

## **Mining for Signals of Future Consumer Expenditure on Twitter and Google Trends**

**Pekar, Viktor**

Finance Department, Business School, University of Birmingham, United Kingdom.

---

### ***Abstract***

*Consumer expenditure constitutes the largest component of Gross Domestic Product in developed countries, and forecasts of consumer spending are therefore an important tool that governments and central bank use in their policy-making. In this paper we examine methods to forecast consumer spending from user-generated content, such as search engine queries and social media data, which hold the promise to produce forecasts much more efficiently than traditional surveys. Specifically, the aim of the paper is to study the relative utility of evidence about purchase intentions found in Google Trends versus those found in Twitter posts, for the problem of forecasting consumer expenditure. Our main findings are that, firstly, the Google Trends indicators and indicators extracted from Twitter are both beneficial for the forecasts: adding them as exogenous variables into regression model produces improvements on the pure AR baseline, consistently across all the forecast horizons. Secondly, we find that the Google Trends variables seem to be more useful predictors than the semantic variables extracted from Twitter posts, the differences in performance are significant, but not very large.*

**Keywords:** *Google Trends and Search Engine data; Social media and public opinion mining; Internet econometrics; Machine learning econometrics; Consumer behavior, eWOM and social media marketing.*

---

## **1. Introduction**

Consumer expenditure constitutes the largest component of Gross Domestic Product in developed countries: in the US, it accounts for about 70% of GDP, in the UK 66%, in Germany 60% (Pistaferry, 2015). Significant changes to consumer spending are key to predict the depth of a recession or the speed of recovery, and central banks use consumer spending forecasts as an important tool for monetary policy-making.

Government institutions and market research agencies compile their consumer spending indices on a regular basis. Among best-known examples are the University of Michigan Consumer Sentiment Index for the US, or the Household Final Consumption Expenditure by the UK Office for National Statistics. Currently, such indices are measured by market research surveys, but these have significant drawbacks: they are expensive to organize, they have sampling problems, the amount of effort required to collect and compile the data often entails that the indices are out of date by the time they are published.

This paper examines the hypothesis that user-generated content, such as search engine queries or social media posts, offers a better alternative to traditional surveys when it comes to estimating consumer expenditure. Effective methods to extract signals about future consumer spending from this data may help to produce forecasts more efficiently, based on much larger data samples, and in near-real time.

Previous work studied models of consumer spending trained on search engine data, based on the intuition that web searches for product names indicate intended purchases (Vosen and Schmidt, 2011; Scott and Varian, 2015; Wu and Brynjolfsson, 2015). Another direction of research has been to estimate economic confidence and purchase intentions of consumers from social media using automatic sentiment analysis (O'Connor et al., 2010; Daas and Puts, 2014; Najafi and Miller, 2016).

In this paper we study the relative utility of evidence about purchase intentions found in search engine queries versus those found in social media, for the problem of forecasting consumer expenditure.

## **2. Google Trends**

The Google Trends (GT) site provides data on the volume of all Google queries based on geographic locations and time, collected since 2004. The frequencies of queries is not the absolute number of actual queries, but a normalized index, such that for any given retrieval criteria, the index is always between 0 and 100, 100 being the count of the most common query in the retrieved data.

GT contains data not only on individual queries, but also on categories of queries. In our study we use the data on search volumes of the 18 subcategories of the top-level "Shopping" category in GT. Examples of the subcategories are "Apparel", "Consumer Electronics", "Luxury Goods", "Ticket Sales". The search volume on each category is used as an exogenous variable in the Support Vector Regression model.

The time period we analyse spans 43 months (from 1st January 2014 to 31.07.2017). Because GT returns only weekly search volumes in one request for periods longer than six months, we are able to retrieve only weekly search volumes for the entire time period. To obtain daily search volumes, we first retrieve daily data in separate queries for each 6-month period. Then, within each such subset, we fit a linear regression on the monthly data and use it to obtain daily volumes for the entire 43 months dataset.

