiple keywords becomes complicated, as new keywords might have higher volume which requires re-scaling. To overcome this we use the same scaling procedure as proposed by Schaer et al. (2019b). They use the neutral scaling keyword “marker” to scale each video game’s popularity accordingly.

4.4.2 Feature estimation

PRB data

Several studies showed that the predictive power of buzz increases towards release (e.g. Kim and Hanssens, 2017; Xiong and Bharadwaj, 2014). Therefore, the decision lead time becomes a trade-off between maximising forecast performance and the management’s operational requirements. For this experiment we set lead time l to 1 as we are interested in the maximum available forecast performance. For a discussion about lead time of PRB on video games we refer the reader to Schaer et al. (2019b) and Xiong and Bharadwaj (2014).

For *PRBfpc* we set the window length w to 26 weeks (182 days) similar to Xiong and Bharadwaj (2014). For all other measures, we use a flexible window length w for each product, dependent on when search traffic becomes available. To avoid any spurious start of PRB, we limit the maximum window length to 40 weeks and require two consecutive observations. The week when PRB becomes available marks our 0% entry for *PRBvelocity*, with further inputs measured at 25%, 50% and 75% of the observed PRB adoption.

All analysis and model estimation is carried out in the statistical programming language R (R Core Team, 2018). We estimate the *PRBtrend* coefficient with ordinary least squares. The *PRBadoption* coefficients for the Bass curve are estimated using non-linear least squares, as implemented in the diffusion package v.0.2.7 (Schaer and Kourentzes, 2018). All PRB curves are smoothed with b-splines, using Akaike Information Criterion to determine the smoothing parameter λ, that is available in the cobs package v.1.3.3 (Ng and Maechler, 2017). For *PRBfpc* we include the first 4 principal components, as this provided the best predictive performance (similar to Xiong and Bharadwaj, 2014; Hann et al., 2011).

Product characteristics data

We encode the video game genre (*PCTgenre*) and November release (*PCTrelease*) as categorical variables. November is the month with the most releases in our dataset. We also experimented in encoding all months individually but this led to worse predictive performance. For each video game we also indicate the sequel number *PCTsequel*.

For the segmentation, we also include for each game the review (*PctReview*) scores from MetaCritic and IGN, which are available publicly on the Kaggle dataset repository (https://www.kaggle.com/datasets).

Sales data

Similarly to the *PRBvelocity* defined in equation (4.3) we measure the *SLSvelocity* of sales as the number of weeks it takes to reach 25%, 50%, 75% and 95% of the overall adoption. We use the diffusion package for R (Schaer and Kourentzes, 2018) to estimate the parameters of the Gompertz curve for the inputs of *SLSadoption*. The last two inputs included in the model are the opening week and total sales (*SLSsales*).

4.4.3 Predictive algorithm

Regression algorithm

To estimate the market potential of competitors' new products with PRB information, we use Random Forest (Breiman, 2001). This machine learning method is an ensemble technique that uses bootstrapping in order to build a large number of decision trees and then select the most voted one. We opt for Random Forest rather than linear regression, as the former can capture flexibly variable instructions beyond the simple linear ones. An additional motivation to use Random Forest is that it is performing well in both regression and classification, simplifying the modelling. The Random Forest algorithm is available in the caret package v.6.0-81 for R (Kuhn et al., 2018).

We train a PRB model for both the *intra-brand* and *cross-brand* scenarios, labelled hereafter as *iRgPRB* and *xRgPRB*, respectively. In the *cross-brand* case, we use a 10-fold cross-validation approach to tune the model parameters. The restrictive *intra-brand* sample size requires the use of Leave-One-Out cross-validation. We considered alternative algorithms such as a gradient boosting in the form of XGBoost (Chen and Guestrin, 2016), but found no benefits. Therefore, here we focus on the Random Forest results.

Segmentation

Since our dataset contains both continuous and categorical data types, we use the Gower similarity coefficient for the distance matrix creation (Gower, 1971). We avoid transforming categories into binary variables as this leads to an information loss (Xu and Wunsch, 2009). For the clustering we use the Partitioning Around Medoids (PAM) algorithm, as suggested by Kaufman and Rousseeuw (2005). In contrast to k-means, k-medoids is not dependent on having squared Euclidean distances and is suitable with the Gower distance (Hastie et al., 2008). Both PAM and Gower similarity coefficient methods are implemented in the cluster package v2.0.7-1 for R (Maechler et al., 2018).

Once the clusters are determined, the second step is to train a classifier that can allocate new pre-launch information that characterises a new product to a cluster. The restricted sample size within clusters makes it difficult to run the proposed regression on clusters. Instead, we directly predict total sales $\hat{Y}_{j,T+h}$ from the median of all video games sales observed at $T + h$ within a cluster.

There are a variety of measures that help in selecting the optimal number of clusters, such as the Gap statistics (Tibshirani et al., 2001) or the Jump method (Sugar and James, 2003). While these measures rely solely on cluster characteristics, our two-stage process has the advantage that it allows us to measure the quality of the clustering directly on the outcome for the target variable. More specifically, we select the cluster with the smallest Mean Squared Error (MSE) on sales using a 10-fold cross validation. We consider up to 30 clusters. We refer to this model as *xClPRB*.

4.4.4 Benchmark models

To assess the predictive value of PRB, we introduce two types of benchmarks that use both *intra-* and *cross-brand* data. The first is based on regression that uses no PRB information and draws upon observed sales and product characteristics (*iRgPCT* & *xRgPCT*). The *iRgPCT* reflects the model that companies would typically base their competitors' forecast when no PRB information is available. The second is a naïve model where we calculate the median of the entire training sample (*xMdSLS* & *iMdSLS*). Despite being trivial to implement, such parsimonious benchmarks are good forecasting practice and often hard to beat (Ord et al., 2017). For convenience we summarise all included features of the different forecasting models in Table 4.2. The first two columns list and describe the input features. The other columns indicate their inclusion into the different forecasting model.

Table 4.2: Overview of features included for different prediction models

		PRB models			Benchmark models	
			ClPRB			
Feature	Input	RgPRB	Clust.	Class.	RgPCT	MdSLS
PRB data based						
PRBadaption	Bass param.	x	x	x	-	-
PRBfpc	FPC scores	x	x	x	-	-
PRBtrend	Trend slope	x	x	x	-	-
PRBvelocity	Time to % adop.	x	x	x	-	-
PRBvolume	Total sum	x	x	x	-	-
Product characteristics data based						
PCTgenre	Genre cat.	x	x	x	x	-
PCTrelease	Release month	x	x	x	x	-
PCTreviews	Review scores	-	x	-	-	-
PCTsequel	Sequel #	x	x	x	x	-
Sales data based						
SLSadoption	Gompertz param.	-	x	-	-	-
SLSsales	Observed sales	-	x	-	-	x
SLSvelocity	Time to % adop.	-	x	-	-	-

4.4.5 Performance evaluation

Our experimental design is based on scenarios (ii) and (iii) with *intra-brand* and *cross-brand* estimation, as illustrated in Figure 4.1. More specifically, we generate individual product forecasts $\hat{Y}_{j,T+h}$ for competitors' title j. We can construct a total of 23 different out-of-sample sets of various sizes. Each out-of-sample is individually predicted by the model introduced in Section 4.3; either based on the full training sample (*cross-brand* case) or as a individual publisher (*intra-brand* case). For the latter, we only consider publishers with more than 6 games, as the estimation becomes very unreliable with even fewer data. This allows to test the *intra-brand* case for 10 publishers, as listed in Table 4.3.

We have a total of 2186 out-of-sample predictions, from different training samples for

Table 4.3: Publishers represented within the training and test-sample

Training & testing		Testing only	
Publisher	# Games	Publisher	# Games
Capcom	16	Level 5	1
Nintendo	18	Codemasters	2
Ubisoft	30	Bethesda Softworks	3
Electronic Arts	61	Eidos Interactive	1
Take-Two Interactive	24	Square Enix	1
Activision	23	Valve	1
Microsoft Game Studios	13	Spike	2
Sega	8	Konami Digital Entertainment	6
Sony	7	MTV Games	1
THQ	14	Deep Silver	1
		Namco Bandai Games	2
		From Software	1
		WB Games	4

each of the previously outlined models. We measure the forecast accuracy using the Geometric Mean Relative Absolute Error (GMRAE):

$$\mathrm{GMRAE}_{i,T+h} = \sqrt[n]{\prod_{i=1}^{n}\left(\frac{\mathrm{AE}_{i,r}}{\mathrm{AE}_{i,b}}\right)},$$

$$\mathrm{AE}_{i,T+h} = \left|\hat{y}_{i,T+h} - y_{i,T+h}\right|,$$

where r is the candidate model that and b is a benchmark forecast, which is the *intra-brand* model without PRB information (iRgPCT), for each series i and horizon $h = \{1, EoL\}$. The GMRAE is an intuitive scale-independent error metric, with errors smaller than 1 meaning that the candidate model outperforms the selected benchmark by (1-GMRAE)*100% (Ord et al., 2017).

4.4.6 Results

Table 4.4 presents the predictive performance of the different forecasting models against the benchmark $iRgPCT$ for the opening week ($h = 1$) and total sales ($h = EoL$) forecasts. The top three rows show the *intra-brand* models, while the bottom 4 rows highlight the *cross-brand* ones. The best performing models for each category are highlighted in bold. In both *intra-brand* and *cross-brand* scenarios, the augmented PRB model perform substantially better than the benchmark.

