POMS PRODUCTION AND OPERATIONS MANAGEMENT

PRODUCTION AND OPERATIONS MANAGEMENT

Vol. 27, No. 10, October 2018, pp. 1770–1774
ISSN 1059-1478 | EISSN 1937-5956 | 18 | 2710 | 1770

DOI 10.1111/poms.12839

Can Google Trends Improve Your Sales Forecast?

Tonya Boone, Ram Ganeshan*
Raymond A. Mason School of Business, The College of William and Mary, Williamsburg, Virginia 23185, USA,
tonya.boone@mason.wm.edu, ram.ganeshan@mason.wm.edu

Robert L. Hicks
Department of Economics, The College of William and Mary, Williamsburg, Virginia 23185, USA, robhicks@mykolab.com

Nada R. Sanders
D'Amor-McKim School of Business, Northeastern University, Boston, Massachusetts 02115, USA, n.sanders@northeastern.edu

In this issue, Cui et al. (2018) show how the quantity and quality of user-generated Facebook data can be used to enhance product forecasts. The intent of this note is to show how another type of user-generated content—customer search data, specifically one obtained from Google Trends—can be used to reduce out-of-sample forecast errors. Based on our work with an online retailer, we bolster Cui et al. (2018) result by showing that adding customer search data to time series models improves out-of-sample forecast errors.

Key words: forecasting; search queries; big data
History: Received: June 2017; Accepted: October 2017 by Kalyan Singhal, after 1 revision.

1. Introduction

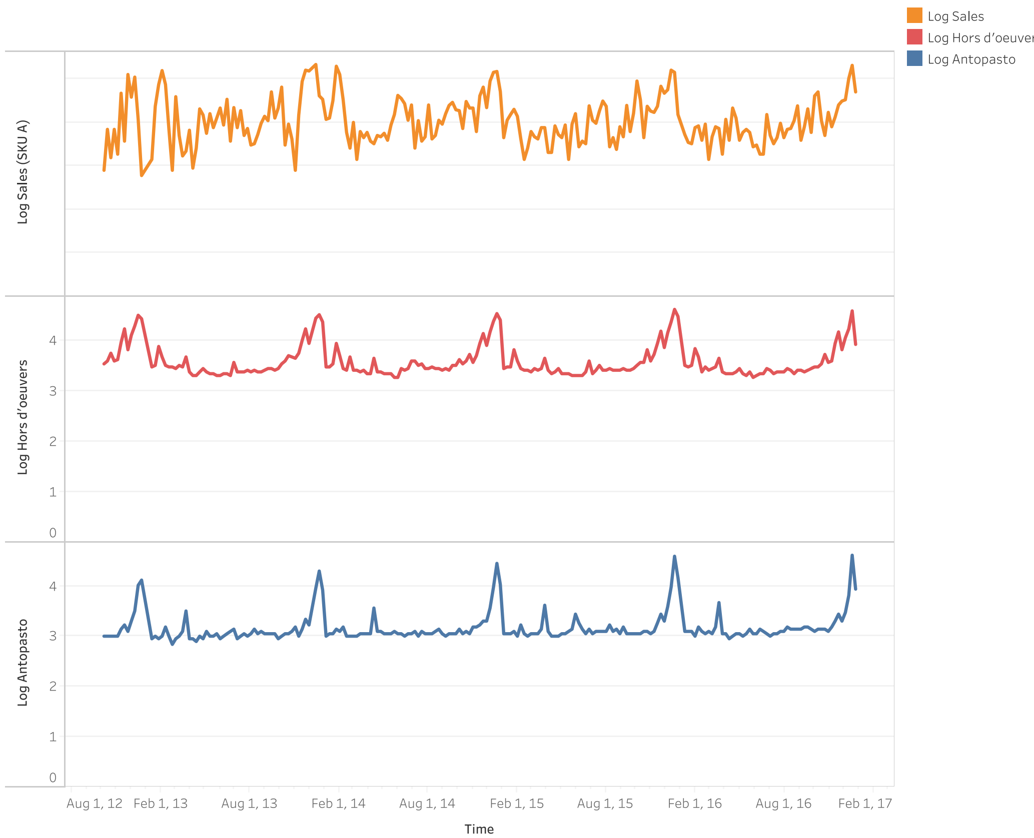
Cui et al. (2018) in this issue shows how user-generated content, specifically user interactions on Facebook, can be used to improve product forecasts. Their social media-enhanced forecasting models show a relative out-of-sample accuracy improvements anywhere from 12.85% to 23.23%. The intent of this note is to bolster the primary idea in Cui et al. (2018) that user-generated content can be helpful in improving product level forecasts. We have been working with an online retailer of speciality food over the last several years. Using operational data provided by the retailer, we explore how another form of user-generated content—customer search queries—can be used to enhance product forecasts. Specifically, we use Google Trends data to incorporate customer queries into our forecasting models.

