Forecasting Sales of Durable Goods – Does Search Data Help?

Carsten D. Schultz

Abstract Search data can be used to forecast macroeconomic measures. The present study extends this research direction by drawing on real sales data from a household panel over two years. Specifically, the study analyzes whether search data improves forecasts for seven products groups of durable goods. The forecast model also includes the average weekly price and a dummy for the Christmas season. Forecast accuracy is indeed improved when search data is included even for product groups that have a short information and search phase. The product groups, however, need to be chosen carefully, because some durable goods show no lag between online search and purchase.

Carsten D. Schultz
University of Hagen, Universitätsstr. 11, 58097 Hagen,
✉ carsten.schultz@fernuni-hagen.de

ARCHIVES OF DATA SCIENCE, SERIES A
(ONLINE FIRST)
KIT SCIENTIFIC PUBLISHING
Vol. 6, No. 2, 2020
DOI: 10.5445/KSP/1000098012/01
ISSN 2363-9881

1 Introduction

One part of the consumer decision process is the information and search phase (Schultz, 2007). Consumers search to retrieve information, e.g. on product features and retail conditions. Consumers consequently base most of their decisions on the information obtained and thus predetermine their choices during this phase (Schultz, 2007). Online searching has become a primary source of information for people, drastically reducing individual search costs. Consumers draw, e.g., on product reviews (Fan et al., 2017), social media (Asur and Huberman, 2010; Hennig-Thurau et al., 2015), and search engines (Kulkarni et al., 2012). In particular, search engines, such as Baidu, Bing, and Google, provide fast access to the information sought for.

Beyond informing consumers, the underlying consumer behavior has become itself a source of information for companies and researchers. One stream of research addresses the potential of such information for the purpose of forecasting. For example, Fan et al. (2017) extract a sentiment index from product reviews to augment product forecasts. Social media has similarly been used to forecast the success of cinema movies (Asur and Huberman, 2010; Hennig-Thurau et al., 2015) and to predict stock performance (Tirunillai and Tellis, 2012; Sul et al., 2017). Another source of information in this research stream is search data.

Search data relates to the search queries users enter in search engines during their search process. Search data has previously been used to forecast automobile sales (Fantazzini and Toktamysova, 2015), cinema admissions (Hand and Judge, 2012; Kulkarni et al., 2012), economic indicators (Choi and Varian, 2012; Vosen and Schmidt, 2011), housing prices (Oestmann and Bennöhr, 2015; Dietzel, 2016), influenza epidemics (Ginsberg et al., 2009), tourism demand (Önder and Gunter, 2016; Park et al., 2017), and unemployment rates (D'Amuri and Marcucci, 2017). For the prediction of risks and returns in stock markets see Calomiris and Mamaysky (2019). These studies indicate that search data may be useful for forecasting macroeconomic measures. For product sales, these forecasts serve as the basis for planning activities concerning multiple company functions, such as production, procurement, communication, and distribution.

The present study explores the use of such search data to forecast sales of durable goods. More precisely, the study draws on panel data for seven product groups: audio books, blu-ray-players, calendars, children's books, hifi systems, smartphones, and televisions from a two-year period. Compared to automobiles and housing, the consumer search process is considerable shorter in these product categories. This study thus contributes to the applicability of search data in case of short search decision phases.

The retail perspective further motivates this approach. Retailers, especially stationary retailers, have limited shelf and storage space. Increasing forecast precision reduces overstock due to overestimating product sales and reduces revenue loss due to product shortage when underestimating product sales. In particular, retailers may increase storage and shelf efficiency. The research question of this study, thus, is: Does search data help forecasting sales in case of product groups with a rather short information and search phase?

The remainder of the paper is organized as follows. Section 2 relates this study to the existing literature on using search data for forecasting purposes. Section 3 first outlines the data sample. This section also addresses the modeling approach. Then, the data is analyzed and the empirical results are presented. The paper concludes with a discussion and an outline for future research directions.

2 Related Literature

Online data have been used to predict product sales, such as product reviews (Fan et al., 2017) and social media (Asur and Huberman, 2010; Hennig-Thurau et al., 2015). The present study relates particularly to research on integrating search data in forecasting tasks. This section, thus, focuses on studies closely related to the research question.

