Applied Economics Letters

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rael20

Can google trends improve sales forecasts on a product level?

Benjamin Fritzsch, Kai Wenger, Philipp Sibbertsen & Georg Ullmann

To cite this article: Benjamin Fritzsch, Kai Wenger, Philipp Sibbertsen & Georg Ullmann (2020) Can google trends improve sales forecasts on a product level?, Applied Economics Letters, 27:17, 1409-1414, DOI: 10.1080/13504851.2019.1686110

To link to this article: https://doi.org/10.1080/13504851.2019.1686110

Published online: 01 Nov 2019.

APPLIED ECONOMICS LETTERS
2020, VOL. 27, NO. 17, 1409–1414
https://doi.org/10.1080/13504851.2019.1686110

Routledge
Taylor & Francis Group

ARTICLE

Can google trends improve sales forecasts on a product level?

Benjamin Fritzsch [a], Kai Wenger [b], Philipp Sibbertsen [b] and Georg Ullmann [a]

[a]Institut für Integrierte Produktion Hannover gGmbH, Hannover, Germany; [b]School of Economics and Management, Leibniz University Hannover, Hannover, Germany

ABSTRACT
Combining standard time series models with search query data can be helpful in predicting sales. We include the search volume of company as well as product-related keywords provided by Google Trends as new predictors in models to forecast sales on a product level. Using weekly data from January 2015 to December 2016 of two products of the audio company Sennheiser we find evidence that using Google Trends data can enhance the prediction performance of conventional models.

KEYWORDS
Google econometrics; forecasting; search query data

JEL CLASSIFICATION
C12; C22; M11

I. Introduction

Econometric modelling to forecast sales on a product level is a widely studied field. On the one hand, models that lack in predictability induce supply shortfalls if the prediction falls below the actual demand. On the other hand, false predictions induce storage costs if the forecast exceeds the actual demand. Therefore, an appropriate econometric model to forecast sales is needed.

Traditional time series models that aim to predict sales usually rely on historical data. Nowadays, online search engines provide new sources of data about real-time economic activity (Ettredge, Gerdes, and Karuga 2005). Here, we examine whether considering Google Trends data improves forecasting sales on a product level. Using Google Trend data was firstly introduced in econometrics by Choi and Varian (2009). Later, they were considered in predicting unemployment, exchange rates, private consumption, return volatility, and so on (see Choi and Varian 2012; Vicente, López-Menéndez, and Pérez 2015; Smith 2016; Bulut 2017; D'Amuri and Marcucci 2017; Perlin et al. 2017; González-Fernández and González-Velasco 2018; among others). Due to a lack of free availability of company's sales data, there is just a little literature about forecasting sales on a product level (Cui et al. (2018), Boone et al. (2018)). The research up to now focuses on data provided by online retailers, which directly sell their products to the customer. Since these companies fully execute their business online, the link between using a search engine and buying the products is close.

In our contribution, we examine a unique set of sales data of two earphones provided by the audio company *Sennheiser*. The principal business of Sennheiser is to sell their products on the spot through department stores. That is they provide their products indirectly and the link between searching for a product and buying it is not as close as for an online retailer. Therefore, it is not clear whether the results obtained in the literature carry over to the products we investigate.

To shed light on the question whether search query data improve forecasts for our data, we apply traditional time series models that additionally include Google Trend data as an explanatory variable. To do so, we first collect time series of Google Trends keywords based on company and product information. Then, we investigate in-sample through Granger causality tests which keywords are related to the sales data of each product. We then apply a model specification procedure to build appropriate time series models which include Google Trends keywords. In an out-of-sample experiment, we split the sample into an estimation and forecasting period and find that the RMSE of the models that include Google Trend data decreases by 4–14%.

CONTACT Kai Wenger wenger@statistik.uni-hannover.de School of Economics and Management, Institute of Statistics, Leibniz University Hannover, Königsworther Platz 1, D-30167, Hannover, Germany

The rest of the paper is structured as follows. Section 2 introduces the data collection and cleaning. Section 3 describes the model and the model specification procedure. Section 4 presents our results and Section 5 concludes.