Google Trends provides free and fine-grained data on customer queries on search terms. It reports the data as an index—the proportion of queries of a particular search term as a fraction of total search in a given geographic region or time frame. The addition of Google Trends' queries to time series models has shown to improve *out-of-sample* forecast errors in a wide variety of contexts. They include predicting economic parameters such as unemployment and consumer sentiment (Choi and Varian 2009, 2012), GDP estimates (Castle et al. 2009), housing (Limnios and You 2016), flu outbreaks (Ginsberg et al. 2009), and financial markets (Dimpfl and Jank 2016). Boone et al. (2015) first posited that Google searches improve *in-sample* forecasts but to the best of our knowledge this note is the first to demonstrate *out-of-sample* performance across multiple product categories.

In our context, we study if the search volumes for certain search terms improve the sales forecasts of specific products. The premise is that search for a certain term in our context of online retail shows an intent to explore and potentially buy the product. Our data are from a speciality food retailer. Figure 1 shows the logged Sales of stock keeping unit (SKU) A, a speciality food item bought and used as an appetizer. The figure also plots the logged values of the search trend indices for the terms "Antipasto" and "Hors d'oeuvres" often used to describe this product. The trends are reported on a 1–100 scale, where a value of 100 indicates peak popularity of a search term whereas 50 indicates that it is half of its peak value. The chart shows how the sales of this SKU matches up with the popularity of these two search terms. For example, sales for SKU A peaks around the Thanksgiving season, about the same time when the searches for the terms "Antipasto" and "Hors d'oeuvres" peak.

Can this query information be used to improve forecast errors? Our results show that the addition of Google Trends data decreases out-of-sample mean absolute percentage errors (MAPE) anywhere from 2.2% to 7.7% for the SKUs we studied.

Figure 1 Sales of SKU A vs. Google Trends [Color figure can be viewed at wileyonlinelibrary.com]



In section 2, we detail our data, models, and compare our baseline model with the Google trend-enhanced model. In section 4 are some of our conclusions.

2. Data and Models

We illustrate our models using data from a speciality retailer of food and cookware. This retailer identifies as a supplier of high quality speciality food items and also stocks a wide variety of cookware. We were given access to sales and pricing data for a number of stock keeping units, ranging from speciality meats, spices, cookware, wine and cheese. We chose five SKUs, including SKU A, that had at least four years of sales and pricing data. For customer query data, we used Google Trends' search index for specific search terms, publicly available from Google's web site.[1] The query terms we use in this note are the most likely candidates that may lead the customer to the SKU as reported by the business owner. To estimate and compare models, we divide the data into training and testing sets. Training sets are all data points prior to January 1, 2016. Data from 2016 and onwards is used to test out-of-sample performance of our models.

2.1. Models

For the base forecasting model, we use an ARIMA (4,0,0) model. The dataset $t = 1, \ldots, T$ is divided into the training set $t = 1, \ldots, t_r$ and validation set $t = t_{r+1}, \ldots, t_T$. If S_t is the sales of any given SKU in week t and p_t is the price of the SKU in week t,

$$\ln S_t = K + \beta_0 \ln p_t + \sum_{j=1}^{4} \beta_j \ln S_{t-j} + \beta_5 D_C + \epsilon_t \quad (1)$$

where K is a constant and ε_t is the error. D_C is a dummy variable that captures the increased sales during the Christmas season. β_0 through β_5 are parameters that we are estimating.