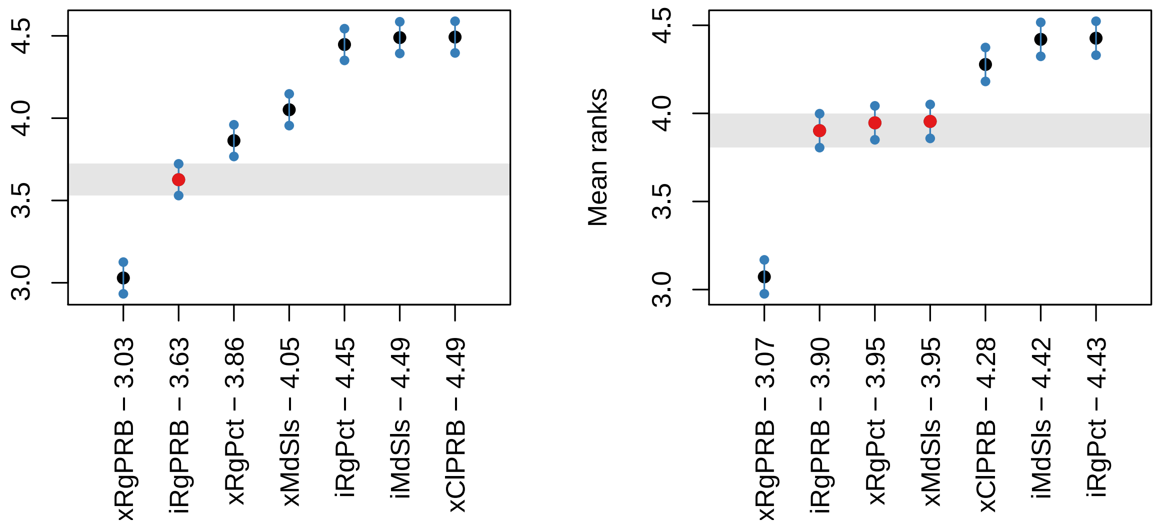
Table 4.4: GMRAE performance of models against iRGPCT (product characteristic only)

Model type			Forecast scenarios	
s	e	v	$h = 1$	$h = EoL$
i	Rg	PRB	**0.797**	**0.812**
i	Rg	PCT	1.000	1.000
i	Md	SLS	1.083	1.026
x	Rg	PRB	**0.631**	**0.624**
x	Cl	PRB	1.189	0.982
x	Rg	PCT	0.893	0.867
x	Md	SLS	0.931	0.869

Column s describes the data sample being intra- (i) or cross-brand (x) based. Column e indicates the estimation method: regression (Rg), segmentation (Cl) and medians (Md). Column v notes the inputs of the model using PRB or only product characteristics (PCT) and sales (SLS).

Between the *intra-brand* models, *iRgPRB* outperforms *iRgPCT* by nearly 20% for both forecasting horizons. *iMdSLS* is outperformed by more than 28% and 22% for the opening and total sales, respectively. This illustrates that PRB contains predictive performance for competitor products compared to solely competitor product characteristics and internal sales data only. To further confirm these findings we test whether differences between models are significant and not due to randomness. We use a non-parametric Friedman and the post-hoc Nemenyi tests as GMRAE errors are non-normally distributed (Hollander et al., 2014) as implemented in tsutils v.0.9.0 package for R (Kourentzes, 2019). The Friedman test results in p-value of 0.000 for both forecast horizon. This indicates that at least one set of results is statistically different. We proceeed with the post-hoc Nemenyi test to identify subgroups results. Figure 4.4 shows the mean ranks of the different models with the whisker being the critical Nemenyi distance of 0.193. A model is considered to be statistically different if there is no overlap between them. We highlight the *iRgPRB* model with a grey bar, indicating that for both forecast horizons it significantly outperforms the *intra-brand* benchmarks and confirms *RQ 1*.

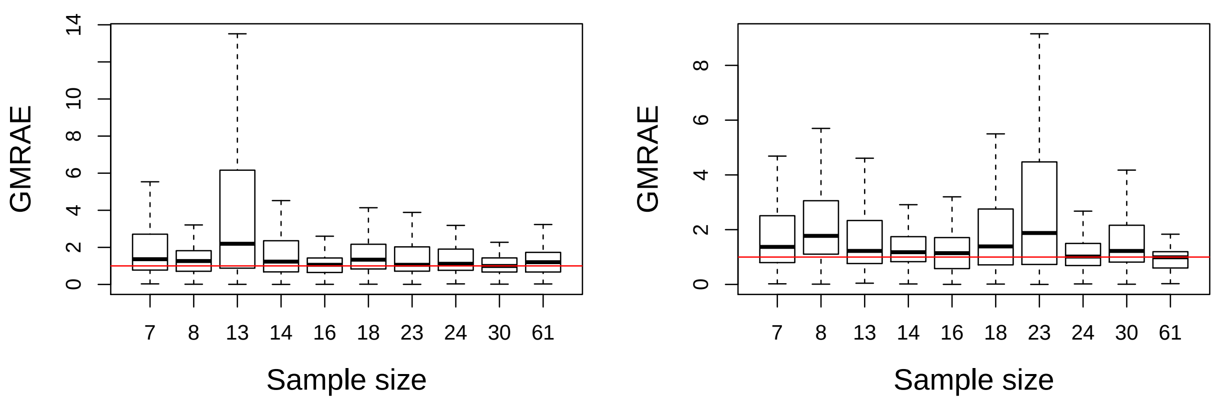
Compared to *cross-brand* data, *iRgPRB* provides better forecast accuracy than the two benchmarks *xRgPCT* and *xMdSLS*, even though the improvements are less impressive com-



(a) Opening sales (b) Total sales

Figure 4.4: Nemenyi test results at 5% significance level

pared to the *intra-brand* benchmarks, with gains of 8.9% and 6.3% for the opening and end-of-life horizon. In the latter case differences are also no longer significantly different. If we compare the *PRB* based models, we see that the *cross-brand* regression model significantly outperforms its *intra-brand* counterpart with more than 18%. It appears that the richer *cross-brand* training can generate substantially more accurate predictions than just *intra-brand PRB*, which confirms *RQ 2*. However, this does not hold for the clustering approach which falls significantly short against both regression *PRB* models. This again supports our *RQ 3*. We attribute this to the relative homogeneity of the video games sector.



(a) Opening sales (b) Total sales

Figure 4.5: Effect of sample size per publisher for iRgPRB against xRgPRB

One obvious question is whether the *intra-brand* sample size affects the predictive performance. We calculate GMRAE to report errors for *iRgPRB* over *xRgPRB* to facilitate the comparison. As Figure 4.5 depicts, there is no visible trend, which would indicate that *intra-brand* errors reduce with larger sample sizes, when compared against the *cross-brand* estimations. We conclude that there is no support for *RQ 4*. The GMRAE of individual publishers are available in the Appendix Table A1.

We proposed including multiple inputs into the regression model and let it flexibly identify the most influential ones. We tested more restricted models, with pre-filtering variables which for some cases, can provide slightly better performance. However, the approach that includes all variables has the overall most stable performance across the different models, and greatly simplifies the modelling process. More details are available in Appendix Table A2.

4.5 Discussion

We set out four research questions to test the suitability of PRB for predicting the success of competitors' new products. We proposed a model able to generate sales forecasts and empirically test its performance on a set of global video games sales. We find support for *RQ 1*, that PRB significantly improves competitors' new product predictions with internal sales data over models that use only internal sales and product characteristics. In doing so, the predictive performance of PRB was found to be independent of the sample size, and we argue against *RQ 4*, that a larger *intra-brand* training sample lead to smaller forecast errors. The literature review in Section 4.2 has highlighted that to date, very little research is concerned with incorporating competitor user-generated content in a predictive context. Our findings support the value that competitor related online content can have, together with internal data.

Beyond the proposed competitor PRB model, we looked at an alternative that can consider the potential heterogeneity in the market using segmentation. Our findings indicated that the video games market is fairly homogeneous, and although this model could outperform benchmarks, it was found to be inferior to the regression based model. In that, this supports *RQ 3*, that separating by brands is better than product segmentation in the case of video games sales.

Arguably, our benchmark model *iRgPCT*, is limited by not considering all possible product categories. We argue that any *cross-brand* dataset is based on information that is not readily available to all organisations. With that in mind, the performance against *xRgPCT* is useful to assess whether PRB can overcome those restrictions, but is otherwise not practical. We found that using PRB outperforms or matches the performance on opening and total sales, respectively, demonstrating the efficacy of the proposed model. All models, however, fall short against the *xRgPRB* model. This is expected as it can draw from a much richer dataset than the *intra-brand* models and confirms *RQ 2*. Nonetheless, the practical usefulness of *xRgPRB* is questionable.

It is worth noting that the forecasting performance of *iRgPRB* is similar for the two different forecast horizons. Similar observations have been reported by Schaer et al. (2019b), where PRB not only provides value for predictions close to release but also contains predictive information for the long-term forecast.

4.5.1 Managerial implications

Our study shows that it is possible to not only gain valuable insights from PRB for own sales product launches, but can also provide valuable insights for new competitor launches. This provides decision-makers with crucial information to better plan their advertising activities to counter-act or support their own sales. As such, it might also provide insights when planning an own product release to monitor competing products that might impact the release success.

PRB such as search traffic information is freely available and can readily be implemented into the predictive framework, as outlined in this research. One of the main benefits is that PRB adds value even when the own product sample size is relatively small. Moreover, PRB also includes information for the near-term open sales but also provides insights on the strategic horizon by explaining some of the overall market potential. In the video game industry, such a PRB signal usually becomes available well in advance, and the forecasting model can be updated once new information arrives (see Schaer et al., 2019b; Xiong and

Bharadwaj, 2014).