### **3. Purchase intentions on Twitter**

Our method aims to predict a consumer spending index from the mentions of purchase intentions in Twitter posts. The method consists of the following steps. First, tweets mentioning a purchase intention are identified. Second, noun phrases referring to the objects of the intended purchases are extracted and represented as semantic vectors using the word2vec method. Finally, a regression model of the consumer spending index is trained that uses semantic vectors as explanatory variables. These steps are detailed below.

#### ***3.1. Detecting purchase intentions***

To obtain tweets mentioning purchase intentions, we issue a set of queries to the tweet collection, which are meant to capture common ways to express an intention to buy something. They are created from combinations of (1) first-person pronouns ("I" and "we"), (2) verbs denoting intentions ("will", "I", "be going to", "be looking to", "want to", "wanna", "gonna"), and (3) verbs denoting purchase ("buy", "shop for", "get oneself"), thus obtaining queries such as "I will buy" or "we are going to buy".

The text of each tweet is then processed with a part-of-speech tagger. PoS tag patterns are then applied to extract the head noun of the noun phrase following the purchase verb (e.g., "headphones" in "I am looking to buy new headphones"). After that, daily counts of the head nouns are calculated.

#### ***3.2. Semantic vectors***

To represent the semantics of the nouns, we use the word2vec method (Mikolov et al., 2013) which has proven to produce accurate approximations of word meaning in different NLP tasks. A word2vec model is a neural network that is trained to reconstruct the linguistic context of words. The model is built by taking a sequence of words as input and

learning to predict the next word, using a feed-forward topology where a projection layer in the middle is taken to constitute a semantic vector for the word, after connection weights have been learned. The semantic vector is a fixed-length, real-valued pattern of activations reaching the projection layer. For each word, the input text originally has a dimensionality equal to the vocabulary size of the training corpus (typically millions of words), but the semantic modelling provides reduction to the size of the vector (typically several hundreds).

For each date, we map each noun that was observed on that day to a semantic vector, using 100-dimensional word2vec vectors trained on a large corpus of Twitter posts. The semantic vectors of all the nouns for each day are then averaged to obtain a single vector. The components of the vectors will then be used as exogenous variables in regression models.

To allow for some time between the stated purchase intention and the actual purchase, we experiment with the “intention lag”, different numbers of days between the day on which intentions were registered and the day for which the value of the consumer spending index is predicted.

## 4. Experiments

### 4.1. Data

**Indicator of Consumer Expenditure.** As the forecast variable in our model, we use the Gallup Consumer Spending Index (CSI)<sup>1</sup>. The index represents the average dollar amount Americans report spending on a daily basis. The eventual index is presented as a 3-day and a 14-day rolling averages of these amounts. In our evaluation, we used the 3-day values of CSI, between January 1, 2014 and July 31, 2017, i.e. 1,310 days in total.

**Twitter.** For the same time period, we collected Twitter posts that originate from the US and that express intentions to buy, obtaining the total of 288,730 messages. Counts of nouns referring to purchases were extracted and rolling averages for each noun for three-day periods were calculated. To eliminate noisy data, we selected the 1000 most common nouns to construct semantic vectors.

**Google Trends.** Also for the same period, we obtain frequencies of searches in the 18 subcategories of the top-level Shopping category from the GT site, limiting the data to the US.

**Train-validation-test split.** The available data was divided into the training, validation and test parts, in proportion 60%-20%-20%.

---

<sup>1</sup> <http://www.gallup.com/poll/112723/gallup-daily-us-consumer-spending.aspx>

#### **4.2. Modelling strategies**

We experiment with four methods to ensure stationarity of the time series data: differencing, detrending, seasonal adjustment and detrending with seasonal adjustment. Detrending and seasonal adjustment are performed using the STL method (Cleveland et al., 1990). Before evaluating the quality of forecasts on test data, the forecasts are de-differenced, and the trend and the seasonal component estimated on training data are added to the forecasts.