Two influential works are Ginsberg et al. (2009) and Choi and Varian (2012). Ginsberg et al. (2009) demonstrate that search data helps forecasting influenza levels in nine regions of the United States. The reporting lag is about one day and, thus, one to two weeks ahead of the competing surveillance report. The authors fit a linear model using the log-odds of the percentage of physician visits V and of the related fraction of search queries Q where α is a coefficient and ϵ is the error term:

$$\mathrm{logit}(V(t) = \alpha \cdot \mathrm{logit}(Q(t)) + \epsilon \ . \tag{1}$$

Based on earlier technical reports, Choi and Varian (2012) show the broad applicability of search data to social and economic indicators. These indicators refer to trends in automotive, unemployment, travel, and consumer confidence indices. For example, search data categories (in particular related to SUVs and automotive insurance) can improve the forecast on monthly sales for vehicles and parts by the Census Bureau of the United States:

$$\begin{aligned} \mathrm{sales}_t = \alpha + \beta_1\, \mathrm{sales}_{t-1} &+ \beta_2\, \mathrm{sales}_{t-12} \\ &+ \beta_3\, \mathrm{search(SUV)} + \beta_4\, \mathrm{search(Insurance)} \\ &+ \epsilon\ . \end{aligned} \tag{2}$$

In these cases, a seasonal autoregressive model including search data performs 5 to 20 percent better than its benchmark without search data. Subsequent research has addressed these areas in more detail.

Using monthly automobile sales, search data is also applicable with forecast horizons of up to two years (Fantazzini and Toktamysova, 2015). The authors also include notable economic indicators and compare a variety of competing models. In these long-term forecasts, search data may explain some part of the nonlinearity of sales data. Similarly, models including search data outperform alternative indicators, improving forecast horizon for unemployment rate (D'Amuri and Marcucci, 2017). For the unemployment rate U, autoregressive models with other explanatory variables X are adopted, including lag polynomials such that $\beta_1(L)U_t = \sum_{j=1}^{p} \beta_{1,j} U_{t-j}$ and $\beta_2(L)X_t = \sum_{j=1}^{q} \beta_{1,j} X_{t-j}$:

$$U_{t+h} = \alpha + \beta_1(L)U_t + \beta_2(L)X_t + \epsilon\ . \tag{3}$$

Search data also augments travel predictions (number of tourists) in comparison to standard time series, for example, to Austria (Önder and Gunter, 2016) and South Korea (Park et al., 2017). Here, Önder and Gunter (2016) find that textual search data performs better than data from image search, but that both may complement each other. The authors base their forecasts for tourists T on autoregressive distributed lag models including previous tourist numbers, web search W, image search I, monthly dummy, and a deterministic linear *trend*:

$$\begin{aligned}\ln(T)_t = \alpha + &\sum_{g=1}^{12} \beta_{1,g} \ln(T)_{t-g} + \sum_{h=1}^{12} \beta_{2,h} \ln(W)_{t-h} \\ &+ \sum_{i=1}^{12} \beta_{3,i} \ln(I)_{t-i} + \sum_{j=1}^{11} \beta_{4,j} \text{month}_j + \beta_5 \text{ trend} \\ &+ \epsilon \ .\end{aligned} \tag{4}$$

In times of economic stability and high internet usage, search volume increases the explanatory power to predict house prices (Oestmann and Bennöhr, 2015; Dietzel, 2016). The authors use quarterly data and thus provide an approach of aggregation for the weekly search index. Oestmann and Bennöhr (2015) use a lagged fixed-effect regression to model house prices *P*. In particular, the authors include independent controls such as inflation, interest rates, and unemployment (denoted as vector *X*), *seasonal* dummies, a *crisis* dummy, and *search* data categories:

$$\begin{aligned}\ln(P)_{i,t} = \alpha + \sum_{g=1}^{3} \beta_{1,g} (X)_{i,g,t-1} + \sum_{h=1}^{3} \beta_{2,h} \text{ season}_{h,t-1} \\ + \beta_3 \text{ crisis} + \beta_4 \text{ search}_{i,t-1} + \epsilon_{i,t} \mu_i \ .\end{aligned} \tag{5}$$