To incorporate customer search data, we add the Google query index to our model (the "trend" model from now on).

$$\ln S_t = K + \beta_0 \ln p_t + \sum_{j=1}^{4} \beta_j \ln S_{t-j} + \beta_5 D_C + \sum_{k=1}^{q} \beta_{5+k} G_{t-1}^k + \epsilon_t \quad (2)$$

K and ε_t as before are the constant and error terms. G_{t-1}^k is the search query index for the term G^k in week $t-1$, $k = 1, \ldots, q$.

Table 1 Estimates of Baseline and Trend models for SKU A

	Baseline model	Trend model
K	7.144429‡	6.568145‡
ln $p_t(\beta_0)$	−1.134831‡	−0.8237727†
ln $S_{t-1}(\beta_1)$	0.4988018‡	0.4661723‡
ln $S_{t-2}(\beta_2)$	0.1523347†	0.1893845†
ln $S_{t-3}(\beta_3)$	−0.1565492††	−0.1440074††
ln $S_{t-4}(\beta_4)$	0.0026355	−0.0022457
$D_C(\beta_5)$	0.4803484‡	0.5744589†
$G_{t-1}^{Antipasto}(\beta_6)$		−0.8570413‡
$G_{t-1}^{Hors\ d'oeuvres}(\beta_7)$		0.6692148††
Number of observations	169	169
Log likelihood	−131.99	−128.69
Wald χ^2	89.69‡	95.05‡

Note: *p*-values: ‡: ≤0.01; † ≤0.05; †† ≤0.10.

Table 1 shows the estimates for both the baseline and the trend model for SKU A. 169 weekly observations spanning October 7, 2013 through December 31, 2015 were used to estimate both the models. Both models give consistent results and the common coefficients are similar. From the table, the search term Antipasto is negatively associated with sales while the term Hors d'oeuvres is positively associated with Sales.

3. The Value of Search Query Information

We use a one-step ahead forecast to evaluate the models.[2] To compute forecasts in week t, we use the estimated parameters in Table 1 and observed values in week $t-1$. In-sample one-step-ahead forecasts are generated using the training set (weekly values in the range 10/7/2012–12/31/2015 (r = 169 observations) in the case of SKU A). One-step-ahead out-of-sample forecasts are generated using the estimated parameters in Table 1 and the observed values from the testing set (1/1/2016–1/1/2017 ($T-r=52$ observations) in the case of SKU A). In line with prior literature on Google trends, we do not reestimate the model each period (see Choi and Varian 2009). We compute the mean absolute deviation (MAPE) as $1/N\sum_{t=1}^{N}|(\log\hat{S}_t - logS_t)|/logS_t$ for both the baseline and the trend models, where $\hat{S}_t$ is the predicted sales in week t, and N is the number of observations used.[3]

We use the metric from Cui et al. (2018), to gauge the relative improvement of using the trend model:

$$\frac{(\text{Baseline MAPE} - \text{Trend MAPE})}{\text{Baseline MAPE}} \qquad (3)$$

For SKU A, the in-sample MAPE was 9.87% for the baseline model and the trend model had an in-sample MAPE of 9.7%. The out-of-sample MAPE for the baseline model was 10.89% while the trend model yielded a MAPE of 10.49%. Using Equation (3), this gives a relative improvement of 3.67% when using the trend model.

