Ann Oper Res (2018) 270:415–431
https://doi.org/10.1007/s10479-016-2296-z

BIG DATA ANALYTICS IN OPERATIONS & SUPPLY CHAIN MANAGEMENT

Customer reviews for demand distribution and sales nowcasting: a big data approach

Eric W. K. See-To[1] · **Eric W. T. Ngai**[2]

Published online: 24 August 2016

Abstract Proliferation of online social media and the phenomenal growth of online commerce have brought to us the era of big data. Before this availability of data, models of demand distribution at the product level proved elusive due to the ever shorter product life cycle. Methods of sales forecast are often conceived in terms of longer-run trends based on weekly, monthly or even quarterly data, even in markets with rapidly changing customer demand such as the fast fashion industry. Yet short-run models of demand distribution and sales forecasting (aka. nowcast) are arguably more useful for managers as the majority of their decisions are concerned with day to day discretionary spending and operations. Observations in the fast fashion market were acquired, for a collection time frame of about 1 month, from a major Chinese e-commerce platform at granular, half-daily intervals. We developed an efficient method to visualize the demand distributional characteristics; found that big data streams of customer reviews contain useful information for better sales nowcasting; and discussed the *current* influence pattern of sentiment on sales. We expect our results to contribute to practical visualization of the demand structure at the product level where the number of products is high and the product life cycle is short; revealing big data streams as a source for better sales nowcasting at the corporate and product level; and better understanding of the influence of online sentiment on sales. Managers may thus make better decisions concerning inventory management, capacity utilization, and lead and lag times in supply-chain operations.

Keywords Big data · Sales nowcasting · Short-run operation · Demand distribution

✉ Eric W. K. See-To
eric.see-to@polyu.edu.hk

Eric W. T. Ngai
eric.ngai@polyu.edu.hk

[1] Department of Industrial and Systems Engineering, Hong Kong Polytechnic University, Kowloon, Hong Kong

[2] Department of Management and Marketing, Hong Kong Polytechnic University, Kowloon, Hong Kong

1 Introduction

This research studies the informational content of big data streams from a big e-commerce platform to support supply chain management. A large amount of data is generated from social media every day (Ekbia et al. 2015), and by the end of 2012, around 2.5 exabytes of data per day were brought, and the number was doubling every 40 months or so (McAfee and Brynjolfsson 2012), which brings challenges and opportunities for organizations in extracting data for knowledge in making decisions (Buhl et al. 2013). In addition, data generated within a company (Babu et al. 2014; Li et al. 2015) and outside the company from the Internet (Sanders and Ganeshan 2015) for supply chain management is increasingly voluminous. These offer tremendous potential for more effective supply chain management based on big data and predictive analytics (Waller and Fawcett 2013), with lower inventory and faster customer response (Christopher and Ryals 2014).

Furthermore, with the large capacity of big data, the prediction is transforming from low frequency (quarterly, monthly) forecasting to a high frequency nowcast (daily) to acquire more accurate of prevision (Hirashima et al. 2015) to support real time decision making, especially for supply chain management. Carriere-Swallow and Labbe (2013) found automobile sales could be nowcast by analyzing the Google trend data. Tan et al. (2015) thought innovation for supply chain management could be achieved with big data analytics so as to enhance the competitive advantages of firms.

In particular, a series of research models for sales forecasting or nowcasting have been proposed so as to reduce the under stock or over stock phenomena and offer precise decision making by applying data mainly from following two channels by examining the literature. First, from the perspective of traditional transactions or financial data, different researchers have developed different models to predict the sales. Second, data from Google search, Internet clickstreams, or social media, such as the Twitter, Facebook, blogs, Sina MicroBlog, and Taobao was extracted to forecast or nowcast the sales of products. However, the effect of the combination of the usage of these two big data was received relatively little concern.

In this paper, we ask three research questions: (1) from the big data streams of sales transactions, how can we extract relevant knowledge about the underlying functional form of customer demand structure? (2) From the big data streams of customer reviews, is there useful information for better sales volume nowcasting? (3) Do the big data streams contain useful information for an understanding of the *current* pattern of the influence of customer reviews on sales volume?

Two particular product-level, high frequency (sampled at half-daily intervals), big data streams from a major Chinese electronic commerce platform, are investigated: one concerning the sales transaction itself; and the other the customer sentiment contained in the reviews. From the big data streams of sales transactions, we aim to provide an efficient visual way to examine the distributional characteristics of product level demand from an online sales platform. From the big data streams of customer reviews, we first test whether there is additional information in it to improve sales nowcasting predictive power. Second, we further analyze the big data to reveal the *current* influence pattern of customer sentiment on product sales.

The rest of the paper is as follows. Section 2 discusses current literature and our contribution to the extant research. Section 3 outlines the data collection process, and the analysis strategy. Section 4 presents the results. Section 5 discusses and concludes the study.

2 Literature review

Since big data appeared in the article of Cox and Ellsworth (1997), the definition of big data has been in dispute. Some scholars define big data as data sets that are large and complicated so as to deal with standard statistical software (Snijders et al. 2012), which is beyond people's ability to comprehend (Weinberger 2014). While others consider Big Data are technologies and capabilities that are able to collect, manage, and analyze those huge data sets (Boyd and Crawford 2012; Ekbia et al. 2015). In addition, big data has five dimensions- volume (large amount), velocity (high speed of data processing), variety (various types of data) (Laney 2001), variability (expansion in the range of data values) and value (useful for business) (Li et al. 2015).

2.1 Nowcasting by using big transactional data

Big data streams provide a possible channel for us to extract relevant knowledge about the underlying functional form of the customer demand structure. The distributional characteristics of market transaction data are fundamental to decision-making in many activities in modern corporations. These provide a foundation for understanding the demand structure, which is critical for supply chain and operational decisions. Extracting and comprehending demand distributional information from online big data streams is challenging for two reasons. First, there are a large number of products being sold at the same time, and it is difficult to process so much information simultaneously. Second, the shelf life of each product tends to be shorter and shorter in this era of electronic commerce, which means the number of per time period data points for each product is small using usual sampling frequencies such as weekly or monthly. In addition, according to the survey of AgilOne (2014), nearly 70 % of ecommerce retailers will apply predictive analytics in their sales channels, because big data has the capability to improve the operation management of organizations.