#### **4.3. Support vector regression**

The Support Vector Machines learning algorithm (Cortes and Vapnik, 1995) is one of the most popular machine learning methods for supervised learning. In our experiments we use Support Vector Regression (SVR), a version of SVM adapted for regression. During evaluation, we experimentally determine free parameters of SVR (the cost parameter, the gamma parameter and the kernel type) on a validation dataset using the grid search technique. The model with the best parameter configuration is then evaluated on the test set.

#### **4.4. Evaluation method**

Once a model was trained on the training set and its parameters optimized on the validation set, it was evaluated on the test set using dynamic forecasting: given the first day  $t$  of the test set, and the forecast horizon  $h$ , the model predicted  $h$  days in the future, for each day from  $t_2$  to  $t_h$  the values predicted by the model for previous days were input as endogenous variables. In the following, we report results for  $h = 7, 14,$  and  $28$ . As evaluation metrics, we use the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE).

As the baselines, we use SVR models trained with the same algorithms but only on endogenous variables, i.e. lagged values of CSI. Because CSI displayed weekly seasonality, we used seven lagged variables in the baseline model.

### **5. Results and discussion**

#### **5.1. Modelling strategies**

An inspection of the correlogram of the CSI time series suggested that it is likely to have weekly seasonality. Furthermore, considering the long time period the data covers, the data may contain a trend. Therefore we first examined the effect of different techniques to “whiten” the time series on the quality of the forecasts. Table 1 details the results (the baseline refers to the raw original data, the best RMSE and MAE scores are in bold).

**Table 1. Forecast accuracy for different time series transformation methods.**

	<b>h=7</b>		<b>h=14</b>		<b>h=28</b>	
	<b>RMSE</b>	<b>MAE</b>	<b>RMSE</b>	<b>MAE</b>	<b>RMSE</b>	<b>MAE</b>
Baseline	12.78	9.9	14.42	11.23	14.42	11.38
Differencing	12.11	9.57	21.51	18.0	11.77	9.04
Detrending	10.64	8.18	11.49	8.9	11	8.42
Deseasonalizing	12.62	9.85	14.57	11.36	14.35	11.33
Detrend+Deseason	<b>10.5</b>	<b>7.96</b>	<b>11.48</b>	<b>8.64</b>	<b>10.96</b>	<b>8.23</b>

These results show that applying both detrending and seasonal adjustment consistently resulted in the best forecasting results, for all forecasting horizons. Therefore, in the subsequent experiments, the CSI data was detrended and deseasonalized.

### 5.2. Effect of Google Trends variables

We next examined the effect of supplying GT data as exogenous variables into the regression model, in addition to the autoregressive variables. The results are shown in Table 2 (improvements on the baseline are in bold).

We find that the GT variables do often perform better than the purely endogenous baseline, for all the three forecast horizons. It appears also that shorter intentions lags (between 0 and 4) produce better quality models. The best-performing model is a lag of one model, which beats the baseline by 3-7% across all the horizons.

**Table 2. Forecast accuracy with GT variables.**

Intention lag	h=7		h=14		h=28	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
0	<b>10.49</b>	8.03	<b>10.9</b>	<b>8.19</b>	<b>10.57</b>	<b>8.0</b>
1	<b>10.21</b>	<b>7.69</b>	<b>10.81</b>	<b>8.06</b>	<b>10.52</b>	<b>7.85</b>
2	<b>10.4</b>	<b>7.64</b>	<b>10.78</b>	<b>7.97</b>	<b>10.7</b>	<b>7.98</b>
3	10.56	8.0	<b>10.78</b>	<b>8.02</b>	<b>10.68</b>	<b>8.02</b>
4	10.84	8.3	<b>11.15</b>	<b>8.34</b>	11.5	8.63
5	10.57	8.06	11.52	8.65	11.62	8.73
6	10.46	7.99	11.48	8.65	<b>10.9</b>	8.24
7	11.19	8.58	<b>11.28</b>	<b>8.39</b>	11.09	8.33

### 5.3. Effect of Twitter variables

Table 3 presents results on the effect of semantic variables extracted from Twitter posts. As with GT variables, improvements are found across all the horizons. However, the baseline is consistently outperformed only when the intention lag is 0, and the improvements are more modest, ranging between 1 and 5%.