4.5.2 Model extensions

In this study, we focused on measuring the future market potential of competitors' product and provide initial insights on the suitability of PRB. A natural next step is to formulate a framework that also measures the competitive impact on their products. Search traffic information has for example been incorporated into market response models that measure impact on sales (e.g. Du et al., 2015; Du and Kamakura, 2012) or advertising (e.g. Hu et al., 2014). An extension into this area of new products would have two potential implications. First, to improve the marketing-mix (for an application without PRB see Luan and Sudhir, 2010) and second, to better manage product release timing. For example, postponing a release is common practice in the movie industry (Einav, 2010), but there is limited work to help identify when the loss in revenue due to competition launches outweights the diminished stock value (Einav and Ravid, 2009) and brand trust (Herm, 2013).

Our modelling approach is based on a set of well-established methods to reduce the dimensionality of PRB. While our results are in favour of a data-driven variable selection, there is room to experiment further with different time series characterisation (e.g. Lubba et al., 2019). It is also possible to expand our framework to include other types of PRB sources, in particular, those involving unstructured data (see Balducci and Marinova, 2018). Although we found little value in clustering video games sales the model might prove valuable on different datasets.

4.6 Conclusions

Although, research on Competitive Intelligence is well established and frequently draws from user-generated content, there is limited focus on new products. However, the market trend to shorter-life cycles leads to increased competition on product launch (Calantone et al., 2010). Our study contributes to this area of research by providing insights on the predictive value of publicly available PRB for competitor products. While most of the literature evaluates the proposed approaches on randomised *cross-brand* datasets, we limit the training of our forecasting model to *internal-brand* information and competitor PRB. This is both harder and more realistic. The results suggest that PRB provides valuable insights for companies to better understand their competitive standing.

In comparison to other market intelligence, PRB is an inexpensive source of information that reflects consumer interest (Houston et al., 2018). Moreover, it has the advantage that is available over time, which also allows capturing its dynamic (Xiong and Bharadwaj, 2014). We show that in a relatively homogeneous market setting, only a few internal sales observations are adequate to produce valuable insights for the competitors' market potential. This information can be taken into consideration when managing the marketing-mix and support new product release planning.

A Additional results

Table A1: GMRAE performance of iRgPRB against various benchmarks at publishers level

	Opening Sales			Total Sales		
Publisher	iRgPct	xRgPct	xRgPRB	iRgPct	xRgPct	xRgPRB
Activision (23)	0.602	0.815	1.178	0.723	1.344	1.854
Capcom (16)	0.772	0.779	1.077	0.773	0.717	0.999
Electronic Arts (61)	0.809	0.761	1.105	0.610	0.652	0.947
Microsoft Game Studios (13)	0.701	1.651	2.417	0.774	0.971	1.318
Nintendo (18)	1.066	0.945	1.296	0.845	1.066	1.387
Sega (8)	0.996	0.807	1.189	1.094	1.111	1.600
Sony Computer Ent. (7)	0.788	1.051	1.484	0.827	1.023	1.420
Take-Two Interactive (24)	0.561	0.785	1.182	0.651	0.797	1.071
THQ (14)	0.999	0.901	1.282	0.871	0.788	1.129
Ubisoft (30)	0.894	0.724	1.038	0.923	0.944	1.360

Table A2: GMRAE comparison of different feature models over iRgPCT (all)

	Intra-brand						Cross-brands							
	iRgPRB		iRgPCT		iMdSLS		xRgPRB		xClPRB		xRgPCT		xMdSLS	
Features	$h = 1$	EoL	$h = 1$	EoL	$h = 1$	EoL	$h = 1$	EoL	$h = 1$	EoL	$h = 1$	EoL	$h = 1$	EoL
All variables	0.797	**0.812**					0.631	0.624	1.189	0.982				
PRBvolume	0.852	0.887					0.609	**0.619**	1.326	0.957				
PRBvelocity	0.934	0.984					0.692	0.861	1.308	1.003				
PRBtrend	0.906	0.932	1.000	1.000	1.083	1.026	0.769	0.759	1.076	0.927	0.893	0.867	0.931	0.869
PRBadoption	0.909	0.927					0.687	0.703	1.182	0.907				
PRBfpc	0.828	0.863					0.615	0.701	**0.961**	**0.807**				
PRBvol. & PRBfpc	**0.793**	0.835					**0.596**	0.662	1.117	1.043				

Best performing model per column in bold

Chapter 5

Discussion and conclusions

The fast-paced business world with changing consumer preferences increases the volatility of product demand and shortens own and competitors' life-cycle. Accurate forecasts become vital for organisations to cope with those challenges and better align their supply chain and marketing activities. Online buzz is an attractive data source with substantial research and practice potential. This thesis contributes to the fast-growing research on incorporating online buzz into predictive models. While the majority of the literature is very positive about the predictive ability of online buzz, our findings are mixed, demonstrating both its potential, but also limitations of its usefulness.

We identified 3 main research questions: (i) how can we use post-release buzz to provide short-term forecasts in the context of operational decision making? (ii) can pre-release buzz help predict the market potential of new products? (iii) can we use pre-release buzz to predict competitors' new product success?

In Chapter 2, we investigated whether search traffic and social network shares are helpful in improving demand forecasting with post-release buzz. We first highlight the limitations on the existing literature regarding their experimental design, both from a statistical and practical point of view. We then attempt to support the key findings of the literature investigating its usefulness in two forecasting case studies. Those findings are in stark contrast to the literature, and we find that established univariate forecasting benchmarks, such as exponential smoothing, consistently perform better than models using online information. We also did not find substantial differences regarding its predictive power in different phases of the life-cycle. Crucially, it is to be severely limited in terms of operational decision making benefits. Business decisions have lead times that any useful predictive variable must follow.

We find that this information has little if any predictive signal left for common businesses' lead times. A contributing factor to this weak signal is the instant buying decision and noise that surrounds this data with pre- and post-purchase online activity (see Houston et al., 2018). Ferrara and Simoni (2019) identify this issue as well, finding the Google Trends signal to no longer providing forecast improvements when additional data sources became available to nowcast GDP.

In Chapter 2 we stress that using pre-release buzz can be a more fruitful as the buzz signal for an unreleased product is anticipatory, as previously confirmed in various studies (e.g. Kim et al., 2015; Xiong and Bharadwaj, 2014; Kulkarni et al., 2012). However, most of these studies only consider opening sales. In Chapter 3 we show that pre-release buzz is not only adding value to predict early week sales, but also in explaining the entire market potential of new products. We propose a model framework that uses the adoption information of previous product generations and pre-release buzz. Our findings support that pre-release buzz affects the long-term sales, similarly to pre-release advertising and publicity (e.g. Burmester et al., 2015; Elberse and Anand, 2007). This finding is expected, given the close dependency of marketing efforts and pre-release buzz (Kim and Hanssens, 2017), but our findings highlight that the benefits from measuring pre-release buzz reach beyond the first week. We find that benefits for predicting the total life-cycle sales occur from about 9 weeks prior to release and onwards. For the opening week, the predictively valuable period is more than 26 weeks prior to release (confirming findings of Xiong and Bharadwaj, 2014). Furthermore, we find that the relationship between pre-release buzz and the market potential is multiplicative.

One of the key strength of the proposed model in Chapter 3 is that it only requires a few observations of previous product generations to generate valuable forecasts for new product. To the best of our knowledge, our study is also the only one which uses pre-release buzz on company internal data. As Chapter 4 discusses in more detail, the richer dataset used in the literature, mixing brand data from multiple companies, typically provides better estimates but, does not reflect the real-world environment. This challenges the reported value of these estimates. Motivated by the findings in Chapter 3, we transfer the idea of forecasting new products with pre-release towards competitive intelligence in attempting to forecast competitors' products. In Chapter 4, we find that pre-release buzz is also able to significantly improve predictions for competitors' new product predictions based on own internal sales data, over models that use only internal sales and product characteristics. This expands the literature on Competitive Intelligence with buzz into a predictive context. A further insight is that the predictive ability does not seem to depend on sample size.

5.1 Managerial implications

In times of intense competition, companies are keen to improve their predictive ability and previous research has hyped the usefulness of online buzz. Our research provides a critical reflection on these developments. In a nutshell, our findings suggest that online buzz is helpful in the pre-release case, but loses predictive strength in the post-release phase. In particular, when decision-makers have sufficient historical sales and other indicators the benefit from post-release online buzz is likely to be marginal (see Ferrara and Simoni, 2019; Schaer et al., 2019a; Goel et al., 2010). In practice, the usefulness will highly depend on product type and the business environment. Markets with long lead times are unlikely to see benefits, while noise around post-release buzz diminish its predictive ability.

On the other hand, for pre-release buzz, we find accuracy improvements for both own and competitors' products of over 18% compared to benchmarks. For own products this allows better planning of stock, procurement and marketing activities, while insights on the competing products help to adjust the marketing-mix and alter own product launch dates, if necessary. For own products, we suggest to increase focus on pre-release advertisement to stimulate pre-release buzz (Kim and Hanssens, 2017; Burmester et al., 2015). Our suggested approach and findings also facilitate forecasting life-cycle sales. This allows companies to cover the initial launch period with a pre-release forecast, until sufficient data is available to estimate conventional time-series models.

However, we underline the need for adequate benchmarking and thorough forecast evaluation. In both, Chapter 3 & 4, we introduced a forecasting framework that shows that pre-release buzz can be helpful even with very limited data points that reflect reality. We also provide in Chapter 2 insights on some practical challenges that can occur when collecting online buzz and point out helpful resources in the literature to mitigate those. An important point for practice is that, although, online buzz is free and easy to collect, there might a significant cost and skill involved in getting the right variables. To facilitate the implementation of our findings in practice, we developed two open-source packages for R: the *diffusion package* contains functions to estimate and forecast from a variety of diffusion curves (Schaer and Kourentzes, 2018) and the *GTT package* provides tools to handle Google Trends data better (Schaer, 2018).