It is believed that accurate sales forecasting is beneficial in reducing the cost of stocking (Chung et al. 2012) and effective in improving the performance of both suppliers (Amornpetchkul et al. 2015; Mishra et al. 2009) and resellers (Chen and Xiao 2012), with value added to each of them (Balar et al. 2013). With the advent of big data, data accelerates at an unprecedentedly speed which offers the raw material for real time and accurate nowcasting. Compared with forecasting, nowcasting deals with higher frequency data, namely data from very narrow time horizons (Banbura et al. 2011), such as daily or half-daily data, to predict the present, the very near future and the very recent past (Banbura et al. 2013) so that nowcasting can support fast decision making for the daily operations management. Therefore, various sales nowcasting models have been developed to predict the sales of various products and services, like automobiles (Carriere-Swallow and Labbe 2013) and telecom products (Bughin 2015).

Chung et al. (2012) developed the networking model and the re-rent model, which could offer excellent prediction for the sales of Blockbuster in practice. They believed their research could improve the inventory management of Blockbuster due to keeping a reasonable stock level. Osadchiy et al. (2013) developed a forecasting model for retailers in the USA based on lagging financial market returns and demonstrated that the market-based forecast achieved an average 15 % improvement in accuracy compared with forecasts given by equity analysts. They believed the proposed model was more appropriate for retailers in jewelry, office supplies, clothing, durable goods, and home furnishings for forecasting sales. In addition, support vector regression (SVR) models combining independent component analysis (ICA) and K-means clustering were applied to forecast the information technology (IT) product agent sales which revealed that the proposed sales forecasting models could provide efficient sales forecasting (Lu and Chang 2014). An extreme learning machine (ELM) model was applied to examine real data sets for fashion products and concluded ELM was fast and effective in sales forecasting (Yu et al. 2011). On the other hand, Guo et al. (2013) developed an multivariate intelligent decision-making (MID) model to offer effective retail forecasting with a harmony search-wrapper-based variable selection (HWVS) module and a multivariate intelligent forecaster (MIF) module. They believed the proposed MID model could outperform the extreme learning machine-based model and generalized linear model by extensively testing typical sales datasets from real world retail industry.

2.2 Nowcasting by using big customer review data

Do the big data streams contain useful information for understanding the *current* pattern of the influence of customer reviews on sales volume? Significant relationships exist between the public mood identified from Sina MicroBlog and the sales of Sony cameras (Ma and Zhang 2015). After analyzing review properties, reviewer characteristics and review influences from an online forum, Chern et al. (2015) found that product sales were significantly associated with the online viewpoints of customers and the sales forecasting model was more effective with a large number of online reviews. Moreover, sentiment analysis of the viewpoints from blogs was effective in forecasting movie sales, which was proven by a sentiment-aware model (Yang et al. 2007). Discounts, provision of free delivery options, and online review information such as the ratings of the products and the percentage of positive and negative reviews on products of Amazon.com are effective product sales predictors that can be used to build a model for sales forecasting (Chong et al. 2015). Huang and Mieghem (2014) formulated a model using clickstream data from non-transactional websites for inventory management. They found that the noisy clickstream data was statistically significant in predicting the propensity, amount, and timing of offline orders. In addition, the demand information extracted from the clickstream data could reduce the inventory holding and backordering cost by 3–5 % in the data set through counter-factual analysis. Bughin (2015) applied an Error Correction Mechanism (ECM) model to nowcast the telecom sales in Belgium with the data from Google searches and social media (Twitter, Facebook and other blogs) comments and found the ECM model with nowcasting variables could improve telecom sales forecasts by about 25 % compared with a naive autoregressive sales model. Lassen et al. (2014) evaluated a linear regression model that transformed iPhone tweets into a prediction of the quarterly iPhone sales. They discovered that the correlation between iPhone tweets and iPhone sales became significantly stronger after incorporating the sentiments in tweets, which indicated that data from social media was valuable for sales forecasting. Kim et al. (2015) examined the data from wireless sensor networks and topic modeling of buzzwords in blog documents and proposed models for demand forecasting of medicines. They found that the content from blogs significantly influenced the performance of demand forecasting. Chern et al. (2015) proposed eWord-of-Mouth Sales Forecasting Algorithm (WOMSFA) to forecast sales by analyzing review properties, reviewer characteristics and review influences from an online forum. By testing the WOMSFA with the data from a popular cosmetic retailer in Taiwan, the empirical results indicated that WOMSFA was more suitable than traditional time series forecasting models for products with sufficient online reviews.

3 Methodology

In this section, we outline first the empirical context of the study. Then the data collection and processing process is described. Finally, the three methods used to answer the three research questions respectively are discussed.

3.1 The empirical context

In this study, we collected data from a well-known in China for business-to-consumer (B2C) online retailing, namely the Tmaill.com, spun off from Taobao. The Tmall.com was well suited for the purpose of this study as the source of big data streams, for both sales transactions and customer reviews.

First, Tmall.com is representative of the current situation of the Chinese e-commerce market, and can provide a sufficiently large volume of sales transaction data, in our selected collection time frame, at the chosen highly frequent, half-daily intervals. Tmall.com is the most well-known B2C retail website in China. It currently features more than 70,000 international and Chinese brands from more than 50,000 merchants and serves more than 180 million buyers (Felix 2015). The daily transaction volume is extremely large, even for different shops or different product categories. As we will show later, the instock situation of Tmall.com is excellent, allowing us to use the sales transaction data confidently as a good proxy measure of the underlying demand.

Second, the customer review data collected on this website is more reliable in reflecting the true sentiment of consumers. Tmall.com has a well-established appraisal system for consumers to evaluate the product they purchased. The strict regulation for evaluation reduces the possibility of baleful or false review. Last but not least, they are ready to use review-based predictors, constructed by Tmall.com, that can be used in this research. For example, it divided customer reviews into several sentiment-based categories, some of which are positive while the others are negative. In this paper, the positive ones are called positive tags, and the negative ones are called negative tags.