Similarly, Hand and Judge (2012) use fixed *monthly* seasonal dummies, a time *trend*, search data, and previous admissions to forecast cinema *admissions*. Here, search data refer to the predicted cinema admissions of the same week, thus nowcasting without any lag:

$$\begin{aligned}\ln(\text{admission}_t) = \alpha + \beta_1 \ln(\text{admission}_{t-1}) + \beta_2 \ln(\text{admission}_{t-12}) \\ + \beta_3 \text{ search}_t + \sum_{j=1}^{11} \beta_{4,j} \text{ season}_j + \beta_5 \text{ trend} + \epsilon_t\end{aligned} \tag{6}$$

In consequence, with considerably shorter search phases, search data increases the accuracy of forecasts on general cinema admissions (Hand and Judge, 2012) and box office revenues (Kulkarni et al., 2012). Similarly, Boone et al. (2015) find no lag for food products in their exemplary introduction. Table 1 presents an overview of related studies that integrate search data in sales forecasting tasks.

Table 1: Overview of Related Literature (W=weeks,M=months,Q=quarters).

Study	Time Period t	Object of Analysis	Method	Key Result
Ginsberg et al. (2009)	128 W	influenza epidemics	log linear regression	Search data can accurately estimate the level of weekly influenza activity in 9 regions of a country, with a reporting lag of one day.
Choi and Varian (2012)	127 M 390 W	economic indicators	seasonal AR	Economic indicators are automotive, unemployment, travel, and consumer confidence index. Simple seasonal AR models with search data show a 5 to 20 percent increase in performance.
Hand and Judge (2012)	60 M	cinema admissions	seasonal AR trend	Search data increases the accuracy of short-term cinema admission.
Hu et al. (2014)	103 M	automobile sales	Bayesian dynamic linear model	21 vehicles in 4 categories. Search data improves goodness-of-fit for sales, both in- and out-of-sample.
Fantazzini and Toktamysova (2015)	162 M	automobile sales	34 competing models	22 car brands. Bayesian VAR models performed well for all car brands and for short- and medium-term forecasts, while parsimonious bivariate models including sales and search data only performed well for long-term forecasts. Search data may explain a part of the non-linearity in sales data.
Oestmann and Bennöhr (2015)	33 Q	housing prices	fixed effect regression	14 European countries. Search data increases the explanatory power of the benchmark model, especially for high Internet use and economic stability.
Önder and Gunter (2016)	60 M	tourism demand	autoregressive distributed lag models	Using search data for tourism demand forecasting is valuable both for seasonal and seasonal adjusted data. Image search performs worse than textual search but may complement it.
D'Amuri and Marcucci (2017)	66 M	unemployement rates	529 models	Models augmented with search data outperform traditional ones in predicting monthly unemployment rates.
Park et al. (2017)	142 M	tourism demand	seasonal ARIMAX	Search data-augmented models performed better than standard time series models.
The present study	105 W	durable goods	ARX	7 product groups. Improvement of forecast for sales reported in panel data.

3 Methodology

3.1 Data Sample

The present study draws on two data sources. The GfK household panel provides sales data for seven groups of durable goods: Audio books, blu-ray players, calendars, children's books, hifi systems, smartphones, and TVs. All of these product groups have rather short search behavior processes, but show differences in other areas. For example, purchases of calendars predominantly happen at the end of a year and purchase decisions for TVs take more time compared to all other product groups. The panel consists of 15,515 households over a two-year period (01.01.2014 – 31.12.2015) and includes 43,720 data entries. A data entry is one transaction for one household and day. Consumers bought 1,955 audio books, 1,595 blu-ray players, 12,732 calendars, 16,007 children's books, 2,265 hifi systems, 6,002 smartphones, and 3,164 TVs.