The suggested approaches and resources provided in this thesis bring new analytical capabilities to organisations, but in return require them to embrace data driven models. They require investing to equip teams with the required know-how and access to data sources to implement these innovations into a productive environment.

5.2 Limitations and future work

The thesis attempted to answer some lay research questions on the use of pre- and post-release online buzz for forecasting purposes. In our view, although we provide evidence to address our questions, the thesis opens multiple questions in all modelling, marketing and decision making. In the previous chapters, we concluded with extensions stemming from addressing each specific research question. Here we return to some major ones and introduce research directions that follow the thesis as a whole.

One aspect that is omitted in the discussion by the literature around online buzz is its impact on forecast uncertainty as reflected in prediction intervals. Forecasts with increased certainty should correspond to decision-making advantages. Recent work by Trapero et al. (2019) highlighted the importance of variability, asymmetry and auto-correlation of the forecasts on the determination of the appropriate safety stock level. In other words, even if the mean forecast of online buzz does not improve over simple benchmarks, the prediction intervals might be narrower and may be beneficial for the decisions supported by the forecasts, such as inventory optimisation and can be translated into a prescriptive analytic solution (Bertsimas and Kallus, 2019).

Most of our research has focused on video games, which together with box office, are very well studied sectors (Hennig-Thurau and Houston, 2019). In particular, for pre-release buzz, it is essential to test our proposed approach to other markets with different adoption characteristics, but also to understand the business challenges better, to improve their practical applicability.

Kahn and Chase (2018) find that a majority of new product forecasts are heavily judgmental. Most of the literature that focuses on judgmental adjustment and model selection is on post-release scenarios (e.g., Fildes et al., 2019; Petropoulos et al., 2018). We know little about how decision-makers interpret online information and its uncertainties. Companies might, for example, include information from an online search or product reviews to support the judgmental decision making (Lee et al., 2007; Goodwin et al., 2007). This needs to be investigated further, especially given the current hype about online buzz

Chapter 2 highlighted various issues of online data, including potential sample bias, dependency on large internet companies and vulnerability to manipulation from a third-party. This does not only apply to the interpretation of statistical models that directly include online buzz but could also be extended to support decisions. Moreover, the results presented in Chapter 3 shed some initial light on the potential underestimation bias of pre-release buzz and the merit that forecast combinations with the optimistic human judgment can have.

While this thesis focused on structured data, we believe that there are substantial benefits to further explore non-structured user-generated content (Balducci and Marinova, 2018). Over recent years, deep learning has evolved substantially and now offers very good performance in domains such as image recognition (LeCun et al., 2015). Recent research by Timoshenko and Hauser (2019) illustrates that online reviews can deliver the same accuracy in identifying customer needs, compared to traditional focus groups. This showcases the cost-saving potential these new algorithms can have in automating previously labour-intensive tasks that are difficult to scale. We see particular applications around user-generated online videos such as to automatically measure their emotions (e.g., Teixeira et al., 2012) and use those as richer buzz inputs to our predictive models. This would also directly feedback into the initial research proposal for this thesis that suggested to investigate the adoption of online videos (cf. Preface). The research journey demonstrated that there were more pertinent research questions to answer prior to embarking in those initial questions. Having provided some responses to these questions, we see this thesis as the starting point to explore a multitude of research questions. Some of which could be seen since the beginning of this works, many of which emerged during these years. The latter is the result of the works presented in this thesis and others stem from progress in both the literature and practice, which relates well with our findings. The use of online buzz in its aggregate form, as done here, or in richer forms, perhaps leveraged by Artificial Intelligence, remains an exciting field of research. The work we have presented in this thesis will support this future work by providing some necessary insights.

Bibliography

Abisheva, A., Garimella, V. R. K., Garcia, D., Weber, I., 2014. Who watches (and shares) what on youtube? and when?: Using twitter to understand youtube viewership. In: Proceedings of the 7th ACM International Conference on Web Search and Data Mining. WSDM '14. ACM, New York, pp. 593–602.

Abrahams, A. S., Jiao, J., Fan, W., Wang, G. A., Zhang, Z., 2013. What's buzzing in the blizzard of buzz? Automotive component isolation in social media postings. Decision Support Systems 55 (4), 871 – 882.

Acimovic, J., Erize, F., Hu, K., Thomas, D. J., Van Mieghem, J. A., 2019. Product life cycle data set: Raw and cleaned data of weekly orders for personal computers. Manufacturing & Service Operations Management 21 (1), 171–176.

Aral, S., 2014. The problem with online ratings. MIT Sloan Management Review 55 (2), 47–52.

Araz, O. M., Bentley, D., Muelleman, R. L., 2014. Using google flu trends data in forecasting influenza-likeillness related ed visits in omaha, nebraska. The American Journal of Emergency Medicine 32 (9), 1016 – 1023.

Armstrong, J., Collopy, F., 1992. Error measures for generalizing about forecasting methods: Empirical comparisons. International Journal of Forecasting 8 (1), 69 – 80.

Armstrong, J. S., 2006. Findings from evidence-based forecasting: Methods for reducing forecast error. International Journal of Forecasting 22 (3), 583 – 598.

Asur, S., Huberman, B., Aug 2010. Predicting the future with social media. In: Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on. Vol. 1. pp. 492–499.

Baardman, L., Levin, I., Perakis, G., Singhvi, D., 2017. Leveraging comparables for new product sales forecasting. Working Paper, 1–36.
URL https://ssrn.com/abstract=2832004

Babic Rosario, A., Sotgiu, F., De Valck, K., Bijmolt, T. H., 2016. The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. Journal of Marketing Research 53 (3), 297–318.

Baccianella, S., Esuli, A., Sebastiani, F., 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In: LREC. Vol. 10. pp. 2200–2204.

Balducci, B., Marinova, D., 2018. Unstructured data in marketing. Journal of the Academy of Marketing Science 46 (4), 557–590.

Bangwayo-Skeete, P. F., Skeete, R. W., 2015. Can Google data improve the forecasting performance of tourist arrivals? mixed-data sampling approach. Tourism Management 46, 454 – 464.

Barash, V., Ducheneaut, N., Isaacs, E., Bellotti, V., 2010. Faceplant: Impression (mis) management in facebook status updates. In: ICWSM. pp. 207–210.

Barreira, N., Godinho, P., Melo, P., Nov 2013. Nowcasting unemployment rate and new car sales in south-western europe with google trends. NETNOMICS: Economic Research and Electronic Networking 14 (3), 129–165.

Basallo-Triana, M. J., Rodrguez-Sarasty, J. A., Benitez-Restrepo, H. D., 2017. Analogue-based demand forecasting of short life-cycle products: A regression approach and a comprehensive assessment. International Journal of Production Research 55 (8), 2336–2350.

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

Bass, F. M., Gordon, K., Ferguson, T. L., Githens, M. L., 2001. Directv: Forecasting diffusion of a new technology prior to product launch. Interfaces 31 (3), S82–S93.

Bass, P. I., Bass, F. M., 2001. Diffusion of technology generations: A model of adoption and repeat sales. Working Paper, 1–29, accessed 30/01/19.
URL https://goo.gl/C8kpaB

Belvedere, V., Goodwin, P., 2017. The influence of product involvement and emotion on short-term product demand forecasting. International Journal of Forecasting 33 (3), 652 – 661.

Bemmaor, A. C., 1994. Modeling the diffusion of new durable goods: Word-of-mouth effect versus consumer heterogeneity. In: Laurent, G., Lilien, G. L., Pras, B. (Eds.), Research traditions in marketing. Springer, Dordrecht, pp. 201–229.

Bemmaor, A. C., Lee, J., 2002. The impact of heterogeneity and ill-conditioning on diffusion model parameter estimates. Marketing Science 21 (2), 209–220.

Bemmaor, A. C., Zheng, L., 2018. The diffusion of mobile social networking: Further study. International Journal of Forecasting 34 (4), 612 – 621.

Bertsimas, D., Kallus, N., 2019. From predictive to prescriptive analytics. Management Science, 1–20, pre-print.

Bhatnagar, A., Ghose, S., 2004. Online information search termination patterns across product categories and consumer demographics. Journal of Retailing 80 (3), 221 – 228.

Bijl, L., Kringhaug, G., Molnár, P., Sandvik, E., 2016. Google searches and stock returns. International Review of Financial Analysis 45, 150 – 156.

Blattberg, R. C., Hoch, S. J., 1990. Database models and managerial intuition: 50% model + 50% manager. Management Science 36 (8), 887–899.

Bollen, J., Mao, H., Zeng, X., 2011. Twitter mood predicts the stock market. Journal of Computational Science 2 (1), 1 – 8.

Boone, T., Ganeshan, R., Hicks, R. L., 2015. Incorporating Google Trends data into sales forecast. International Journal of Applied Forecasting 38, 9–14.

Boone, T., Ganeshan, R., Hicks, R. L., Sanders, N. R., 2018. Can google trends improve your sales forecast? Production and Operations Management 27 (10), 1770–1774.

Boone, T., Ganeshan, R., Jain, A., Sanders, N. R., 2019. Forecasting sales in the supply chain: Consumer analytics in the big data era. International Journal of Forecasting 35 (1), 170 – 180.

Boylan, J. E., Goodwin, P., Mohammadipour, M., Syntetos, A. A., 2015. Reproducibility in forecasting research. International Journal of Forecasting 31 (1), 79 – 90.

Breiman, L., 2001. Random forests. Machine Learning 45 (1), 5–32.

Broxton, T., Interian, Y., Vaver, J., Wattenhofer, M., 2013. Catching a viral video. Journal of Intelligent Information Systems 40 (2), 241–259.

Brynjolfsson, E., Geva, T., Reichman, S., 2016. Crowd-squared: Amplifying the predictive power of search trend data. MIS Quarterly 40 (4), 941–962.