3.2 Data collection and processing

We collected the transaction data of four clothing shops on Tmall.com in the period from 25th October 2015 to 30th November 2015. The transaction data was downloaded from Tmall.com. These four shops are Metersbonwe, Jeanswest, Wacoal and Triumph. We choose these four shops based on two reasons. Firstly, data collection was supported by a related project for Jeanswest and Triumph, where we are developing an intelligent decision support system for real time social media sentiment analytics. Secondly, Wacoal and Metersbonwe are the major competitors of Triumph and Jeanswest respectively. During the data collection period, there were 10002 transactions and 9514 customer reviews involved in 2527 different products in the 4 selected shops. Since the time for data collection was quite short and the number of shops involved was small, it is considered that the volume and velocity of the data is big. The formats of the data include structured transaction records and unstructured customer reviews. Some of the customer reviews provide valuable information for describing the quality of the products, but some of them are less useful or even can be classified as spam messages which are excluded in this study. We expected that the model established in this study could support the intelligent system development and contribute to academia as well. Metersbonwe and Jeanswest both sell casual wear, and Wacoal and Triumph both sell under wear. We intentionally included in our dataset two very different, but related (both fast

Table 1 Summary of variables after data processing

	Variable name	Notation
Variables from the sales transaction big data stream	Sales volume	S_{it}
	Shop_id	I_i
	Stock	K_{it}
	New/Hot	N_i
	AM/PM	A_t
	Price	P_{it}
	Double-eleven	D_i
Variables from the customer review big data stream	Cumulative review sentiment	R_{it}
	Number of words for positive review	
	Number of words for negative review	
	Number of positive review	
	Number of negative review	
	Number of positive tag	
	Number of negative tag	

fashion) business categories. We would also like to explore the similarities and differences for useful knowledge contained in their big data streams. For example, it may be possible that product sales of under wear are more sensitive to reviews than that of casual wear. In this study, we collected and processed data in three steps: Firstly, based on the suggestions of the management of the test-bed companies, we divided one day into halves with the break point at 12:00 noon. Before this time point, we denoted the time bucket as AM, while denoting the time bucket after 12:00 noon as PM. One half of a day is defined as one time period in our dataset. The management of the test-bed companies choose half day intervals because it is their common practice in reviewing their sales. Secondly, we sorted out all the hot or new products from the merchandise of these four shops. Thirdly, for each product, we collected their sales volume and associated variables for each AM and PM time bucket during the designated data collection time periods. Table 1 summarizes all the variables in our data set after processing as described below.

The variables available from the big data stream of the sales transactions are as follow. For a product i sold in a time period t, we have following variables. (1) *Sales volume* (S_{it}): this is the volume of sales during a time period t for a product i. (2) *Shop_id* (I_i): this feature is the ID for these four shops. In our model, shop id is a dummy variable. (3) *Stock* (K_{it}): stock here represents the real time inventory of the product i at time t. This information is available online. When consumers intend to purchase a product, they can see the figure for the remaining stock. This figure is commonly seen in most retail websites. (4) *New/Hot* (N_i): this feature indicates whether this merchandise is a new product or a hot product. (5) *AM/PM* (A_t): this feature indicates a different time bucket for a transaction. In this study, we divided one day into two transaction periods, namely AM (morning) and PM (afternoon). (6) *Price* (P_{it}): this variable represents the price for a product i when it is sold in time period t. (7) *Double-Eleven* (D_i): this dummy variable indicates whether the data points are collected during the Double Eleven Shopping Festival—a big sales event in China (on the 11th of November).

The variables available from the big data stream of the customer reviews are as follow. We collectively denote all the seven variables here as the *customer review factor* $(R_{it}$. The factor includes the following variables: (1) *Number of positive tags*: Number of customer reviews tagged as positive by Tmall.com at the product level. (2) *Number of negative tags*: Number of customer reviews tagged as negative by Tmall.com at the product level. (3) *Cumulative review sentiment*: We collected all the customer reviews available from 25th October 2015. The reviews were in Chinese. HowNet is a commonly used dictionary to categorize words contained in reviews into different sentiment categories, some being positive with others being negative (http://www.keenage.com/html/e_index.html). Having done the categorization, we counted the number of words in each category for all reviews. Original review text would be converted into a "time period-category" matrix that shows the word count of each category on each half-day period. Principal component analysis (PCA) was then used on the sentiment categories to obtain a principal component. The principal component is found to be positively correlated to positive sentiment categories and negatively correlated to negative categories, and thus measures how positive a review is. By mapping the original matrix onto the factor, we are able to obtain a set of numbers measuring the sentiments indicated by the reviews for each time period. An overall score was then calculated to represent cumulative review sentiment. This cumulative sentiment score is then be used to categorize reviews into being either positive or negative. (4) *Number of positive reviews*: In this dataset, we find that the number of Tmall.com tagged positive reviews amounted to 86.7 % out of the total. In addition, from our discussion with frequent Tmall.com shoppers and industry experts, positive reviews always dominate, being about 90 % out of the total. Based on this, we sorted customer reviews by descending order of their sentiment score. The top 90 % of the sorted reviews are considered as positive. (5) *Number of negative reviews*: The bottom 10 % of all the sorted reviews are defined as negative. (6) *Number of words for positive review*: This variable counts the average number of words for all the positive reviews for a product. (7) *Number of words for negative review*: This variable counts the average number of words for all the negative reviews for a product.

3.3 The distributional fit visualization method

The number and category of online products increases sharply and their life cycle is relatively shorter compared with normal goods (Cui et al. 2012). Both—a large number of products plus a short life cycle per product—make it particularly difficult to deal with when one tries to visualize produce demand characteristics. If we can visualize the distributional characteristic of product demand, businesses are able to adopt corresponding strategies to face challenges for future demand. This study employs Akaike information criterion (AIC) weights to measure the fit between the actual sales data and the chosen functional form of the demand distribution. Our objective is to see if this method is suitable for visualizing the distributional characteristic of demand for a large number of products, each with a short life cycle.

First introduced by Akaike (1973), AIC is generally employed to choose the model with lowest expected information loss, given that we have a set of models (M_i) for some collected data and the number of models is K. We denote L as the maximum likelihood for the model and V as the number of estimated parameters in the model,then we can calculate the AIC value for the model i:

$$AIC_i = -2logL_i + 2V_i$$

Following the the method introduced by Akaike, we then calculate the AIC weight for model $i$according to the following formula (Wagenmakers and Farrell 2004):

$$w_i\,(AIC) = \frac{\exp\left\{-\frac{1}{2}\Delta_i\,(AIC)\right\}}{\sum_{i=1}^{K}\exp\left\{-\frac{1}{2}\Delta_i\,(AIC)\right\}}$$

$w_i\,(AIC)$ can be interpreted as the probability that M_i is the best model and from the formula we can observe that the sum of $w_i\,(AIC)$ is 1. Thus, if the Akaike weight is closer to 1, the model or distribution is more fitted to real data. In this study, the AIC weight can help us visualize the distributional characteristics of product demand for different products over a range of candidate distributions. Comparing with other model selection methods, AIC penalizes the number of parameters less strongly than does the Bayesian information criterion (BIC). In Burnham and Anderson (2002), the authors show that AIC has theoretical advantages over BIC. First, AIC is derived from information principles. Second, they present a few simulation studies that suggest AIC tends to have practical and performance advantages over BIC (Burnham and Anderson 2004).