II. Data

We use Google Trends data for search activity because Google is by far the most popular online search engine. Since we want to forecast sales on a product level we choose keywords which combine company as well as product information with terms that reflect purchase intention (e.g. *review, store*, and so on). We start our analysis with 40 keywords and filter them in a procedure similar to Afkhami, Cormack, and Ghoddusi (2017) which is described in the next section until we are left with five keywords for product I and four keywords for product II (see Table A1 in the appendix). For each selected keyword Google does not provide the exact number of queries, but an index which is normalized from 0 to 100 (which represents the search intensity for that keyword in the searching period). For the time period, we have chosen only weekly search query data is available.

Google calculates each Google Trends index on a randomly drawn fraction, which changes daily, of all searches. Consequently, the query requests on identical keywords lead to slightly different series on different days. Therefore, we consider the request submitted on a single day as a perturbed series by some measurement error. However, query requests for the same search item on different days should have the same underlying signal. To approximate this signal for each search term series, we compute the indices by averaging over keyword requests submitted over seven sub-sequential days in August 2018 (cf. D'Amuri and Marcucci 2017; Bulut 2017, among others).

The sales data are provided by the audio company Sennheiser for two of their products (product I and product II). The raw data are discontinuous in time since they contain each single order for each product. Furthermore, the data includes missing as well as negative values which are related to reshipments. Therefore, in a first step, the data are cleaned by these values and afterwards summarized on a weekly basis. All in all, we obtain 101 sales data for each product from 01/04/2015 to 12/04/2016.

In a second step, a linear trend is removed from both time series. To check whether the series are I (0), i.e. stationary, ADF and KPSS tests are performed on the detrended time series, which show at a 1% significance level that the data is stationary.

III. Methodology

An autoregressive moving average (ARMA) model for the detrended I(0) sales data is specified because past sales should contain information about future sales. The model which possibly includes Google Trends keywords is given as:

$$y_t = \nu + \sum_{j=1}^{p} \phi_j y_{t-j} + \sum_{j=1}^{q} \zeta_j u_{t-j} + \sum_{k=1}^{z} \sum_{i=1}^{s} \theta_i^k GT_{t-i}^k + u_t, \quad t \in \mathbb{N}_0, \quad p, q, s \in \mathbb{N}, \tag{1}$$

where y_t are the sales series at $t = 1, \dots, T$, ν is a constant, ϕ_j and ζ_j are the AR and MA coefficients, and θ_i the coefficients of the lagged Google Trends keyword GT^k for $k = 1, \dots, z$. The orders p, q, and s show how many lags of the specific variable are included in the model and z gives the number of keywords added. The case $z, s = 0$ is referred to as the benchmark ARMA model. We assume that the error term u_t is White noise with $E(u_t) = 0$, $E(u_t u_t{}') = \sigma_u$ and $E(u_t u_r{}') = 0$ for $r \neq t$.

Before model (1) can be estimated the orders p, q, s and z have to be determined. It must be specified which Google Trends keywords are included in the regression and at which lag as well as how many lags are included for the past sales. We choose the order p and q according to the AIC since we are interested in forecasting. The selection of the keywords and at which lag length they are included in (1) is done by the following procedure:

(1) In a first step, we check whether and at which lag the Google Trend keywords Granger cause the sales of each product. The Granger causality test can be considered as a joint test on the coefficients of each keyword at all lags in model (1). The testing problem is

$$H_0: \theta_1^k = \ldots = \theta_s^k = 0$$

against the alternative

$$H_1: \theta_i^k \neq \theta_j^k \quad \text{for} \quad 1 \leq j \neq i \leq s.$$

for each keyword k. In the next step, we just proceed for product I with search terms that Granger cause sales of product I and for product II with those that Granger cause sales of product II.

(2) In a second step, we build combinations of two Granger causal Google Trends keywords for each of the products and include them in model (1). Then we check whether these models improve the model accuracy compared to the more parsimonious models with fewer (or none) search terms included. Since adding explanatory variables always increases the model accuracy, we just consider models including combinations of two keywords that decrease the AIC compared to the models with just one (or none) keyword.
Additionally, we just consider keywords at a specific lag i if their coefficient θ_i^k in (1) is statistically significant at the $\alpha = 10\%$ significance level.

(3) In a last step, we form for each product models that combine three and more Granger causal Google Trends keywords. As in the second step, we just consider models where on the one hand the AIC decreases compared to the more parsimonious models, and on the other hand at least one coefficient of the added keyword is significant.