Search volume is registered by Google Trends. The data service provides a time series index [0...100] of query volume that users enter into the search engine (Choi and Varian, 2012). Each time series is normalized by dividing the count for each query in a week by the total amount of online search queries (Ginsberg et al., 2009). The higher the index, the higher the search volume for the query selected. These query fractions are provided on a weekly basis stemming from Germany. The search volume is retrieved for the seven product categories introduced above (search queries (in German): *Hörbücher*, *Blu-ray Player*, *Kalender*, *Kinderbücher*, *Hifi Anlage*, *Smartphone*, and *Fernseher*). A single search term represents these seven product groups well linguistically. To consider a corresponding lag between search volume and sales, the data period runs from 01.12.2013 to 31.12.2015.

Table 2 provides an overview of the sales and search data. Following the search volume, data is aggregated on a weekly base. For example, 967 households bought a total of 1,955 audio books with an average weekly price between 5.08 and 25.20 Euros over the 105 weeks. The search index ranged from 56 to 100 in those two years. Figure 1 displays the sales and search series for the seven product groups.

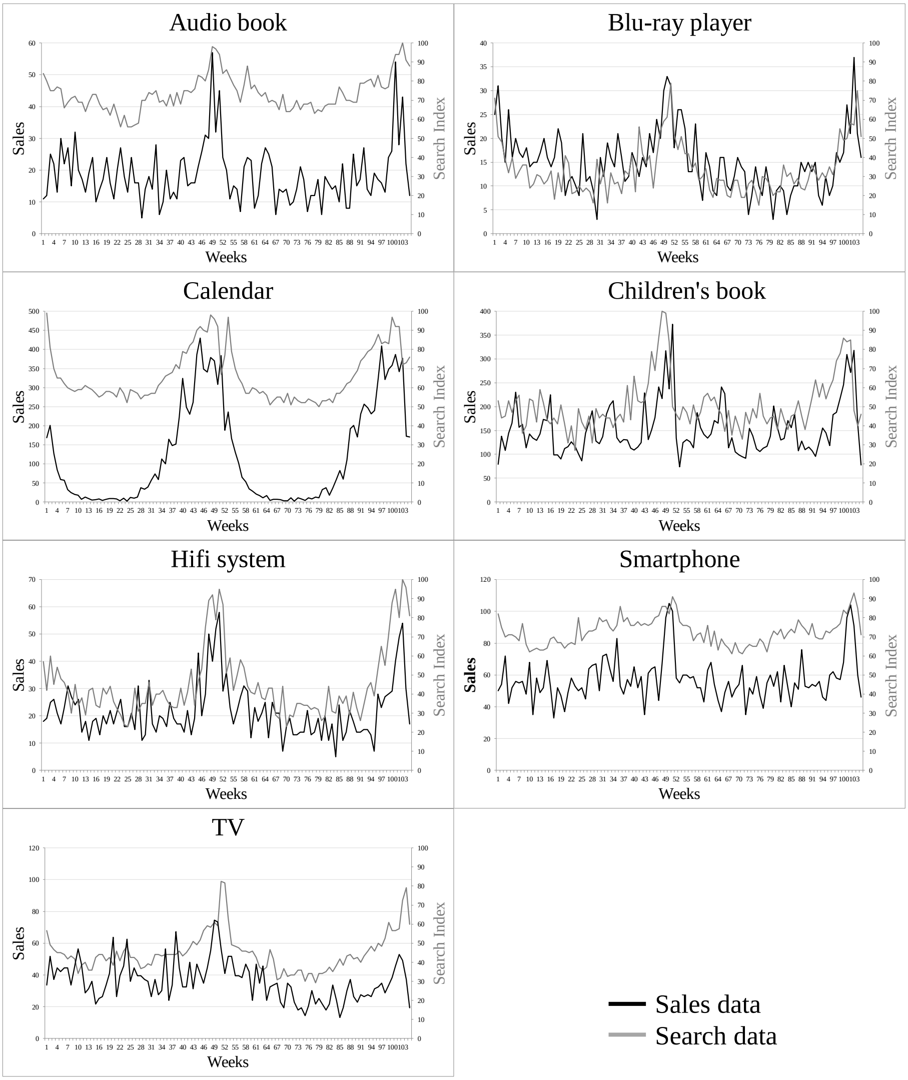

Figure 1: Sales and Search Series for the Product Groups.

Table 2: Weekly Data Overview.