3.4 Test of predictive information from big data

To test if the big data stream of customer reviews contains useful information for sales nowcasting, we present here one baseline model (Model 0), one model using only sales transaction information (Model 1), and one model using both sales transaction and customer review information (Model 2). We compare their nowcast performances. If the customer review big data does contain significant nowcast information, we expect Model 2 will outperform the other two.

For a product i, denote the sales volume in the *t-th* period as $\{\mathrm{S}_{it}\colon t = 1, 2, \cdots, T\}$ and the customer review information until the *t-th* period as $\{\mathrm{R}_{it}\colon t = 1, 2, \cdots, T\}$. Firstly, we start with a simple baseline model: sales volume of this period that is predicted using the unweighted mean of sales volume in the previous four periods. This method is generally known as the Simple Moving Average (SMA).

$$Model\,0\colon St = \frac{S_{t-1} + S_{t-2} + S_{t-3} + S_{t-4}}{4}$$

As a simple method of forecasting, this model is used as a baseline measure to compare its performance with the other models incorporating big data of sales transactions and customer reviews (Booth et al. 2006).

To include big data from sales transactions, we build an OLS regression model to involve all the sales related variables (see Table 1). We also involve the sales volume of the previous four periods as variables in this model.

$$Model\,1\colon S_{it} \sim A_t + P_{it} + N_i + K_{it} + I + S_{t-1} + S_{t-2} + S_{t-3} + S_{t-4} + \mathrm{e}_t$$

The variable et is an error term. Since we involved four previous time period data of the dependent variable in this model, it is in fact an autoregressive model or an AR model. Adding the customer review factor to Model 1, we obtain:

$$Model2\colon S_{it} \sim A_t + P_{it} + N_i + K_{it} + I + S_{t-1} + S_{t-2} + S_{t-3} + S_{t-4} + R_{it} + \mathrm{e}_t$$

3.5 Current form of sales influence from customer reviews

Once it is established that customer review big data carries important information about sales, given the speed of change of online markets and their competitiveness, a useful question to ask is: what is the *current* form of influences from customer reviews on sales volume? The word "current" is key since customer taste and preferences are changing rapidly over time. A simple and quick method would be most useful. We would like to reveal the current impact pattern of customer reviews by looking at the significances of the product level influences from the seven review based variables on sales volume. We attempt to use Analysis of Variance (ANOVA) to evaluate all the seven variables available from the big data stream. ANOVA was used by Choi and Varian (2012) to determine if the variance of the dependent variable can be explained by specific predictors. In our case, through observing the significance of each predictor in customer review information, we can obtain an impact pattern of our review based predictors useful to both managers and researchers.

4 Results

4.1 Visualization of demand distributional characteristics

Here we demonstrate our visualization methodology for product level demand. In our dataset, the four shops on Tmall.com showed exceptional performance in managing inventory, maintaining an over 95 % product level in-stock probability. Even on 11th November, the Double Eleven Shopping Festival—the big Chinese Sales event, an over 85 % in-stock probability was maintained. This enabled us to use the product level sales volume data as a proxy measure for the true demand. Figure 1 presents the AIC weights for each product fitted in six different distributions. We can see that it provides a concise summary of the distributional characteristics of the sales volume data. Since the higher the AIC weight is, the better the distribution fits the data, this result suggests that the lognormal distribution fits the Tmall dataset best in this study, while Weibull distribution ranks the second. The Gamma distribution may be the third most suitable distribution, but most of the AIC weights of the Gamma distribution are already close to 0.

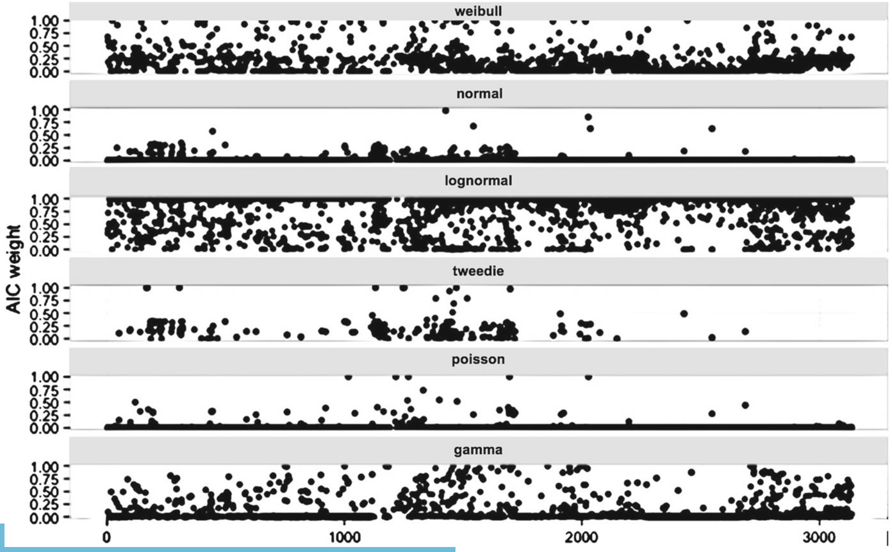

Fig. 1 Visual comparison of product demand fit across candidate distributions

4.2 Significance of predictive information from big data

Sales volume is the dependent variable in our analysis of nowcasting the effects of customer reviews. To do this we use an autoregressive model, involving S_{it-1} to S_{it-4} as four control variables. For this part of analysis, we subdivide the dataset we collected. We want to see if our results are robust across different datasets where the overall intensity of the customer reviews differs. The whole dataset, denoted as D_1, has all the data points we collected in this study. Then, we divide D_1 into two datasets according the number of reviews. The first dataset contains the data points whose number of reviews is smaller than 10. The second dataset D_2 contains the data points whose number of reviews is larger than 10. D_1 contains 59217 data points. D_2 contains 8455 data points. Each data point corresponds to information of the above variables for a transaction period (AM or PM) in one of the four shops.