Bughin, J., 2015. Google searches and twitter mood: nowcasting telecom sales performance. NETNOMICS: Economic Research and Electronic Networking 16 (1), 87–105.

Bulut, L., 2017. Google trends and the forecasting performance of exchange rate models. Journal of Forecasting 37 (3), 303–315.

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

Butler, D., 2013. When google got flu wrong. Nature 494 (7436), 155.

Calantone, R. J., Yeniyurt, S., Townsend, J. D., Schmidt, J. B., 2010. The effects of competition in short product life-cycle markets: The case of motion pictures. Journal of Product Innovation Management 27 (3), 349–361.

Calof, J. L., Wright, S., 2008. Competitive intelligence: A practitioner, academic and interdisciplinary perspective. European Journal of marketing 42 (7/8), 717–730.

Carrière-Swallow, Y., Labbé, F., 2013. Nowcasting with Google Trends in an emerging market. Journal of Forecasting 32 (4), 289–298.

Chen, J., Chen, H., Wu, Z., Hu, D., Pan, J. Z., 2017. Forecasting smog-related health hazard based on social media and physical sensor. Information Systems 64 (Supplement C), 281 – 291.

Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system. In: Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '16. ACM, New York, NY, USA, pp. 785–794.

Chintagunta, P. K., Lee, J., 2012. A pre-diffusion growth model of intentions and purchase. Journal of the Academy of Marketing Science 40 (1), 137–154.

Choi, H., Varian, H., 2009. Predicting the present with Google Trends. http://ssrn.com/abstract=1659302, online; accessed 26/08/2019.

Choi, H., Varian, H., 2012. Predicting the present with Google Trends. Economic Record 88, 2–9.

Claeskens, G., Magnus, J. R., Vasnev, A. L., Wang, W., 2016. The forecast combination puzzle: A simple theoretical explanation. International Journal of Forecasting 32 (3), 754 – 762.

Clement, M., Fabel, S., Schmidt-Stolting, C., 2006. Diffusion of hedonic goods: A literature review. International Journal on Media Management 8 (4), 155–163.

Coldwell, W., 2018. Wish you weren't here: how the tourist boom and selfies are threatening Britain's beauty spots. The Guardian, online; accessed 27/08/2019.
URL https://bit.ly/2w8mmp6

Conlon, S., 2018. The It dress that shows how to promote brands and influence people. The Guardian, online; accessed 27/08/2019.
URL https://bit.ly/2HC509s

Craig, C. S., Greene, W. H., Versaci, A., 2015. E-word of mouth: Early predictor of audience engagement. Journal of Advertising Research 55 (1), 62–72.

Crane, R., Sornette, D., 2008. Robust dynamic classes revealed by measuring the response function of a social system. Proceedings of the National Academy of Sciences 105 (41), 15649–15653.

Cui, R., Gallino, S., Moreno, A., Zhang, D. J., 2018. The operational value of social media information. Production and Operations Management 27 (10), 1749–1769.

Da, Z., Engelberg, J., Gao, P., 2011. In search of attention. The Journal of Finance 66 (5), 1461–1499.

D'Amuri, F., Marcucci, J., 2017. The predictive power of google searches in forecasting us unemployment. International Journal of Forecasting 33 (4), 801 – 816.

Davydenko, A., Fildes, R., 2013. Measuring forecasting accuracy: The case of judgmental adjustments to SKU-level demand forecasts. International Journal of Forecasting 29 (3), 510 – 522.

Dellarocas, C., 2003. The digitization of word of mouth: Promise and challenges of online feedback mechanisms. Management Science 49 (10), 1407–1424.

Dellarocas, C., Zhang, X. M., Awad, N. F., 2007. Exploring the value of online product reviews in forecasting sales: The case of motion pictures. Journal of Interactive Marketing 21 (4), 23 – 45.

Demšar, J., 2006. Statistical comparisons of classifiers over multiple data sets. Journal of Machine learning research 7, 1–30.

Deusser, C., Jansen, N., Reubold, J., Schiller, B., Hinz, O., Strufe, T., 2018. Buzz in social media: Detection of short-lived viral phenomena. In: Companion Proceedings of the The Web Conference 2018. WWW '18. pp. 1443–1449.

Dhar, V., Chang, E. A., 2009. Does chatter matter? The impact of user-generated content on music sales. Journal of Interactive Marketing 23 (4), 300 – 307.

Dimpfl, T., Jank, S., 2016. Can internet search queries help to predict stock market volatility? European Financial Management 22 (2), 171–192.

Ding, C., Cheng, H. K., Duan, Y., Jin, Y., 2017. The power of the like button: The impact of social media on box office. Decision Support Systems 94, 77 – 84.

Divakaran, P. K. P., Palmer, A., Sndergaard, H. A., Matkovskyy, R., 2017. Pre-launch prediction of market performance for short lifecycle products using online community data. Journal of Interactive Marketing 38, 12 – 28.

Du, R. Y., Hu, Y., Damangir, S., 2015. Leveraging trends in online searches for product features in market response modeling. Journal of Marketing 79 (1), 29–43.

Du, R. Y., Kamakura, W. A., 2012. Quantitative trendspotting. Journal of Marketing Research 49 (4), 514–536.

Einav, L., 2010. Not all rivals look alike: Estimating an equilibrium model of the release date timing game. Economic Inquiry 48 (2), 369–390.

Einav, L., Ravid, S. A., 2009. Stock market response to changes in movies' opening dates. Journal of Cultural Economics 33 (4), 311–319.

Elberse, A., Anand, B., 2007. The effectiveness of pre-release advertising for motion pictures: An empirical investigation using a simulated market. Information Economics and Policy 19 (3), 319 – 343.

Eliashberg, J., Jonker, J.-J., Sawhney, M. S., Wierenga, B., 2000. MOVIEMOD: An implementable decision-support system for prerelease market evaluation of motion pictures. Marketing Science 19 (3), 226–243.

Elshendy, M., Colladon, A. F., Battistoni, E., Gloor, P. A., 2018. Using four different online media sources to forecast the crude oil price. Journal of Information Science 44 (3), 408–421.

Erevelles, S., Fukawa, N., Swayne, L., 2016. Big data consumer analytics and the transformation of marketing. Journal of Business Research 69 (2), 897 – 904.

ESA, 2018. Essential facts about the computer and video Game industry.
URL https://goo.gl/Kis3BT

Fan-Osuala, O., Zantedeschi, D., Jank, W., 2018. Using past contribution patterns to forecast fundraising outcomes in crowdfunding. International Journal of Forecasting 34 (1), 30 – 44.

Fantazzini, D., Toktamysova, Z., 2015. Forecasting German car sales using Google data and multivariate models. International Journal of Production Economics 170, Part A, 97 – 135.

Fast, S. M., Kim, L., Cohn, E. L., Mekaru, S. R., Brownstein, J. S., Markuzon, N., 2018. Predicting social response to infectious disease outbreaks from internet-based news streams. Annals of Operations Research 263 (1), 551–564.

Fehringer, D., Hohhof, B., Johnson, T., 2006. State of the art: Competitive intelligence. Tech. rep., Competitive Intelligence Foundation, Alexandria VA.

Ferrara, L., Simoni, A., 2019. When are google data useful to nowcast gdp? an approach via pre-selection and shrinkage. Banque de France Working Paper 717, 1–30.
URL https://ssrn.com/abstract=3370917

Fildes, R., Goodwin, P., Önkal, D., 2019. Use and misuse of information in supply chain forecasting of promotion effects. International Journal of Forecasting 35 (1), 144 – 156.

Fildes, R., Ord, J. K., 2002. Forecasting competitions: Their role in improving forecasting practice and research. In: Clements, M. P., Henry, D. F. (Eds.), A Companion to Economic Forecasting. Blackwell Publishing, Malden, pp. 322–353.

Foutz, N. Z., Jank, W., 2010. Research note–prerelease demand forecasting for motion pictures using functional shape analysis of virtual stock markets. Marketing Science 29 (3), 568–579.

Friedman, J., Hastie, T., Simon, N., Tibshirani, R., 2016. glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models. R package version 2.0-5.
URL https://cran.r-project.org/web/packages/glmnet/index.html

Fulcher, B. D., 2018. Feature-based time-series analysis. In: Dong, G., Liu, H. (Eds.), Feature Engineering for Machine Learning and Data Analytics. CRC press, Boca Raton, pp. 87–143.

Gelper, S., Peres, R., Eliashberg, J., 2015. Pre-release word-of-mouth dynamics: The role of spikes. Working Paper, 1–40, accessed 30/01/19.
URL https://goo.gl/csebdR

Genre, V., Kenny, G., Meyler, A., Timmermann, A., 2013. Combining expert forecasts: Can anything beat the simple average? International Journal of Forecasting 29 (1), 108 – 121.

Geva, T., Oestreicher-Singer, G., Efron, N., Shimshoni, Y., 2017. Using forum and search data for sales prediction of high/involvement products. MIS Quarterly 41 (1), 65–82.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., Brilliant, L., 2009. Detecting influenza epidemics using search engine query data. Nature 457 (7232), 1012–1014.

Godes, D., Silva, J. C., 2012. Sequential and temporal dynamics of online opinion. Marketing Science 31 (3), 448–473.

Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., Watts, D. J., 2010. Predicting consumer behavior with web search. Proceedings of the National Academy of Sciences 107 (41), 17486–17490.

Goia, A., May, C., Fusai, G., 2010. Functional clustering and linear regression for peak load forecasting. International Journal of Forecasting 26 (4), 700 – 711.

Goodwin, P., Dyussekeneva, K., Meeran, S., 2013. The use of analogies in forecasting the annual sales of new electronics products. IMA Journal of Management Mathematics 24 (4), 407.

Goodwin, P., Fildes, R., Lawrence, M., Nikolopoulos, K., 2007. The process of using a forecasting support system. International Journal of Forecasting 23 (3), 391 – 404.