Product Group	Households	Sales				Price				Search Volume			
		min	max	avg	sd	min	max	avg	sd	min	max	avg	sd
Audio book	967	5	57	18.6	9.06	5.08	25.20	13.68	3.89	56	100	73.5	9.16
Blu-ray player	1,372	3	37	15.2	6.47	28.56	170.55	95.17	31.49	15	79	33.7	12.85
Calendar	6,568	2	430	121.3	130.89	2.05	14.98	8.03	2.51	50	99	66.4	13.61
Children's book	6,159	74	373	152.5	55.07	5.92	13.48	8.93	1.53	27	100	50.8	13.83
Hifi system	1,865	5	58	21.6	9.91	42.00	339.24	142.67	51.34	23	100	45.1	18.12
Smartphone	4,522	33	105	57.2	13.88	111.62	263.90	172.66	28.82	61	93	72.8	7.02
TV	2,838	11	62	30.1	10.35	362.68	949.54	517.34	96.76	35	99	54.0	12.13

Notes: N = 105 weeks per product group.

3.2 Model

For the comparison of sales forecasting with and without search data, the present study draws on simple autoregressive models (e.g., Choi and Varian, 2012). The baseline model A refers to sales in a given week as a function of sales in the previous week. For model B, we additionally include a dummy to account for the peaks in the Christmas season. The dummy constitutes the month of December. An alternative model C includes the average weekly price per product group as managers deliberately determine prices in advance. The last model D includes all explanatory variables. The model takes on the following form including the search volume:

$$\mathrm{sales}_t = \alpha + \beta_1\,\mathrm{sales}_{t-1} + \beta_2 \mathrm{Christmas} + \beta_3 \mathrm{price} + \beta_4\,\mathrm{search}_{t-d} + \epsilon \qquad (7)$$

where α and β are parameters to be estimated and ϵ is the estimation error. d is at least 1 to secure practical application. Based on a simple linear regression of sales as a function of search volume, the product groups audio books, calendars, hifi systems, and smartphones show the highest R^2 for the same week compared to a lag d of 1, 2, 3, and 4 – termed nowcasting by Choi and Varian (2012).

$$\mathrm{sales}_t = \alpha + \beta\,\mathrm{search}_{t-d}\,\epsilon \qquad (8)$$

However, to compare models under practical application, we use search_{t-1} for model estimation. Blu-ray players and children's books demonstrate highest R^2 (d = 1), whereas R^2 for TVs was highest for d = 3.

4 Results

To compare the fit with and without search data, model estimates are calculated on the entire period as well as a split of 88 weeks for testing and 17 weeks for validation. The study also uses the SPSS expert modeler to identify an alternative ARIMA model for sales of every product group.

For the seven product groups, Table 3 displays the results for Model D with and without search data over the entire data period of 105 weeks. The table also reports R^2, the p value for the Ljung-Box Q statistics, the Root Mean Square Error (RMSE), the Mean Absolute Percentage Error (MAPE), and Normalized Bayesian Information Criterion (BIC).

Even though forecasting accuracy should be improved further, search data considerably improves sales explanations. The average improvement in R^2 is 17.03 %, ranging from 3.49 % for calendars to 39.64 % for children's books. The relative percentage improvement in RMSE is -7.68 % and -7.92 % in MAPE. The average weekly price in a product group does not provide any substantial forecast improvement. A potential explanation is that this study used the average weekly price in a product group. Retailers may specifically include the actual planned prices. The results are altogether similar for models A, B, and C without and with search data. Similar results are also obtained in all cases for the test and validation split.

For the product group calendar, Ljung-Box test ($p = .001$) indicates some form of residual autocorrelation. The present study considers calendars as somewhat of a control group because the purchasing behavior is very predictable. Calendars are predominantly bought at the end of the old year and the beginning of the new year. Consequently, Figure 1 shows a two-u-shaped figure for calendars. Sales also spike in every other product group at the end of the year, although these spikes are less pronounced for television sales. The intra year movement could alternatively be explained with seasonal or monthly dummy variables. Similarly to previous results (Fantazzini and Toktamysova, 2015), search data explains part of the variance in sales data, as search data significantly adds to the results in all product groups. However, further research should address whether alternative approaches beyond the straightforward approach confirm this result. Even though these approaches will not identify relevant consumer trends, they will provide some insights into the development of product demand with short lead time.