For D_2, we would like to do further analysis to examine the impact pattern of the customer reviews. Firstly, we calculate the median for both the number of positive reviews (M_P) and the number of negative reviews (M_N). Based on these two medians, we cluster D_2 into following datasets:

D_3: data points whose number of positive review is larger than the median ($\mathrm{N}_P > \mathrm{M}_P$)
H_4: data points whose number of positive review is smaller than the median ($\mathrm{N}_P =< \mathrm{M}_P$)
D_5: data points whose number of negative review is larger than the median ($\mathrm{N}_N > \mathrm{M}_N$)
D_6: data points whose number of negative review is larger than the median ($\mathrm{N}_N < \mathrm{M}_N$)

As a first step to establish the significance of customer review information, we fitted Model 1 and Model 2 on $\mathrm{D}_1 - \mathrm{D}_6$ using OLS regression. Through comparing their adjusted R^2, we could measure the differential impact of additional customer review information on the nowcasting sales volume. If the adjusted R^2 increases after adding customer review information and the results are consistent for all the datasets, then we can conclude that customer reviews have impact on the sales volume and Model 2 will be selected. Otherwise, Model 1 will be selected.

In the second step, we would like to evaluate the performance of the selected model. We employ root-mean-square deviation (RMSD) to measure the accuracy of forecasting. In order to perform this evaluation, we divide our dataset into a training set and test set. We use the training dataset to calculate the coefficient of each variable and use test dataset to conduct model evaluation. In this study, we pick the last 10 days data as the test dataset (2015-11-19 to 2015-11-30) and the previous days data (2015-10-26 to 2015-11-18) as the training dataset. The training dataset is dynamic in this study, for example, if we would like to calculate the RMSE of forecasting on 2015-11-19, then the training dataset will be 2015-10-26 to 2015-11-18. However, if we would like calculating RMSE of forecasting on 2015-11-20, then the training dataset is 2015-10-26 to 2015-11-19. The training dataset becomes larger when we keep the forecasting sales volume. This method is commonly known as dynamic forecasting. Through comparing the RMSE of Model 0 and the selected model in step 1, we can evaluate the accuracy and predictive power of each model. Particularly, we are interested in the performance of adding big data information on sales transactions and customer reviews.

We compare the linear model without review information and linear model with review information using the adjusted R^2. OLS with review information outperforms OLS with only control variables across all of the datasets. The increase in adjusted R^2 ranges from 0.04 to 0.37 %. As a result, we select OLS with review information to perform the following analysis (Table 2).

To evaluate the prediction power of the selected model, we compare the model with the moving average result. The table below shows the RMSE of the selected model and moving

Table 2 Adj. R^2 for OLS model and OLS model with review information

	$D1$	$D2$	$D3$	$D4$	$D5$	$D6$
Adj. R^2−OLS	0.3836	0.6846	0.8322	0.6492	0.7565	0.6696
Adj. R^2−OLS with review	0.3840	0.6858	0.8343	0.6502	0.7602	0.6705
ΔAdj. R^2	0.0004	0.0012	0.0021	0.0010	0.0037	0.0009

average on each dataset and on each period. The RMSE of the selected OLS model outperform the benchmark MA model in nearly all cases, and the overall RMSE of the selected model is lower than that of MA for all six datasets (Table 3).

4.3 What is the current influence pattern of customer reviews on sales volume?

The following table shows the results of ANOVA tests across six datasets. Although the effects differ from dataset to dataset, most of the variables have a significant influence on the transaction volume. As for sales transaction related variables, AM/PM is useful for predicting the transaction volume except for items with few reviews (D4 and D6). The Stock level is proven to be influential in all six datasets. Whether the product is new or hot matters when the item receives many reviews (D3 and D5), and it does not matter when the number of reviews is less than its median. Past sales (Sold1–Sold4) are highly related to the current sales, and Double 11 does boost up the transaction volume significantly. The transaction volume is sensitive to price, only when the number of good (bad) reviews exceeds the median number of good (bad) reviews, and the brand effect is significant for items with more reviews and the whole dataset.

In the review related variables, the number of bad review that customers can see at time t is significantly related to the transaction volume on items with large numbers of reviews. The same effect holds for the number of good reviews. The average length of the review does not seem to make any difference, except that the average length for positive reviews may have some explanatory effect to the transaction volume on the whole dataset. As expected, the number of bad tags will influence the transaction volume, but the effect mainly appears in items with fewer reviews than an average item. The number of good tags may be useful on the whole dataset. Finally, the sentiment mined from the review text only has little explanatory effect on the transaction volume, and the effect only appears in items whose number of bad reviews is less than the median number of bad reviews (Table 4).

5 Discussion and conclusion

Knowledge about the demand distribution is critical for supply chain and operational decisions. For example, it is necessary to know the functional form of the demand distribution in order to apply the classical newsvendor model for inventory management (Khouja 1999). Moon and Choi (1995) assumed that the first two moments of the distribution are known, and others (Liao et al. 2011; Mostard et al. 2005) make assumptions based on the known mean and the variance of the distribution of demand. Normal, exponential and Gamma distributed demand have also been used (Kumaran and Achary 1996). Better inventory management can be achieved by evaluating the distribution model (Wiesemann et al. 2014). Therefore, it is necessary to identify the demand distribution as an input to the newsvendor model (Levi

Table 3 RMSE comparison between OLS with review information and moving average across six datasets