Goodwin, P., Meeran, S., Dyussekeneva, K., 2014. The challenges of pre-launch forecasting of adoption time series for new durable products. International Journal of Forecasting 30 (4), 1082 – 1097.

Gopinath, S., Chintagunta, P. K., Venkataraman, S., 2013. Blogs, advertising, and local-market movie box office performance. Management Science 59 (12), 2635–2654.

Gower, J. C., 1971. A general coefficient of similarity and some of its properties. Biometrics 27 (4), 857–871.

Green, K. C., Armstrong, J. S., 2015. Simple versus complex forecasting: The evidence. Journal of Business Research 68 (8), 1678 – 1685, special Issue on Simple Versus Complex Forecasting.

Hamid, A., Heiden, M., 2015. Forecasting volatility with empirical similarity and google trends. Journal of Economic Behavior & Organization 117, 62 – 81.

Hand, C., Judge, G., 2012. Searching for the picture: forecasting UK cinema admissions using Google Trends data. Applied Economics Letters 19 (11), 1051–1055.

Hann, I.-H., Oh, J., James, G., 2011. Forecasting the sales of music albums: A functional data analysis of demand and supply side p2p data. Working Paper, 1–31, accessed 30/01/2019.
URL https://goo.gl/goWrKN

Hastie, T., Tibshirani, R., Jerome, 2008. The elements of statistical learning. Data mining, inference, and prediction, 2nd Edition. Springer, New York.

Hastie, T., Tibshirani, R., Wainwright, M., 2015. Statistical Learning with Sparsity. The Lasso and Generalizations. Monographs on Statistics and Applied Probability 143. CRC Press, Boca Raton.

He, W., Wu, H., Yan, G., Akula, V., Shen, J., 2015. A novel social media competitive analytics framework with sentiment benchmarks. Information & Management 52 (7), 801 – 812.

He, W., Zha, S., Li, L., 2013. Social media competitive analysis and text mining: A case study in the pizza industry. International Journal of Information Management 33 (3), 464 – 472.

Heinonen, K., 2011. Consumer activity in social media: Managerial approaches to consumers' social media behavior. Journal of Consumer Behaviour 10 (6), 356–364.

Hennig-Thurau, T., Houston, M. B., 2019. Entertainment science. Data analytics and practical theory for movies, games, books, and music. Springer, Switzerland.

Hennig-Thurau, T., Houston, M. B., Heitjans, T., 2009. Conceptualizing and measuring the monetary value of brand extensions: The case of motion pictures. Journal of Marketing 73 (6), 167–183.

Herm, S., 2013. When things go wrong, don't rely on committed consumers: Effects of delayed product launches on brand trust. Journal of Product Innovation Management 30 (1), 70–81.

Hewett, K., Rand, W., Rust, R. T., van Heerde, H. J., 2016. Brand buzz in the echoverse. Journal of Marketing 80 (3), 1–24.

Ho, C.-S., Damien, P., Gu, B., Konana, P., 2017. The time-varying nature of social media sentiments in modeling stock returns. Decision Support Systems 101, 69 – 81.

Hollander, M., Wolfe, D. A., Chicken, E., 2014. Nonparametric Statistical Methods, 3rd Edition. Wiley, New Jersey.

Hou, F., Li, B., Chong, A. Y.-L., Yannopoulou, N., Liu, M. J., 2017. Understanding and predicting what influence online product sales? a neural network approach. Production Planning & Control 28 (11-12), 964–975.

Houston, M. B., Kupfer, A.-K., Hennig-Thurau, T., Spann, M., 2018. Pre-release consumer buzz. Journal of the Academy of Marketing Science 46 (2), 338–360.

Hu, K., Acimovic, J., Erize, F., Thomas, D. J., Van Mieghem, J. A., 2019. Forecasting new product life cycle curves: Practical approach and empirical analysis. Manufacturing & Service Operations Management 21 (1), 66–85.

Hu, Y., Du, R. Y., Damangir, S., 2014. Decomposing the impact of advertising: Augmenting sales with online search data. Journal of Marketing Research 51 (3), 300–319.

Huang, X., Zhang, L., Ding, Y., 2017. The baidu index: Uses in predicting tourism flows a case study of the forbidden city. Tourism Management 58, 301 – 306.

Huberty, M., 2015. Can we vote with our tweet? on the perennial difficulty of election forecasting with social media. International Journal of Forecasting 31 (3), 992 – 1007.

Hyndman, R. J., 2016. forecast: Forecasting functions for time series and linear models. R package version 7.2.
URL http://github.com/robjhyndman/forecast

Hyndman, R. J., Khandakar, Y., 7 2008. Automatic time series forecasting: The forecast package for R. Journal of Statistical Software 27 (3), 1–22.

Hyndman, R. J., Koehler, A. B., Ord, J. K., Snyder, R. D., 2008. Forecasting with Exponential Smoothing. The State Space Approach. Springer Science & Business Media, Berlin.

Johansson, E., 2014. Creating daily search volume data from weekly and daily data. Online; accessed 14/12/2017.
URL https://bit.ly/34dlQXl

Jun, S.-P., Sung, T.-E., Park, H.-W., 2017. Forecasting by analogy using the web search traffic. Technological Forecasting and Social Change 115, 37 – 51.

Kahn, K. B., Chase, C. W., 2018. The state of new-product forecasting. Foresight: The International Journal of Applied Forecasting 3 (51), 24–31.

Kalampokis, E., Tambouris, E., Tarabanis, K., 2013. Understanding the predictive power of social media. Internet Research 23 (5), 544–559.

Kaufman, L., Rousseeuw, P. J., 2005. Finding Groups in Data. An Introduction to Cluster Analysis, 2nd Edition. John Wil, New Jersey.

Kim, H., Hanssens, D. M., 2017. Advertising and word-of-mouth effects on pre-launch consumer interest and initial sales of experience products. Journal of Interactive Marketing 37, 57 – 74.

Kim, S., Shin, D. H., 2016. Forecasting short-term air passenger demand using big data from search engine queries. Automation in Construction 70, 98 – 108.

Kim, T., Hong, J., Kang, P., 2015. Box office forecasting using machine learning algorithms based on sns data. International Journal of Forecasting 31 (2), 364–390.

Kim, T., Hong, J., Kang, P., 2017. Box office forecasting considering competitive environment and word-of-mouth in social networks: A case study of korean film market. Computational Intelligence and Neuroscience, 1–16.

Kim, T., Hong, J., Koo, H., 2013. Forecasting diffusion of innovative technology at prelaunch: A surveybased method. Industrial Management & Data Systems 113 (6), 800–816.

Koning, A. J., Franses, P. H., Hibon, M., Stekler, H., 2005. The m3 competition: Statistical tests of the results. International Journal of Forecasting 21 (3), 397 – 409.

Kourentzes, N., 2019. tsutils: time series exploration, modelling and forecasting. R package version 0.9.0.
URL https://cran.r-project.org/package=tsutils

Kourentzes, N., Svetunkov, I., 2016. TStools: Time Series Analysis Tools and Functions. R package version 2.1.0.
URL https://github.com/trnnick/TStools

Kübler, R. V., Colicev, A., Pauwels, K., 2017. Social media's impact on consumer mindset: When to use which sentiment extraction tool? Marketing Science Institute Working Paper Series 17 (122), 1–99.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., Hunt, T., 2018. caret: classification and regression training. R package version 6.0-81.
URL https://cran.r-project.org/package=caret

Kuksov, D., Shachar, R., Wang, K., 2013. Advertising and consumers' communications. Marketing Science 32 (2), 294–309.

Kulkarni, G., Kannan, P., Moe, W., 2012. Using online search data to forecast new product sales. Decision Support Systems 52 (3), 604–611.

Kumar, V., Choi, J. B., Greene, M., 2016. Synergistic effects of social media and traditional marketing on brand sales: capturing the time-varying effects. Journal of the Academy of Marketing Science, 1–21.

Lampos, V., Cristianini, N., 2012. Nowcasting events from the social web with statistical learning. ACM Trans. Intell. Syst. Technol. 3 (4), 1–22.

Lau, R. Y. K., Zhang, W., Xu, W., 2018. Parallel aspect-oriented sentiment analysis for sales forecasting with big data. Production and Operations Management 27 (10), 1775–1794.

Lazer, D., Kennedy, R., King, G., Vespignani, A., 2014. The parable of google flu: Traps in big data analysis. Science 343 (6176), 1203–1205.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.

Lee, H., Kim, S. G., woo Park, H., Kang, P., 2014. Pre-launch new product demand forecasting using the bass model: A statistical and machine learning-based approach. Technological Forecasting and Social Change 86, 49 – 64.

Lee, J., Cho, Y., Lee, J.-D., Lee, C.-Y., 2006. Forecasting future demand for large-screen television sets using conjoint analysis with diffusion model. Technological Forecasting and Social Change 73 (4), 362 – 376.

Lee, J. E., Watkins, B., 2016. Youtube vloggers' influence on consumer luxury brand perceptions and intentions. Journal of Business Research 69 (12), 5753 – 5760.

Lee, S.-Y., Qiu, L., Whinston, A., 2018. Sentiment manipulation in online platforms: An analysis of movie tweets. Production and Operations Management 27 (3), 393–416.

Lee, W. Y., Goodwin, P., Fildes, R., Nikolopoulos, K., Lawrence, M., 2007. Providing support for the use of analogies in demand forecasting tasks. International Journal of Forecasting 23 (3), 377 – 390.

Lenk, P. J., Rao, A. G., 1990. New models from old: Forecasting product adoption by hierarchical bayes procedures. Marketing Science 9 (1), 42–53.