We evaluate the model accuracy in-sample by using the AIC, the R^2, the adjusted R^2 as well as the RMSE. Additionally, in our out-of-sample forecasting experiment, we conduct Mincer-Zarnowitz regressions of the form

$$y_t = b_1 + b_2\hat{y}_t + e_t, \quad t = 1, \ldots, T$$

with $e_t \overset{iid}{\sim} (0, \sigma_e^2)$ to test whether the forecast has predictive power ($b_2 > 0$) and whether the forecast is unbiased ($b_1 = 0$ and $b_2 = 1$). Furthermore, we apply Diebold-Mariano tests (Diebold and Mariano (1995)) to verify if the models that include Google Trends keywords have a significant higher predictive accuracy than the benchmark ARMA model without search query information.

IV. Results

At the beginning of our analysis, we compare the models that include Google trends keywords with the benchmark ARMA model. We show the in-sample estimation results as well as the measures of accuracy for product I in Table 1 and for product II in Table 2.

According to the AIC as the benchmark models, an ARMA(2,0) is fitted for product I and an ARMA (1,0) for product II. The AR coefficients for both series are all positive and statistically significant at a 5% level. They range roughly between 0.2 and 0.45. Therefore, sales of the previous week (and for product I the week before the previous) have a positive effect on sales of this week.

For product I, we consider Google trends keywords up to two lags while for product II we consider keywords up to three lags based on the AIC. After we applied our keyword selection procedure described in Section 3 we end up with four models that include two (models a to c) or three (model d) Google trends keywords. The selected search terms are given in Table A1.

We observe that all keywords are statistically significant at some lag and have the same sign and roughly the same magnitude for the same keyword at the same lag across the models. However, different keywords have different signs at different lags.

For both sales series, it is obvious that the inclusion of Google Trends keywords increases the predictive accuracy compared to the benchmark ARMA model: The (adjusted) R^2 increases and the AIC as well as the RMSE decrease. Relative to the benchmark model, the RMSE improves by 6% to 12% for product I and by 3% to 17% for product II. For both sales series, the model with three keywords offers the best accuracy in-sample.

We also perform an out-of-sample experiment where we split the series into an estimation and a forecasting period. First, the parameters of the models are estimated from January 2015 to February 2016. Second, the one-step sales forecasts for every product up to December 2016 are determined for each model. Then, the forecast errors are

Table 1. In-sample results of product I. In brackets below the estimated coefficients are the p-values, where *** denotes significance at 1%, ** significance at 5% and * significance at 10%. The variable ΔRMSE is the relative improvement of the RMSE of the models that include Google Trends keywords with the benchmark ARMA model (left column).

	ARMA	(a)	(b)	(c)	(d)
Intercept	−30.153	50.971	−43.373	54.290	−24.236
	(0.265)	(0.154)	(0.142)	(0.167)	(0.362)
AR1	0.261	0.333	0.370	0.458	0.328
	(0.014***)	(0.001***)	(0.000***)	(0.000***)	(0.001***)
AR2	0.217	0.204	0.307	0.255	0.220
	(0.023***)	(0.035***)	(0.003***)	(0.011***)	(0.021***)
GT1 lag1			−7.287		
			(0.011***)		
GT2 lag1		5.696			5.217
		(0.000***)			(0.001***)
GT3 lag1				−7.309	−7.865
				(0.018***)	(0.003***)
GT3 lag2				7.336	
				(0.019***)	
GT4 lag2		−5.501		−5.458	
		(0.012***)		(0.031***)	
GT5 lag2			4.933		3.637
			(0.007***)		(0.036***)
R^2	0.338	0.457	0.416	0.436	0.491
Adj R^2	0.324	0.434	0.391	0.405	0.463
AIC	1406.522	1390.978	1398.099	1396.731	1386.612
RMSE	282.676	256.107	265.486	261.007	248.004
ΔRMSE	0.000	0.094	0.061	0.077	0.123

Table 2. In-sample results of product II. In brackets below the estimated coefficients are the p-values, where *** denotes significance at 1%, ** significance at 5% and * significance at 10%. The variable ΔRMSE is the relative improvement of the RMSE of the models that include Google Trends keywords with the benchmark ARMA model (left column).