Period	D1		D2		D3		D4		D5		D6	
	OLS+R	MA	OLS+R	MA	OLS+R	MA	OLS+R	MA	OLS+R	MA	OLS+R	MA
20/11/2015 am	10.29	21.97	7.63	41.96	9.79	39.66	5.01	7.13	4.66	43.33	3.15	7.13
20/11/2015 pm	7.83	16.86	8.34	30.44	8.86	40.44	8.15	7.35	8.42	44.06	6.45	7.07
21/11/2015 am	9.24	16.81	9.04	31.06	11.69	40.79	4.40	7.32	11.49	44.59	4.13	7.96
21/11/2015 pm	8.91	12.00	10.38	21.59	13.17	28.06	7.37	6.03	13.13	30.78	7.06	6.44
22/11/2015 am	7.73	15.52	4.45	27.56	6.23	35.69	2.85	7.91	5.16	39.05	2.33	9.13
22/11/2015 pm	8.33	10.40	9.19	18.06	11.86	23.29	5.19	6.35	12.30	25.27	4.57	7.19
23/11/2015 am	8.61	21.29	5.27	36.65	6.18	47.95	2.70	11.49	6.14	51.83	2.47	13.33
23/11/2015 pm	8.01	16.96	9.11	29.29	11.50	38.13	5.74	10.09	11.90	41.00	5.88	11.22
24/11/2015 am	8.35	14.97	4.60	25.74	6.40	33.72	2.33	6.54	5.50	36.49	1.76	7.24
24/11/2015 pm	9.31	11.27	9.06	18.56	11.23	24.73	4.89	4.79	11.59	26.19	4.53	5.39
25/11/2015 am	14.38	14.73	23.85	25.70	35.36	37.09	8.54	8.97	38.63	40.63	8.52	8.93
25/11/2015 pm	7.36	8.02	12.76	13.34	20.64	19.38	4.86	4.89	21.56	20.96	5.26	5.07
26/11/2015 am	4.93	21.05	2.02	36.63	3.79	58.27	1.54	10.90	3.61	62.47	1.68	10.87
26/11/2015 pm	5.80	17.58	5.91	30.72	9.71	47.44	3.72	9.43	10.31	52.50	4.07	9.73
27/11/2015 am	7.17	14.31	3.85	21.79	4.80	29.36	4.71	13.09	4.63	30.51	4.78	16.08
27/11/2015 pm	4.21	7.86	3.51	11.76	4.50	16.21	3.08	6.66	4.11	16.18	2.98	8.48
28/11/2015 am	4.56	5.79	1.77	8.63	2.73	12.03	1.40	4.99	2.05	11.72	1.19	5.44
28/11/2015 pm	4.55	4.32	3.02	6.75	3.46	9.05	2.78	4.21	3.23	9.86	2.50	4.23
29/11/2015 am	4.78	3.61	1.41	6.12	2.68	8.71	1.14	2.61	2.32	9.38	0.83	2.99
29/11/2015 pm	4.06	3.44	2.98	5.79	4.26	8.35	2.54	1.91	4.20	8.74	2.27	2.32
30/11/2015 am	5.03	5.35	2.20	9.28	3.16	13.40	1.21	2.91	3.11	14.08	0.82	2.93
Overall	7.78	13.95	8.27	24.27	11.35	32.49	3.83	7.47	11.70	35.16	4.09	8.31

Table 4 F-values and significance level for ANOVA test across six datasets

	D1	D2	D3	D4	D5	D6
AM/PM	15.48***	14.53***	63.3***	0.93	48.95***	1.95
stock_1	3114.16***	412.77***	244.64***	394.54***	74.74***	415.35***
newHot	6.99***	0.94	10.49***	1.35	12.99***	0.88
sold_1	7643.81***	1487.45***	1303.74***	696.29***	1297.03***	693.56***
sold_2	1491.75***	176.06***	16***	185.07***	0.15	283.38***
sold_3	139.84***	62.47***	52.18***	102.79***	71.79***	5.65**
sold_4	147.79***	23.47***	176.76***	4.79*	116.04***	1.78
db11	18408.29***	16176.23***	19201.69***	6610.24***	9379.66***	8743.53***
price	0.36	0.81	8.8***	0.13	11.52***	0.02
shop_id	9.56***	3.17*	6.73***	0.65	6.18***	0.48
badRev_1	32.42***	15.88***	56.1***	0.67	46.22***	0.1
goodRev_1	0.06	10.37**	41.8***	0.03	31.13***	0.38
numWord_neg_1	0.06	0.93	0.09	1.67	0.09	0.75
numWord_pos_1	6.11**	0	2.52	0.01	0.06	0
badTag_1	2.35	14.43***	1	21.41***	0.15	26.61***
goodTag_1	3.97**	0.52	0.02	1.46	0.37	0.06
cumSent_1	0.74	0	1.22	0.6	5.03**	0.61

*** 99% confidence level, ** 95% confidence level, * 90% confidence level

et al. 2015; Choy and Cheong 2011) for accurate decision making (Olivares et al. 2008). Levi et al. (2015) proposed that if a distribution has a large weighted mean spread, then the sample average approximation (SAA) solution is near-optimal with relatively few samples to draw from the true demand distribution from random samples. Choy and Cheong (2011) suggested that demand the distribution should be identified prior to forecasting product demand for successful supply chain management. In this study, we firstly discovered that the datasets of Tmall conform to the lognormal distribution, which can predict the sales more precisely. In addition, we suggest that a methodology to establish the significance of big data information in sales nowcasting models be based on empirical comparisons.

Similar to the research of Chong et al. (2015), customer reviews have significant relationships for sales prediction. In this study, we further reveal the detailed information content of these reviews and find that brand value, price, and product type—the three critical managerial variables—are only important when a product gets sufficient customer reviews. The number of negative reviews has significant impact unless the total number of reviews is low. The number of positive reviews is important only when we have a good total number of reviews. The actual content of the customer reviews (i.e. the average number of words in a review and the sentiment mined from the review text) is not critical in most cases (except for the average length of positive reviews in the "Whole" dataset and average sentiment of customer reviews in D3). The number of bad reviews as classified by Tmall (badTag_1) is significant for nowcasting when you have a good number of reviews in total (D2)—but out of which a smaller number contributes to either being positive or negative—that is, only when the distribution of reviews is unbalanced (D4 and D6). Finally, the number of good reviews as classified by Tmall (goodTag_1) is only important in the general case (D1).

The findings from this research offer great insights for the managerial implications in following aspects. First, sales of online stores can be nowcasted with customers' reviews from the online platform with even the simple models we proposed; second, with the real time sales nowcasting, it offers an optimized solution for the online stores to manage their inventory, therefore reducing the cost of stocking and avoiding the under stocking; third, the online stores should encourage consumers to offer the reviews of their products, since the number of reviews is essential for sales nowcasting. Forth, it is beneficial for marketing managers to understand the popularity of different products so as to plan the corresponding marketing activities, such as advertisements, for the products and new products. Fifth, it helps finance managers to modify investment activities, and develop different strategies and plans based on the nowcasting results, so as to develop the short to medium term financial goals of the organization.

Moreover, big data based high frequency nowcasting can serve both the purposes of getting good short term forecasts to support the relevant supply chain operation decisions, as well as understanding the current form and influence of possible management intervention. The methodology to do so, however, is not the same for both the forecasting and understanding tasks. To understand the current influence pattern of the impact of possible management interventions, a different strategy is needed.