For an alternative ARIMA model, SPSS expert modeler is applied to model B – thus dropping the average weekly price. The dummy for the Christmas holiday season is treated as an event handle. An ARIMA (0,0,0) may represent sales of audio books and hifi systems. The ARIMA (1,0,0) is suggested for children's books and smartphones. The product group of blu-ray players seems to follow an ARIMA (1,0,1) process in the present sample, whereas TVs follow an (0,1,1) process. The modeler proposes an ARIMA (0,1,4) for calendars. Overall, the data-driven approach suggests directions for modeling improvement indicating various processes for the different product groups. However, the data-driven results also indicate opportunities for improvement - most apparently apparent in the constant estimation in the case of audio books. Table 4 presents these model results.

5 Outlook

Motivated by the retail perspective to improve storage and shelf efficiency, search volume is considered in forecasting durable product sales. The results indicate that search data from user queries in search engines is useful for sales forecasting. In general, search data from search engines is freely available, detailed, and available in real time. As demonstrated, the application is simple and it increases forecast quality. However, an important point to note is that search data is only indicative of consumer interest, but is not necessarily causal. In particular, the shorter the information and decision phases, the lesser the likelihood of search data to add forecasting accuracy. In other words, the expected variance of such data increases and, simultaneously, the forecast horizon shortens. These effects pose some primary challenges for companies, when applying search data in the case of durable goods with short decision phases.

Researchers and managers could probably improve the data quality by drawing on search data from online retailers, such as Amazon or Alibaba. Instead of household panel data, similar data is alternatively available from customer cards, point-of-sale scanning, and inventory management systems.

Naturally, additional data can improve the forecasting accuracy further. Researchers and managers will include additional and available causal data that affects either sales or search volume. Advertising, for example, can affect both sales and search volume. Print advertising has been shown to influence sales directly as well as indirectly via search engines (Olbrich and Schultz, 2014).

Table 3: Model D Results without and with Search Data.

	Audio book		Blu-ray player		Calendar		Children's book		Hifi system		Smartphone		TV	
	w/o	with	w/o	with	w/o	with	w/o	with	w/o	with	w/o	with	w/o	with
$Sales_{t-1}$	0.26*	0.22*	0.38*	0.25*	0.94*	0.88*	0.55*	0.26*	0.30*	-0.01	0.28*	0.21*	0.46*	0.35*
Christmas	9.54*	4.65	7.86*	3.98	-0.01	27.59	13.33	7.78	14.26*	3.25	18.45*	12.90*	6.15	4.27
Price	-0.14	-0.18	-0.00	-0.01	-1.71	-0.90	-1.63	-2.26	-0.01	-0.02	-0.06	-0.08	-0.01	-0.01
$Search_{t-d}$	-	0.27*	-	0.18*	-	4.41*	-	2.57*	-	0.33*	-	0.50*	-	0.33*
Constant	19.47*	0.43	14.65*	9.19*	145.36*	-171.85*	163.84*	41.09	21.72*	8.93*	65.97*	32.34	35.61*	16.49*
R^2	.197	.227	.364	.424	.884	.916	.300	.497	.349	.440	.269	.298	.304	.372
Ljung–Box p	.982	.931	.286	.515	.198	.001*	.256	.297	.145	.369	.491	.469	.168	.144
RMSE	8.24	8.12	5.24	5.01	45.18	38.68	46.76	39.83	8.12	7.56	12.04	11.86	8.76	8.37
MAPE	40.97	40.50	35.39	33.04	88.67	82.11	22.98	20.44	33.96	29.57	17.34	16.65	24.70	22.93
BIC	4.40	4.41	3.49	3.44	7.80	7.53	7.87	7.59	4.37	4.27	5.15	5.17	4.52	4.47

Notes: $^*p < .05$; $d = 3$ for TV and $d = 1$ in all other product groups; n: 105 weeks.

Table 4: Data-driven ARIMA Models suggested by SPSS Expert Modeler.