Leung, L., 2013. Generational differences in content generation in social media: The roles of the gratifications sought and of narcissism. Computers in Human Behavior 29 (3), 997 – 1006.

Li, X., 2016. Nowcasting with big data: is google useful in presence of other information? London Business School Mimeo, 1–41.
URL https://bit.ly/2ZzdPYS

Li, X., Hitt, L. M., 2008. Self-selection and information role of online product reviews. Information Systems Research 19 (4), 456–474.

Li, X., Hitt, L. M., Zhang, Z. J., 2011. Product reviews and competition in markets for repeat purchase products. Journal of Management Information Systems 27 (4), 9–42.

Li, X., Pan, B., Law, R., Huang, X., 2017. Forecasting tourism demand with composite search index. Tourism Management 59, 57 – 66.

Limnios, C., You, H., 2016. Can google trends actually improve housing market forecasts?, (December, 2016), available at SSRN: http://dx.doi.org/10.2139/ssrn.2886705.

Liu, B., 2015. Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. Cambridge University Press.

Liu, X., Singh, P. V., Srinivasan, K., 2016. A structured analysis of unstructured big data by leveraging cloud computing. Marketing Science 35 (3), 363–388.

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

Liu-Thompkins, Y., 2012. Seeding viral content: The role of message and network factors. Journal of Advertising Research 52 (4), 465–478.

Luan, Y. J., Sudhir, K., 2010. Forecasting marketing-mix responsiveness for new products. Journal of Marketing Research 47 (3), 444–457.

Lubba, C. H., Sethi, S. S., Knaute, P., Schultz, S. R., Fulcher, B. D., Jones, N. S., 2019. catch22: Canonical time-series characteristics. Selected through highly comparitive time-series analysis. Data Mining and Knowledge Discovery, 1–32, pre-print.

Luo, X., Zhang, J., 2013. How do consumer buzz and traffic in social media marketing predict the value of the firm? Journal of Management Information Systems 30 (2), 213–238.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., 2018. cluster: cluster analysis basics and extensions. R package version 2.0.7-1.
URL https://cran.r-project.org/package=cluster

Mahajan, V., Sharma, S., Sharma, S., 1986. A simple algebraic estimation procedure for innovation diffusion models of new product acceptance. Technological Forecasting and Social Change 30 (4), 331 – 345.

Marchand, A., 2016. The power of an installed base to combat lifecycle decline: The case of video games. International Journal of Research in Marketing 33 (1), 140 – 154.

Marchand, A., Hennig-Thurau, T., 2013. Value creation in the video game industry: Industry economics, consumer benefits, and research opportunities. Journal of Interactive Marketing 27 (3), 141 – 157.

Marchand, A., Hennig-Thurau, T., Wiertz, C., 2017. Not all digital word of mouth is created equal: Understanding the respective impact of consumer reviews and microblogs on new product success. International Journal of Research in Marketing 34 (2), 336 – 354.

Markham, S. K., Lee, H., 2013. Product development and management association's 2012 comparative performance assessment study. Journal of Product Innovation Management 30 (3), 408–429.

Markovitch, D. G., Steckel, J. H., Michaut, A., Philip, D., Tracy, W. M., 2014. Behavioral reasons for new product failure: Does overconfidence induce overforecasts? Journal of Product Innovation Management 32 (5), 825–841.

Mavragani, A., Tsagarakis, K. P., 2016. Yes or no: Predicting the 2015 greferendum results using google trends. Technological Forecasting and Social Change 109 (Supplement C), 1 – 5.

Mayzlin, D., Dover, Y., Chevalier, J., 2014. Promotional reviews: An empirical investigation of online review manipulation. The American Economic Review 104 (8), 2421–2455.

Meade, N., Islam, T., 2006. Modelling and forecasting the diffusion of innovation - a 25-year review. International Journal of Forecasting 22 (3), 519–545.

Meeran, S., Jahanbin, S., Goodwin, P., Neto, J. Q. F., 2017. When do changes in consumer preferences make forecasts from choice-based conjoint models unreliable? European Journal of Operational Research 258 (2), 512 – 524.

Moe, W. W., Fader, P. S., 2002. Fast-track: Article using advance purchase orders to forecast new product sales. Marketing Science 21 (3), 347–364.

Moe, W. W., Schweidel, D. A., 2012. Online product opinions: Incidence, evaluation, and evolution. Marketing Science 31 (3), 372–386.

Moe, W. W., Trusov, M., 2011. The value of social dynamics in online product ratings forums. Journal of Marketing Research 48 (3), 444–456.

Moon, S., Kim, J., Bayus, B. L., Yi, Y., 2016. Consumers pre-launch awareness and preference on movie sales. European Journal of Marketing 50 (5/6), 1024–1046.

Mostard, J., Teunter, R., de Koster, R., 2011. Forecasting demand for single-period products: A case study in the apparel industry. European Journal of Operational Research 211 (1), 139 – 147.

Muchnik, L., Aral, S., Taylor, S. J., 2013. Social influence bias: A randomized experiment. Science 341 (6146), 647–651.

Mülbacher, H., Füller, J., Huber, L., 2011. Online forum discussion-based forecasting of new product market performance. Marketing ZFP 33 (3), 221–234.

Ng, P. T., Maechler, M., 2017. COBS – constrained b-splines (sparse matrix based). R package version 1.3.3.
URL https://CRAN.R-project.org/package=cobs

Nguyen, H. T., Chaudhuri, M., 2018. Making new products go viral and succeed. International Journal of Research in Marketing, 1–24, pre-print.

Norton, J. A., Bass, F. M., 1987. A diffusion theory model of adoption and substitution for successive generations of high-technology products. Management Science 33 (9), 1069–1086.

Önder, I., 2017. Forecasting tourism demand with google trends: Accuracy comparison of countries versus cities. International Journal of Tourism Research 19 (6), 648–660.

Önder, I., Gunter, U., 2016. Forecasting tourism demand with google trends for a major european city destination. Tourism Analysis 21 (2-3), 203–220.

Önder, I., Gunter, U., Gindl, S., 2019. Utilizing facebook statistics in tourism demand modeling and destination marketing. Journal of Travel Research, 1–14, pre-print.

Onishi, H., Manchanda, P., 2012. Marketing activity, blogging and sales. International Journal of Research in Marketing 29 (3), 221 – 234.

Orbach, Y., Fruchter, G. E., 2011. Forecasting sales and product evolution: The case of the hybrid/electric car. Technological Forecasting and Social Change 78 (7), 1210 – 1226.

Ord, J. K., Fildes, R., Kourentzes, N., 2017. Principles of business forecasting, 2nd Edition. Wessex, New York.

Padhi, S. S., Pati, R. K., 2017. Quantifying potential tourist behavior in choice of destination using google trends. Tourism Management Perspectives 24, 34 – 47.

Pan, B., Wu, D. C., Song, H., 2012. Forecasting hotel room demand using search engine data. Journal of Hospitality and Tourism Technology 3 (3), 196–210.

Pan, B., Yang, Y., 2016. Forecasting destination weekly hotel occupancy with big data. Journal of Travel Research, 1–14.

Park, S., Lee, J., Song, W., 2017. Short-term forecasting of japanese tourist inflow to south korea using google trends data. Journal of Travel & Tourism Marketing 34 (3), 357–368.

Peng, G., Liu, Y., Wang, J., Gu, J., 2016. Analysis of the prediction capability of web search data based on the HE-TDC method: prediction of the volume of daily tourism visitors. Journal of Systems Science and Systems Engineering, 1–20.

Perera, H. N., Hurley, J., Fahimnia, B., Reisi, M., 2019. The human factor in supply chain forecasting: A systematic review. European Journal of Operational Research 274 (2), 574 – 600.

Perlin, M. S., Caldeira, J. F., Santos, A. A. P., Pontuschka, M., 2017. Can we predict the financial markets based on google's search queries? Journal of Forecasting 36 (4), 454–467.

Petropoulos, F., Kourentzes, N., Nikolopoulos, K., Siemsen, E., 2018. Judgmental selection of forecasting models. Journal of Operations Management 60, 34 – 46.

Phillips, L., Dowling, C., Shaffer, K., Hodas, N. O., Volkova, S., 2017. Using social media to predict the future: A systematic literature review. CoRR abs/1706.06134.
URL http://arxiv.org/abs/1706.06134

Preis, T., Moat, H. S., 2014. Adaptive nowcasting of influenza outbreaks using google searches. Royal Society Open Science 1 (2), 1–5.

R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL http://www.R-project.org/

R Core Team, 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
URL https://www.R-project.org/

Ramsay, J., Silverman, B. W., 2005. Functional data analysis, 2nd Edition. Springer Series in Statistics. Springer, New York.

Rao, T., Srivastava, S., 2013. Modeling movements in oil, gold, forex and market indices using search volume index and twitter sentiments. In: Proceedings of the 5th Annual ACM Web Science Conference. WebSci '13. ACM, New York, NY, USA, pp. 336–345.

Ravi, K., Ravi, V., 2015. A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. Knowledge-Based Systems 89, 14 – 46.

Rhyn, L., 2019. Der Tourist von heute dokumentiert jeden seiner Schritte auf Social Media. Das verändert Dörfer, Bergbahnen und Restaurants nicht nur zum Guten. Neue Zürcher Zeitung, online; accessed 27/08/2019.
URL https://bit.ly/2L7ZHRh

Rivera, R., 2016. A dynamic linear model to forecast hotel registrations in puerto rico using Google Trends data. Tourism Management 57, 12 – 20.

Rogers, E. M., 2003. Diffusion of Innovations, 5th Edition. Free Press, New York.

Ruohonen, J., Hyrynsalmi, S., 2017. Evaluating the use of internet search volumes for time series modeling of sales in the video game industry. Electronic Markets, 1–20.

Ruths, D., Pfeffer, J., 2014. Social media for large studies of behavior. Science 346 (6213), 1063–1064.