In our dataset, the numbers of customer reviews across product items fit well into a long tailed distribution. This implies a significant proportion, much higher than the 80/20 Pareto Principle would suggest, of product items has a good number of customer reviews. This has the implication that the effect from reviews as a whole is much higher than if the customer reviews number distribution is exponential. After evaluating a sales forecasting algorithm with real data, Chern et al. (2015) found that the proposed method was more effective when the online reviews were adequate. However, the influence of a single review was continuous but was decaying over time. The overwhelming significance of the current stock level highlights the importance of inventory and logistics management for online businesses.

To conclude, this study shows that the combination of information of big data streams of customer reviews and transactions is more effective than using the data separately on sales nowcasting. This study contributes to supply chain management under the background of big data in the following ways: first, we construct a methodology to visualize the distributional characteristics of demand at the product level, which is critical to models for supply chain management, such as the classic newsvendor model, from online big data streams. It offers insights for further research into using this new information. Second, we tested whether there is significant useful information of big data streams of customer reviews for sales nowcasting, and found it to be positive. Third, we analyzed the patterns of influences of online sentiment on sales. Big data streams overall seem to be a promising source for better sales nowcasting and deeper understanding of the demand structure. With online sales data available, it is theoretically possible, and perhaps useful, to sample observations at a high time frequency, such as half-daily. This presents an additional possibility for mangers to deal with the ever faster changing customer demand and ever shorter product life cycles. This work demonstrates that such high frequency data may actually contain useful information, albeit there may be noise as well. The short-run customer review data in fact improve the ability of nowcasting even for simple regression models.

In the current study, we proposed several simple models for nowcasting sales of online stores. This is a pioneer study for testing the significance of big data and should be useful for nowcasting. For future studies, researchers may proceed from there to produce more accurate models for sales nowcasting in other online markets. It is suggested to apply machine learning approaches, such as artificial neural networks, Bayesian networks, support vector machines, and genetic algorithms, to nowcast the sales in online stores. The data were collected from 25th October 2015 to 30th November 2015, which is a peak season in online stores. It is also suggested carrying out analysis of using different dataset throughout the year, so as to have a deeper understanding of the effects of online big data streams on sales.

References

AgilOne. (2014). *AgilOne posts new data-driven marketing survey results*. http://search.proquest.com.ezproxy.lb.polyu.edu.hk/docview/1476226999?OpenUrlRefId=info:xri/sid:primo&accountid=16210.

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, *60*, 255–265.

Amornpetchkul, T., Duenyas, I., & Şahin, Ö. (2015). Mechanisms to induce buyer forecasting: Do suppliers always benefit from better forecasting? *Production and Operations Management*, *24*, 1724–1749.

Antipa, P., Barhoumi, K., Brunhes-Lesage, V., et al. (2012). Nowcasting German GDP: A comparison of bridge and factor models. *Journal of Policy Modeling*, *34*, 864–878.

Babu, M. S. P., Sastry, S. H., IEEE. (2014). Big data and predictive analytics in ERP systems for automating decision making process. *2014 5th IEEE international conference on software engineering and service science (ICSESS)*, pp 259–262.

Balar, A., Malviya, N., Prasad, S., & Gangurde, A. (2013). Forecasting consumer behavior with innovative value proposition for organizations using big data analytics. In *2013 IEEE international conference on computational intelligence and computing research (ICCIC)* (pp. 1–4). IEEE.

Banbura, M., Giannone, D., Modugno, M., & Reichlin, L. (2013). *Now-casting and thereal-time data flow*. European Central Bank (ECB), Working Paper No. 1564.

Banbura, M., Giannone, D., & Reichlin, L. (2011). *Nowcasting with daily data*. European Central Bank, Working Paper.

Booth, E., Mount, J., & Viers, J. H. (2006). Hydrologic variability of the Cosumnes River floodplain. *San Francisco Estuary and Watershed Science* 4.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information Communication and Society*, *15*, 662–679.

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

Buhl, H. U., Roglinger, M., Moser, F., et al. (2013). Big Data a fashionable topic with(out) sustainable relevance for research and practice?(Editorial). *Business and Information Systems Engineering*, *5*, 65.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). Berlin: Springer.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods and Research*, *33*, 261–304.

Camacho, M., & Martinez-Martin, J. (2014). Real-time forecasting US GDP from small-scale factor models. *Empirical Economics*, *47*, 347–364.

Carriere-Swallow, Y., & Labbe, F. (2013). Nowcasting with Google trends in an emerging market. *Journal of Forecasting*, *32*, 289–298.

Chen, Y. J., & Xiao, W. (2012). Impact of reseller's forecasting accuracy on channel member performance. *Production and Operations Management*, *21*, 1075–1089.

Chern, C.-C., Wei, C.-P., Shen, F.-Y., & Fan, Y. N. (2015). A sales forecasting model for consumer products based on the influence of online word-of-mouth. *Information Systems and e-Business Management*, *13*(3), 445–473.

Choi, H., & Varian, H. (2012). Predicting the present with google trends. *Economic Record*, *88*, 2–9.

Chong, A. Y. L., Ch'ng, E., Liu, M. J., & Li, B. (2015). Predicting consumer product demands via Big Data: the roles of online promotional marketing and online reviews. *International Journal of Production Research*. doi:10.1080/00207543.2015.1066519.

Choy, M., Cheong, ML. (2011). Identification of demand through statistical distribution modeling for improved demand forecasting. arXiv:1110.0062

Christopher, M., & Ryals, L. J. (2014). The supply chain becomes the demand chain. *Journal of Business Logistics*, *35*, 29–35.

Chung, C., Niu, S.-C., & Sriskandarajah, C. (2012). A sales forecast model for short-life-cycle products: New releases at blockbuster. *Production and Operations Management*, *21*, 851–873.

Cox, M., Ellsworth, D. (1997). Application-controlled demand paging for out-of-core visualization. *Proceedings of the 8th conference on Visualization'97*. IEEE Computer Society Press, 235-ff.

Cui, G., Lui, H.-K., & Guo, X. (2012). The effect of online consumer reviews on new product sales. *International Journal of Electronic Commerce*, *17*, 39–58.

Dias, F., Pinheiro, M., & Rua, A. (2015). Forecasting Portuguese GDP with factor models: Pre- and post-crisis evidence. *Economic Modelling*, *44*, 266–272.