Product Group	ARIMA (p,d,q)	Suggestion		Var(E_t)
Audio book	(0,0,0)	sales_t	$= 2,773 + 0.483\ \text{Christmas}$	–
Blu-ray player	(1,0,1)	sales_t	$= 14,976 + 0,899\ \text{sales}_{t-1} + E_t + 0.654 E_{t-1} + 7.085\ \text{Christmas}$	24.737
Calendar	(0,1,4)	$\Delta\sqrt{\text{sales}_t}$	$= E_t - 0.275 E_{t-4} + 2.753 \Delta\sqrt{\text{sales}_{t-1}} - 1.811 \Delta\sqrt{\text{sales}_{t-3}}$	1.663
Children's book	(1,0,0)	$\ln \text{sales}_t$	$= 2.134 + 0.352 \ln(\text{sales}_{t-1}) + 0.407 \ln(\text{search}_{t-1}) + E_t + 1.251 E_{t-1} - 0.811 E_{t-2}$	0.046
Hifi system	(0,0,0)	$\ln \text{sales}_t$	$= 0.796 \ln(\text{search}_{t-1})$	–
Smartphone	(1,0,0)	sales_t	$= 55.354 + 0.279\ \text{sales}_{t-1} + 17.088\ \text{Christmas}$	–
TV	(0,1,1)	$\Delta\ \text{sales}_t$	$= E_t + 0.687 E_{t-1}$ 78.058	–

Notes: *n*: 105 weeks; Christmas represented as an event.

Consumer characteristics, such as purchase experience, further impact buying decisions. For example, as purchase experience increases or similar purchase involvement decreases, consumers reduce their information and decision time and are less likely to employ search engines. One challenge of consumer characteristics is the availability of such data in good time, especially for short decision phases. Potential ways to infer such data are household panel data and customer cards. Customer relationship management systems may also provide such characteristics, but are generally more available for durable goods with longer decision phases or in business to business settings.

As subsequently pointed out, the forecasting accuracy will additionally depend on characteristics of the product groups. To increase shelf and storage efficiency, target product groups need to be of relevance regarding their profit margin as well as taking up shelf or storage space. Consequently, the saving potential from shelf and storage efficiency is expected to be higher in product groups such as hifi systems and televisions than with audio and children's books. Similarly, products should not only possess some decision time but also have a reasonable purchase frequency, generally measured by their turnover rate. Product groups thus need to be relevant considering their contribution margin, turnover rate, and space requirements to have the potential to create monetary efficiency through shelf and storage management.

Using household panel data to predict sales of durable goods follows the idea of general economic indicators as discussed in previous studies (see Table 1). However, the present study moves one step further by drawing on real sales data on a household level. Household panel data provides a predictive indicator of market development. Further research is directed at the retailer level.

From the retail perspective, research should address the short-term applicability of the approach proposed here. For example, does search data also improve existing forecasting approaches and what is its relative contribution to these approaches? Research may further provide some insights about the potential volume of (monetary) storage and shelf efficiency. Another open question is what are appropriate product groups? Space requirements and turnover ratios are likely to influence the choice of product groups beyond the extent of the information and search phase.

As for the application of the approach presented, retailers need to identify product groups with the discussed attributes, such as decision time, space requirements, turnover rates, and contribution margins. Search data can then be included in the forecasting approach regarding procurement and distribution. Retailers can also extend the time series and thus improve forecasting. Shelf efficiency can be realized if retailers are able to assign the necessary space for the product group and potentially create new space to extend their product range. Storage efficiency, on the other hand, may be realized if storage space can be freed to store alternative products that are likely to run out of stock.

Acknowledgements The author acknowledges the support of Philipp Brüggemann in data processing and thanks the reviewers and participants of the European Conference on Data Analysis 2019 for their helpful comments. The author is also thankful for the feedback of two anonymous reviewers that helped to improve the manuscript further.

References

Asur S, Huberman BA (2010) Predicting the future with Social Media. In: Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 01, IEEE Computer Society, Washington, DC, USA, WI-IAT '10, pp. 492–499. ISBN: 978-0-769541-91-4, DOI: 10.1109/WI-IAT.2010.63.