Sagaert, Y. R., Aghezzaf, E.-H., Kourentzes, N., Desmet, B., 2018. Tactical sales forecasting using a very large set of macroeconomic indicators. European Journal of Operational Research 264 (2), 558–569.

Santillana, M., Nguyen, A. T., Dredze, M., Paul, M. J., Nsoesie, E. O., Brownstein, J. S., 2015. Combining search, social media, and traditional data sources to improve influenza surveillance. PLOS Computational Biology 11 (10), 1–15.

Sawhney, M. S., Eliashberg, J., 1996. A parsimonious model for forecasting gross box-office revenues of motion pictures. Marketing Science 15 (2), 113–131.

Schaer, O., 2018. GTT: Functions for analysis with Google Trends search traffic information. R package version 0.0.1.
URL https://github.com/mamut86/GTT

Schaer, O., Kourentzes, N., 2018. diffusion: Forecast the Diffusion of New Products. R package version 0.2.7.
URL https://cran.r-project.org/package=diffusion

Schaer, O., Kourentzes, N., Fildes, R., 2019a. Demand forecasting with user-generated online information. International Journal of Forecasting 35 (1), 197 – 212.

Schaer, O., Kourentzes, N., Fildes, R., 2019b. Estimating the market potential with pre-release buzz. Lancaster University Management School, Management Science Paper Series 2019:1, 1–38, accessed 24/08/2019.
URL https://ssrn.com/abstract=3325878

Schneider, M. J., Gupta, S., 2016. Forecasting sales of new and existing products using consumer reviews: A random projections approach. International Journal of Forecasting 32 (2), 243 – 256.

Schroll, R., Grohs, R., 2018. Uncertainty in prerelease advertising. Journal of Advertising, 1–14, pre-print.

Scott, S. L., Varian, H. R., 2015. Bayesian variable selection for nowcasting economic time series. In: Goldfarb, A., Greenstein, S. M., Tucker, C. E. (Eds.), Economic Analysis of the Digital Economy. University of Chicago Press, pp. 119–135.

See-To, E. W. K., Ngai, E. W. T., 2016. Customer reviews for demand distribution and sales nowcasting: a big data approach. Annals of Operations Research, 1–17.

Seebach, C., Pahlke, I., Beck, R., 2011. Tracking the digital footprints of customers: How firms can improve their sensing abilities to achieve business agility. In: ECIS 2011 Proceedings. Vol. 258. pp. 1–13.

Seifert, M., Siemsen, E., Hadida, A. L., Eisingerich, A. B., 2015. Effective judgmental forecasting in the context of fashion products. Journal of Operations Management 36, 33 – 45.

Seiler, S., Yao, S., Wang, W., 2017. Does online word of mouth increase demand? (and how?) Evidence from a natural experiment. Marketing Science 36 (6), 838–861.

Siliverstovs, B., Wochner, D. S., 2018. Google Trends and reality: Do the proportions match? Appraising the informational value of online search behavior: Evidence from Swiss tourism regions. Journal of Economic Behavior & Organization 145, 1 – 23.

Situmeang, F. B., Leenders, M. A., Wijnberg, N. M., 2014. The good, the bad and the variable: How evaluations of past editions influence the success of sequels. European Journal of Marketing 48 (7/8), 1466–1486.

Skodda, C., Benthaus, J., 2015. Investigating consumer information search behavior and consumer emotions to improve sales forecasting. In: 21st Americas Conference on Information Systems. pp. 1–12.

Smith, P., 2016. Google's MIDAS touch: Predicting UK unemployment with internet search data. Journal of Forecasting 35 (3), 263–284, for.2391.

Sood, A., James, G. M., Tellis, G. J., 2009. Functional regression: A new model for predicting market penetration of new products. Marketing Science 28 (1), 36–51.

Sood, S., Drèze, X., 2006. Brand extensions of experiential goods: Movie sequel evaluations. Journal of Consumer Research 33 (3), 352–360.

Srinivasan, S., Rutz, O. J., Pauwels, K., 2016. Paths to and off purchase: quantifying the impact of traditional marketing and online consumer activity. Journal of the Academy of Marketing Science 44 (4), 440–453.

Statista, 2017. Breakdown of U.S. computer and video game sales from 2009 to 2015, by delivery format. Online; accessed 27/01/2017.
URL https://bit.ly/2nT3ick

Stephen, A. T., Galak, J., 2012. The effects of traditional and social earned media on sales: A study of a microlending marketplace. Journal of Marketing Research 49 (5), 624–639.

Sugar, C. A., James, G. M., 2003. Finding the number of clusters in a dataset. Journal of the American Statistical Association 98 (463), 750–763.

Tashman, L. J., 2000. Out-of-sample tests of forecasting accuracy: an analysis and review. International Journal of Forecasting 16 (4), 437 – 450.

Teixeira, T., Wedel, M., Pieters, R., 2012. Emotion-induced engagement in internet video advertisements. Journal of Marketing Research 49 (2), 144–159.

Teo, T. S., Choo, W. Y., 2001. Assessing the impact of using the internet for competitive intelligence. Information & Management 39 (1), 67 – 83.

Tian, C. H., Wang, Y., Mo, W. T., Huang, F. C., Dong, S., Huang, 2014. Pre-release sales forecasting: A model-driven context feature extraction approach. IBM Journal of Research and Development 58 (5/6), 8:1–8:13.

Tibshirani, R., Walther, G., Hastie, T., 2001. Estimating the number of clusters in a data set via the gap statistic. Journal of the Royal Statistical Society: Series B 63 (2), 411–423.

Timoshenko, A., Hauser, J. R., 2019. Identifying customer needs from user-generated content. Marketing Science 38 (1), 1–20, , pre-print.

Tirunillai, S., Tellis, G. J., 2012. Does chatter really matter? Dynamics of user-generated content and stock performance. Marketing Science 31 (2), 198–215.

Toubia, O., Goldenberg, J., Garcia, R., 2014. Improving penetration forecasts using social interactions data. Management Science 60 (12), 3049–3066.

Trapero, J. R., Cardós, M., Kourentzes, N., 2019. Empirical safety stock estimation based on kernel and GARCH models. Omega 84, 199 – 211.

Trusov, M., Rand, W., Joshi, Y. V., 2013. Improving prelaunch diffusion forecasts: Using synthetic networks as simulated priors. Journal of Marketing Research 50 (6), 675–690.

Tyebjee, T. T., 1987. Behavioral biases in new product forecasting. International Journal of Forecasting 3 (3), 393 – 404.

van der Lans, R., van Bruggen, G., Eliashberg, J., Wierenga, B., 2010. A viral branching model for predicting the spread of electronic word of mouth. Marketing Science 29 (2), 348–365.

Varian, H. R., 2017. Can google trends data be replicated? http://econbrowser.com/archives/2017/06/guest-contribution-can-google-trends-data-be-replicated, online; accessed: 28/08/2017.

Vaughan, L., Chen, Y., 2015. Data mining from web search queries: A comparison of google trends and baidu index. Journal of the Association for Information Science and Technology 66 (1), 13–22.

Von Graevenitz, G., Helmers, C., Millot, V., Turnbull, O., 2016. Does online search predict sales? evidence from big data for car markets in Germany and the UK. CRG Working Paper 71, 1–32.

Vosen, S., Schmidt, T., 2011. Forecasting private consumption: survey-based indicators vs. Google Trends. Journal of Forecasting 30 (6), 565–578.

Wang, F., Zhang, Y., Li, X., Zhu, H., 2010. Why do moviegoers go to the theater? The role of prerelease media publicity and online word of mouth in driving moviegoing behavior. Journal of Interactive Advertising 11 (1), 50–62.

Wang, X., Yu, C., Wei, Y., 2012. Social media peer communication and impacts on purchase intentions: A consumer socialization framework. Journal of Interactive Marketing 26 (4), 198–208.

Won, H.-H., Myung, W., Song, G.-Y., Lee, W.-H., Kim, J.-W., Carroll, B. J., Kim, D. K., 2013. Predicting national suicide numbers with social media data. PLOS ONE 8 (4), 1–6.

Wu, L., Brynjolfsson, E., 2015. The future of prediction: How google searches foreshadow housing prices and sales. In: Goldfarb, A., Greenstein, S. M., Tucker, C. E. (Eds.), Economic analysis of the digital economy, 1st Edition. University of Chicago Press, Chicago, pp. 89–118.

Xiong, G., Bharadwaj, S., 2014. Prerelease buzz evolution patterns and new product performance. Marketing Science 33 (3), 401–421.

Xu, Q., Gel, Y. R., Ramirez Ramirez, L. L., Nezafati, K., Zhang, Q., Tsui, K.-L., 2017. Forecasting influenza in hong kong with google search queries and statistical model fusion. PLOS ONE 12 (5), 1–17.

Xu, R., Wunsch, D. C., 2009. Clustering. John Wiley & Sons, Inc., New Jersey.

Yi, J., Lee, Y., Kim, S.-H., 2019. Determinants of growth and decline in mobile game diffusion. Journal of Business Research 99, 363 – 372.

Yu, L., Zhao, Y., Tang, L., Yang, Z., 2018. Online big data-driven oil consumption forecasting with google trends. International Journal of Forecasting, 1–11, pre-print.

Zeynalov, A., 2017. Forecasting tourist arrivals in prague: Google econometrics, (December, 2017), available at MPR: https://bit.ly/2L1MPgQ.

Zheng, Z. E., Fader, P., Padmanabhan, B., 2012. From business intelligence to competitive intelligence: Inferring competitive measures using augmented site-centric data. Information Systems Research 23 (3-part-1), 698–720.

Zhu, F., Zhang, X. M., 2010. Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. Journal of Marketing 74 (2), 133–148.