Ekbia, H., Mattioli, M., Kouper, I., et al. (2015). Big data, bigger dilemmas: A critical review. *Journal of the Association for Information Science and Technology*, *66*, 1523–1545.

Fang, H., Zhang, Z. Y., Wang, C. J., et al. (2015). A survey of big data research. *IEEE Network*, *29*, 6–9.

Felix S. (2015). Top online marketplaces for small businesses selling internationally. *The Endica Blog*. http://online-shipping-blog.endicia.com/top-online-marketplaces-for-small-businesses-selling-internationally/

Guo, Z., Wong, W. K., & Li, M. (2013). A multivariate intelligent decision-making model for retail sales forecasting. *Decision Support Systems*, *55*, 247–255.

Hirashima, A., Jones, J., Bonham, CS., et al. (2015). Nowcasting tourism industry performance using high frequency covariates (No. 2015-3). University of Hawaii Economic Research Organization, University of Hawaii at Manoa.

Huang, T., & Van Mieghem, J. A. (2014). Clickstream data and inventory management: Model and empirical analysis. *Production and Operations Management*, *23*, 333–347.

Johansson, M. A., Powers, A. M., Pesik, N., Cohen, N. J., & Staples, J. E. (2014). Nowcasting the spread of chikungunya virus in the Americas. *PloS one*, *9*(8), e104915.

Khouja, M. (1999). The single-period (news-vendor) problem: Literature review and suggestions for future research. *Omega*, *27*, 537–553.

Kim, W., Won, J. H., Park, S., & Kang, J. (2015). Demand forecasting models for medicines through wireless sensor networks data and topic trend analysis. *International Journal of Distributed Sensor Networks*, *2015*, 36.

Kumaran, M., & Achary, K. K. (1996). On approximating lead time demand distributions using the generalised λ-type distribution. *Journal of the Operational Research Society*, *47*(3), 395–404.

Lampos, V., Miller, AC., Crossan, S., et al. (2015). Advances in nowcasting influenza-like illness rates using search query logs. *Scientific Reports* 5.

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note*, *6*, 70.

Lassen, NB., Madsen, R., Vatrapu, R. (2014). Predicting iPhone Sales from iPhone Tweets. In: Reichert, M., Rinderle-Ma, S. and Grossmann, G. (Eds.), *Proceedings of the 2014 IEEE 18th international enterprise distributed object computing conference*, pp 81–90.

Levi, R., Perakis, G., & Uichanco, J. (2015). The data-driven newsvendor problem: New bounds and insights. *Operations Research*, *63*(6), 1294–1306.

Li, J. R., Tao, F., Cheng, Y., et al. (2015). Big data in product lifecycle management. *International Journal of Advanced Manufacturing Technology*, *81*, 667–684.

Liao, Y., Banerjee, A., & Yan, C. (2011). A distribution-free newsvendor model with balking and lost sales penalty. *International Journal of Production Economics*, *133*, 224–227.

Lu, C.-J., & Chang, C.-C. (2014). A hybrid sales forecasting scheme by combining independent component analysis with K-means clustering and support vector regression. *The Scientific World Journal*, *55*, 231–238.

Ma, Q., & Zhang, W. (2015). Public mood and consumption choices: Evidence from sales of sony cameras on taobao. *PloS one*, *10*(4), e0123129.

McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*, *90*, 60–68.

Mishra, B. K., Raghunathan, S., & Yue, X. (2009). Demand forecast sharing in supply chains. *Production and Operations Management*, *18*, 152–166.

Moon, I., & Choi, S. (1995). The distribution free newsboy problem with balking. *Journal of the Operational Research Society*, *46*(4), 537–542.

Mostard, J., De Koster, R., & Teunter, R. (2005). The distribution-free newsboy problem with resalable returns. *International Journal of Production Economics*, *97*, 329–342.

Olivares, M., Terwiesch, C., & Cassorla, L. (2008). Structural estimation of the newsvendor model: an application to reserving operating room time. *Management Science*, *54*, 41–55.

Osadchiy, N., Gaur, V., & Seshadri, S. (2013). Sales forecasting with financial indicators and experts' Input. *Production and Operations Management*, *22*, 1056–1076.

Puts, M., Daas, P., & de Waal, T. (2015). Finding errors in big data. *Significance*, *12*, 26–29.

Sanders, N. R., & Ganeshan, R. (2015). Special issue of production and operations management on big data in supply chain management. *Production and Operations Management*, *24*, 852–853.

Snijders, C., Matzat, U., & Reips, U.-D. (2012). Big data: Big gaps of knowledge in the field of internet science. *International Journal of Internet Science*, *7*, 1–5.

Su, Z. F., Wang, X., & He, K. (2014). Nowcasting and short-term forecasting of Chinese quarterly GDP: Mixed frequency approach. *Anthropologist*, *17*, 53–63.

Tan, K. H., Zhan, Y., Ji, G., et al. (2015). Harvesting big data to enhance supply chain innovation capabilities: An analytic infrastructure based on deduction graph. *International Journal of Production Economics*, *165*, 223–233.

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin and Review*, *11*, 192–196.

Waller, M. A., & Fawcett, S. E. (2013). Click here for a data scientist: Big data, predictive analytics, and theory development in the era of a maker movement supply chain. *Journal of Business Logistics*, *34*, 249–252.

Walsh, B. (2014). Google's Flu Project shows the failings of big data. *Time.com*: 1.

Weinberger, D. (2014). *Too big to know: Rethinking knowledge now that the facts aren't the facts, experts are everywhere, and the smartest person in the room is the room*. New York: Basic Books.

Wiesemann, W., Kuhn, D., & Sim, M. (2014). Distributionally robust convex optimization. *Operations Research*, *62*, 1358–1376.

Yang, L., Xiangji, H., & Aijun, A. (2007). A sentiment-aware model for predicting sales performance using blogs. *Proc SIGIR.* pp. 607–615.

Yu, Y., Choi, T.-M., & Hui, C.-L. (2011). An intelligent fast sales forecasting model for fashion products. *Expert Systems With Applications*, *38*, 7373–7379.

Zhou, Y., Wei, M., Cheng, Z. J., et al. (2013). The wind and temperature information of AMDAR data applying to the analysis of severe weather nowcasting of airport. *International Conference on Information Science and Technology*, *2013*, 1005–1010.