Comparing the performance of the models with GT variables and with Twitter variables, we observe that the GT model tends to fare better, but the gain on the Twitter model is not more than 0.3 points (3.6%) in either RMSE and MAE. Still, the differences in forecasts between the two types of models at corresponding horizons are statistically significant.

## 6. Conclusions

In this paper we presented a study comparing indicators of purchase intentions obtained from Google Trends to those obtained from Twitter using NLP analysis of the messages, on the task of forecasting consumer expenditure. Our main findings are that, firstly, both kinds of purchase intention indicators are beneficial for the forecasts: the improvements on the baseline are consistent across all the forecast horizons and in terms of both evaluation metrics. Secondly, the study found the Google Trends variables seem to be more useful predictors than the semantic variables extracted from Twitter posts, although the differences in performance are not very large.

**Table 3. Forecast accuracy with Twitter variables.**

Intention lag	h=7		h=14		h=28	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
0	<b>10.34</b>	<b>7.76</b>	<b>10.92</b>	<b>8.21</b>	<b>10.88</b>	<b>8.15</b>
1	10.73	7.99	<b>11.37</b>	<b>8.49</b>	11.09	8.29
2	10.75	8.06	<b>11.46</b>	<b>8.61</b>	11.14	8.37
3	10.57	8.01	<b>11.3</b>	<b>8.55</b>	10.97	8.34
4	10.65	8.25	<b>11.33</b>	<b>8.57</b>	11.04	8.42
5	10.71	8.03	<b>11.31</b>	<b>8.46</b>	11.51	8.67
6	10.92	8.15	<b>11.32</b>	<b>8.43</b>	11.52	8.69
7	10.86	8.14	11.54	8.68	11.53	8.7

Future directions for this work may involve a further analysis of models that use Google Trends data, such as an analysis of more fine-grained Google Trends subcategories, automatic selection of the most relevant predictors among them, and their semantic clustering.

## References

- Cleveland R., Cleveland W., McRae J., & Terpenning I. (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics*, Vol.6, No.1, 1990. pp. 3–73.
- Cortes C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*. 20 (3): 273–297.
- Daas P., & Puts M. (2014). Social media sentiment and consumer confidence. In *Workshop on using Big Data for forecasting and statistics*.
- Mikolov T., Chen K., Corrado K., & Dean J. (2013). Efficient estimation of word representations in vector space. In *Proceedings of CoRR*.
- Najafi H. and Miller D. (2016). Comparing analysis of social media content with traditional survey methods of predicting opening night box-office revenues for motion pictures. *Journal of Digital and Social Media Marketing*, 3(3):262–278.

- O'Connor B., Balasubramanian R., Routledge R., & Smith N. (2010). From tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings of ICWSM*.
- Pistaferri L. (2015). Household consumption: Research questions, measurement issues, and data collection strategies. *Journal of Economic and Social Measurement*.
- Scott S. & Varian H. (2015). Bayesian Variable Selection for Nowcasting Economic Time Series. In *Economic Analysis of the Digital Economy*, pages 119–135. University of Chicago Press.
- Vosen S. & Schmidt T. (2011). Forecasting private consumption: survey-based indicators vs. Google trends. *Journal of Forecasting*, 30(6):565–578.
- Wu L. & Brynjolfsson E. (2015). The future of prediction: How Google searches foreshadow housing prices and sales. In *Economic Analysis of the Digital Economy*, pages 89–118. University of Chicago Press.