Boone T, Ganeshan R, Hicks RL (2015) Incorporating Google Trends data into sales forecasting. Foresight: The International Journal of Applied Forecasting 38(3):9–14. URL: `https://foresight.forecasters.org/product/foresight-issue-38/`.

Calomiris CW, Mamaysky H (2019) How news and its context drive risk and returns around the world. Journal of Financial Economics 133(2):299–336. DOI: 10.1016/j.jfineco.2018.11.009.

Choi H, Varian H (2012) Predicting the present with Google Trends. Economic Record 88(Special Issue):2–9. DOI: 10.1111/j.1475-4932.2012.00809.x

Dietzel MA (2016) International sentiment-based predictions of housing market turning points with Google Trends. Journal of Housing Markets and Analysis 9(1):108–136. DOI: 10.1108/IJHMA-12-2014-0058.

D'Amuri F, Marcucci J (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.

Fan ZP, Che YJ, Chen ZY (2017) Product sales forecasting using online reviews and historical sales data: A method combining the Bass model and sentiment analysis. Journal of Business Research 74:90–100. ISSN: 0148-2963, DOI: 10.1016/j.jbusres.2017.01.010.

Fantazzini D, Toktamysova Z (2015) Forecasting German car sales using Google data and multivariate models. International Journal of Production Economics 170, Part A:97–135. DOI: 10.1016/j.ijpe.2015.09.010.

Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L (2009) Detecting influenza epidemics using search engine query data. Nature 457(7232):1012–1014. DOI: 10.1038/nature07634.

Hand C, Judge G (2012) Searching for the picture: Forecasting UK cinema admissions using Google Trends data. Applied Economics Letters 19(11):1051–1055. DOI: 10.1080/13504851.2011.613744.

Hennig-Thurau T, Wiertz C, Feldhaus F (2015) Does Twitter matter? The impact of microblogging word of mouth on consumers' adoption of new movies. Journal of the Academy of Marketing Science 43(3):375–394. DOI: 10.1007/s11747-014-0388-3.

Hu Y, Du RY, Damangir S (2014) Decomposing the impact of advertising: Augmenting sales with online search data. Journal of Marketing Research 51(3):300–319. DOI: 10.1509/jmr.12.0215.

Kulkarni G, Kannan PK, Moe W (2012) Using online search data to forecast new product sales. Decision Support Systems 52(3):604–611. DOI: 10.1016/j.dss.2011.10.017.

Oestmann M, Bennöhr L (2015) Determinants of house price dynamics. What can we learn from search engine data? Review of Economics 66(1):92–127. DOI: 10.1515/roe-2015-0106.

Olbrich R, Schultz CD (2014) Multichannel advertising: Does print advertising affect search engine advertising? European Journal of Marketing 48(9/10):1731–1756. DOI: 10.1108/EJM-10-2012-0569.

Önder I, Gunter U (2016) Forecasting tourism demand with Google Trends for a major European city destination. Tourism Analysis 21(2-3):203–220. DOI: 10.1002/jtr.2137.

Park S, Lee J, Song W (2017) Short-term forecasting of Japanese tourist inflow to South Korea using Google Trends data. Journal of Travel & Tourism Marketing 34(3):357–368, Routledge. DOI: 10.1080/10548408.2016.1170651.

Schultz CD (2007) Consumer Trust in E-Commerce - An Analysis of Means Communicating Trustworthiness From a Buying Transaction Life Cycle Perspective. Dr. Kovač. ISBN: 978-3-830030-14-0.

Sul HK, Dennis AR, Yuan L (2017) Trading on twitter: Using Social Media sentiment to predict stock returns. Decision Sciences 48(3):454–488. DOI: 10.1111/deci.12229.

Tirunillai S, Tellis GJ (2012) Does chatter really matter? Dynamics of user-generated content and stock performance. Marketing Science 31(2):198–215. DOI: 10.1287/mksc.1110.0682.

Vosen S, Schmidt T (2011) Forecasting private consumption: Survey-based indicators vs. Google Trends. Journal of Forecasting 30(6):565–578. DOI: 10.1002/for.1213.