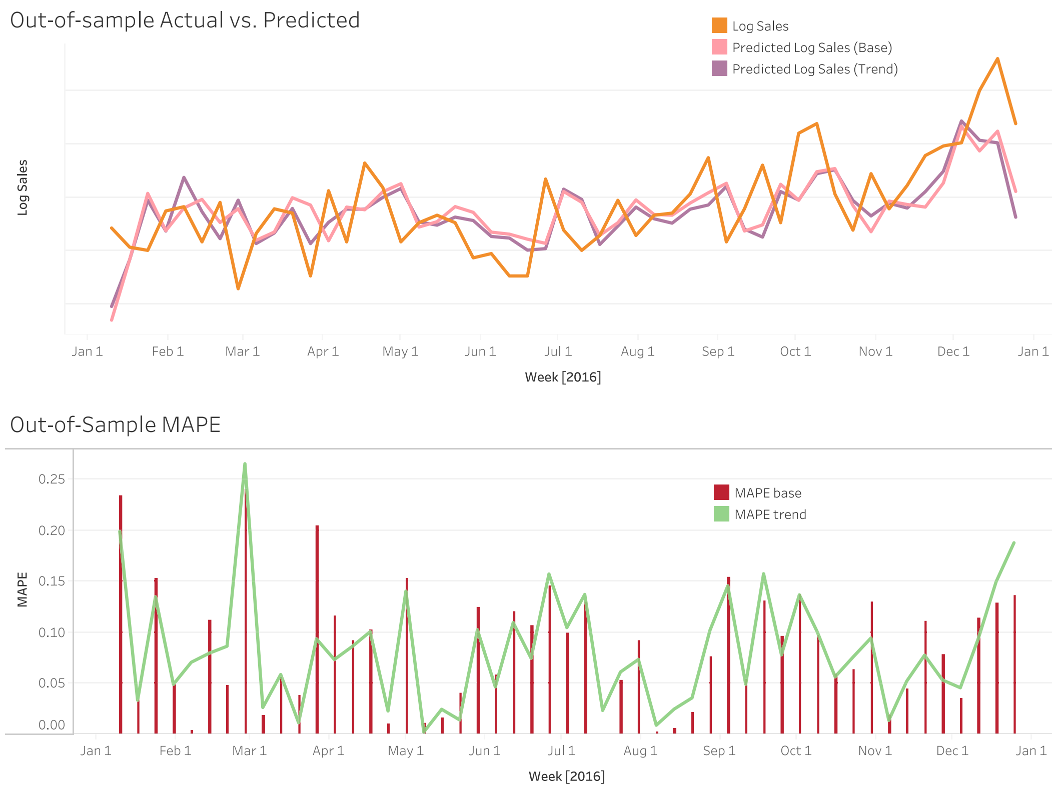
Figure 2 shows the Actual vs. the estimated values for Sales for both models for out-of-sample values. It also shows the MAPE for each week for both the baseline and the trend model. The figure uses a bar for the baseline MAPE and a line for the trend MAPE to show the contrast. The trend model has a smaller error 28 of the 52 observations (53.4% of the time). We note here that 2 weeks in the out-of-sample dataset (1st week of April and 1st week of November) contribute a big part in the reduction of MAPE for the trend model. On one hand, the MEDAPE (median absolute percentage error) of out-of-sample errors[4] for the baseline model was 9.14% while the trend model had a MEDAPE of 7.63% suggesting that the median performance of the trend model for SKU A was better than the baseline. On the other hand, such outlying differences in the absolute percentage errors could suggest that that the trend model may be picking up certain micro-trends from Google Trends resulting in a better prediction. For example, from Figure 1, there is an uptick in searches for the search term "Antipasto" and "Hors d'oeuvres" on March 20, 2016. The trend model predicted the Sales to be relatively higher in the succeeding week (as it was) than the baseline model. While we can only speculate if such short-term trends actually impact sales and accompanying forecasting models, this remains a ripe area for future research.

We used data from four other SKUs (B through E) to compute the out-of-sample percentage improvement in MAPE. Table 2 describes the SKU categories and Google search terms used. SKUs A, D, & E had weekly data available from 10/7/2012; SKU B had data from 1/1/2012; and SKU C had weekly data from 10/6/2013. In all cases, we used weeks prior to 2016 for estimation (the table identifies number of weeks used for estimation) and the 52 observations in 2016 sales to test out-of-sample performance. The results are fairly consistent, with improvements from 2.2% to 7.66%.

4. Conclusions

The primary intent of this note was to bolster the result in Cui et al. (2018) that user-generated content —in our case customer queries—can help reduce forecasting errors. Incorporating customer queries into the sales and operations process (SOP) is intuitive and relatively easy. We also echo much of what Cui et al. (2018) say about operational efficiencies that can

Figure 2 Sales of SKU A vs. Google Trends [Color figure can be viewed at wileyonlinelibrary.com]



accrue from incorporating such information—more efficient pricing and promotion; and effective management of inventory and replenishment.