	ARMA	(a)	(b)	(c)	(d)
intercept	38.539	−23.425	7.836	50.028	−4.222
	(0.597)	(0.755)	(0.910)	(0.478)	(0.952)
AR1	0.325	0.336	0.309	0.353	0.419
	(0.001***)	(0.001***)	(0.001***)	(0.000***)	(0.000***)
GT6 lag3				15.348	
				(0.003***)	
GT7 lag1					−14.895
					(0.004***)
GT7 lag3		9.113			19.480
		(0.074**)			(0.000***)
GT8 lag1			18.222		21.732
			(0.002***)		(0.000***)
GT8 lag3					−11.146
					(0.049***)
GT9 lag1		−13.628	−15.315	−13.612	−13.078
		(0.039***)	(0.018***)	(0.034***)	(0.032***)
GT9 lag2					14.853
					(0.022***)
GT9 lag3					−14.688
					(0.019***)
R^2	0.103	0.168	0.222	0.215	0.391
Adj R^2	0.094	0.141	0.197	0.190	0.337
AIC	1570.527	1567.242	1560.606	1561.531	1546.558
RMSE	708.6359	682.781	660.052	663.171	583.839
ΔRMSE	0.000	0.036	0.069	0.064	0.176

calculated and measures of accuracy are determined. We further conduct Mincer-Zarnowitz regressions as well as Diebold-Mariano tests.

The out-of-sample results for products I and II are given in Table 3. In case of product I, we observe that models (a) to (d) deliver unbiased sales forecasts. F-tests of H_0: $b_0 = 0$ and $b_1 = 1$ cannot be rejected at a 5% significance level for all models.

For product II we observe mixed results. On the one hand, the intercept is not statistically significant, too. On the other hand, the slope

Table 3. Out-of-sample results for products I and II. The upper part of each panel shows the coefficients of the Mincer-Zarnowitz regressions with standard deviations in brackets below. Here, *** denotes significance at 1%, ** significance at 5% and * significance at 10%. The lower part of each panel shows the measures of accuracy where ΔRMSE is the relative improvement of the RMSE compared to the benchmark ARMA model. $teststat_{DM}$ is the test statistic of the Diebold-Mariano test.

	ARMA	(a)	(b)	(c)	(d)
product I					
b_0		−10.058	−8.885	−34.75	−14.407
		(54.115)	(55.532)	(55.434)	(52.439)
b_1		1.108	1.049	1.030	1.079
		(0.200)***	(0.202)***	(0.193)***	(0.182)***
R^2	0.314	0.435	0.408	0.417	0.471
RMSE	372.754	338.404	346.281	343.676	327.377
ΔRMSE		0.092	0.071	0.078	0.122
$teststat_{DM}$		−2.375**	−1.978*	−2.112**	−2.224**
product II					
b_0		−110.406	−133.835	−134.520	−75.448
		(100.635)	(97.204)	(94.604)	(93.439)
b_1		0.592	0.499	0.608	0.741
		(0.389)	(0.321)	(0.292**)	(0.249***)
R^2	0.114	0.231	0.195	0.237	0.343
RMSE	661.066	615.80	630.183	613.48	569.301
ΔRMSE		0.068	0.047	0.072	0.139
$teststat_{DM}$		−2.334**	−1.043	−1.387	−2.213**

coefficient is just for models (c) and (d) significant. Moreover, the F-test is rejected for all models. Therefore, we conclude that the forecasts have predictive power, but do not deliver unbiased sales predictions.

Considering the measures of accuracy we observe that for product I the out-of-sample R^2 of models (a) to (d) that include Google Trends keywords increase compared to the benchmark model and that the RMSE decreases by 7% to 12%. The Diebold-Mariano test rejects for all models that the benchmark model has higher or equal predictive power than models (a) to (d). We therefore conclude that models which utilize search query data better forecast sales for product I.

The results regarding product II are equivocal. The out-of-sample R^2 increases and the RMSE decreases by 4% to 14% compared to the benchmark model. However, the Diebold-Mariano test only rejects for (a) and (d) that the benchmark model has equal or higher forecasting power.

Overall, we find that including Google Trends keywords in the regression increases the model accuracy in-sample. We further observe that we significantly increase the predictive power of the benchmark model out-of-sample. The sales series of product I seems to be better predictable than the sales series of product II.