While intuitive, Google Trends can be challenging. First, it is not obvious how to choose search terms. Terms that relate to the product (such as the ones used in this note) are good starting points but seemingly unrelated terms can have a lot of explanatory power. The choice of terms that signal an intent to buy is often a blend of trial and error and manager intuition.[5] Second, incorporating Google Trends data into the SOP does requires a level of "trust" in the trends data, at least for short-term forecasts—without a clear knowledge of how Google computes the query index, consistency of the index over time is an issue. Finally, such "nowcasting" models have short forecasting horizons while supply chains are typically plagued by significantly longer lead times. A further area of research is to explore how such models can aid supply chain planning in the longer-term.

In summary, whether it be Facebook interactions, retweets, or customer queries, user-generated content has a role to play in the SOP process. It gives planners another source of input that can inform future predictions.

Table 2 Out-of-Sample MAPE Improvement for Selected SKUs

SKU	Category	Search terms used	$r/T - r$	Improvement (%)
A	Appetizer	Antipasto, Hors d'oeuvres	169/52	3.65
B	Cookware	Williams Sonoma coupon	209/52	7.66
C	Wine/Cheese	Gifts for colleagues	117/52	5.33
D	Appetizer	Stonewall kitchen	169/52	2.73
E	Appetizer	Sauce for shrimp	169/52	2.21

Notes

[1]See http://www.google.com/trends.

[2]To evaluate forecast errors, we can use "dynamic" or "static" predictions. Dynamic predictions proceeds by only using known $t \leq t_r$ information on sales (S_t) for forming the basis of future predictions. So in period t_{r+L} sales data for period t_{r+L-1} are themselves forecasted. Since we are using a contemporaneous independent variable G^k_{t-1} we cannot use a true dynamic prediction since our model assumes that any future value (or out of sample value) of G_{r+L-1} is known with certainty, so at best we can have is a partially dynamic model. By contrast, Cui et al. (2018) sidestep this issue by retraining their model based on varying look ahead values L. Their method assures that they are using information in a way consistent with what the business owner would have, by using information in period t_r to predict sales in period t_{r+L}. The most direct comparison we can have to their model is by calculating future sales, using a one-week-ahead static model. This

essentially uses information in period t_r to calculate predictions in period t_{r+1} and continues to do L = 1 predictions until the end of the out of sample data.

[3]Since the time dimension of our data is one week (rather than Cui et al.'s 1 day), our one period look-ahead model has the most direct analog to Cui et al.'s L = 1 (1 period) and L = 7 (1 week) models.

[4]We calculate the MEDAPE as simply the median value of the absolute percentage errors in the validation dataset.

[5]We note that an interesting extension to our model in a Google Trends context is to use machine learning techniques to aid in the selection of most relevant search terms for forecasting.

References

Boone, T., R. Ganeshan, R. L. Hicks. 2015. Incorporating Google Trends Data into Sales Forecasting. *Foresight: Int. J. Appl. Forecast*, Summer (38): 9–14.

Castle, J. L., N. W. P. Nicholas, D. F. Hendry. 2009. Nowcasting is not just contemporaneous forecasting. *Natl Inst. Econ. Rev.* **210** (1): 71–89.

Choi, H., H. Varian. 2009. Predicting the present with Google Trends. Technical Report, Google Inc. Available at http://google.com/googleblogs/pdfs/google_predicting_the_present.pdf (accessed date May 1, 2017).

Choi, H., H. Varian. 2012. Predicting the present with Google Trends. *Econ. Rec.* **88**(s1):2–9.

Cui, R., S. Gallino, A. Moreno, D. Zhang. 2018. The operational value of social media information. *Prod. Oper. Manage.* **27**(10): 1749–1769.

Dimpfl, T., S. Jank. 2016. Can internet search queries help to predict stock market volatility? *Eur. Financ. Manag.* **22**(2): 171–192.

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

Limnios, C., H. You. 2016. Can Google Trends actually improve housing market forecasts? SSRN Working paper. Available at https://ssrn.com/abstract=2886705 or https://doi.org/10.2139/ssrn.2886705 (accessed date May 1, 2017).