V. Conclusion

In this paper, we investigate whether Google Trends data can enhance forecast performance for sales on a product level. Compared to other studies we examine products which are not mainly sold online, but on the spot through department stores. The results indicate that including Google Trends keywords can be helpful in getting more refine predictions of future sales for a company.

Acknowledgments

This paper has been composed as part of the research project “Development of a forecasting model to determine short- and medium-term sales using search engine data” (reference number Ul419/7 - 1), which is funded by the German Research Foundation (DFG). We are also grateful to Sennheiser electronic GmbH & Co. KG for sharing their sales data with us. We would like to thank the anonymous referee for his review. We highly appreciate his comments and suggestions.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Deutsche Forschungsgemeinschaft [Ul419/7 - 1].

ORCID

Benjamin Fritzsch http://orcid.org/0000-0001-7897-5340
Kai Wenger http://orcid.org/0000-0002-6566-0427

References

Afkhami, M., L. Cormack, and H. Ghoddusi. 2017. "Google Search Keywords that Best Predict Energy Price Volatility." *Energy Economics* 67: 17–27. doi:10.1016/j.eneco.2017.07.014.

Boone, T., R. Ganeshan, R. L. Hicks, N. R. Sanders. 2018. "Can Google Trends Improve Your Sales Forecast?". *Production and Operations Management* 27: 1770–1774. doi:10.1111/poms.2018.27.issue-10.

Bulut, L. 2017. "Google Trends and Forecasting Performance of Exchange Rate Models." *Journal of Forecasting* 37 (3): 303-315. doi:10.1002/for.2500.

Choi, H., and H. Varian. 2009. *Predicting the Present with Google Trends*, 1–23. Technology report Google, Google Inc.

Choi, H., and H. Varian. 2012. "Predicting the Present with Google Trends." *Economic Record* 88 (s1): 2–9. doi:10.1111/ecor.2012.88.issue-s1.

Cui, R., S. Gallino, A. Moreno, D. J. Zhang. 2018. "The Operational Value of Social Media Information." *Production and Operations Management* 27 (10): 1749–1769. doi:10.1111/poms.2018.27.issue-10.

D'Amuri, F., and J. Marcucci. 2017. "The Predictive Power of Google Searches in Forecasting US Unemployment." *International Journal of Forecasting* 33 (4): 801–816. doi:10.1016/j.ijforecast.2017.03.004.

Diebold, F. X., and R. S. Mariano. 1995. "Comparing Predictive Accuracy." *Journal of Business & Economic Statistics* 13 (3): 253–263. doi:10.1080/07350015.1995.10524599.

Ettredge, M., J. Gerdes, and G. Karuga. 2005. "Using Web-based Search Data to Predict Macroeconomic Statistics." *Communications of the ACM* 48 (11): 87–92. doi:10.1145/1096000.

González-Fernández, M., and C. González-Velasco. 2018. "Can Google Econometrics Predict Unemployment? Evidence from Spain." *Economics Letters* 170: 42–45. doi:10.1016/j.econlet.2018.05.031.

Perlin, M. S., J. F. Caldeira, A. A. P. Santos, and M. Pontuschka. 2017. "Can We Predict the Financial Markets Based on Google's Search Queries?" *Journal of Forecasting* 36 (4): 454–467. doi:10.1002/for.2446.

Smith, P. 2016. "Google's MIDAS Touch: Predicting UK Unemployment with Internet Search Data." *Journal of Forecasting* 35 (3): 263–284. doi:10.1002/for.v35.3.

Vicente, M. R., A. J. López-Menéndez, and R. Pérez. 2015. "Forecasting Unemployment with Internet Search Data: Does It Help to Improve Predictions When Job Destruction Is Skyrocketing?" *Technological Forecasting and Social Change* 92: 132–139. doi:10.1016/j.techfore.2014.12.005.

Appendix

Table A1. Keywords selected by the procedure in Section 3 for products I and II. The actual product names are subject to a non-disclosure agreement and can therefore not be shared.

Product I	keyword	product II	keyword
GT1	headphones	GT6	sennheiser store
GT2	<product name> compare	GT7	<product name> headphones
GT3	sennheiser sound	GT8	<product name> model
GT4	<product name> in-ear	GT9	sennheiser compare
GT5	<product name> store		