



UNIVERSITY OF
BIRMINGHAM

COLLEGE OF
SOCIAL SCIENCES

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

Viktor Pekar

Business School, University of Birmingham

v.pekar@bham.ac.uk

Why forecast consumer expenditure?

- Consumer spending constitutes the largest component of GDP in developed countries
 - E.g., 70% in the US, 66% in UK, 60% in Germany (Pistaferri, 2015)
- An important component in models of various macroeconomic variables, e.g., inflation or money supply
- The forecasts are used by governments and central banks for different policy-making decisions
- Government organizations and market research agencies measure consumer spending on a regular basis (e.g., in the UK, Office for National Statistics compiles Household Final Consumption Expenditure Index).

Traditional surveys of consumer expenditure

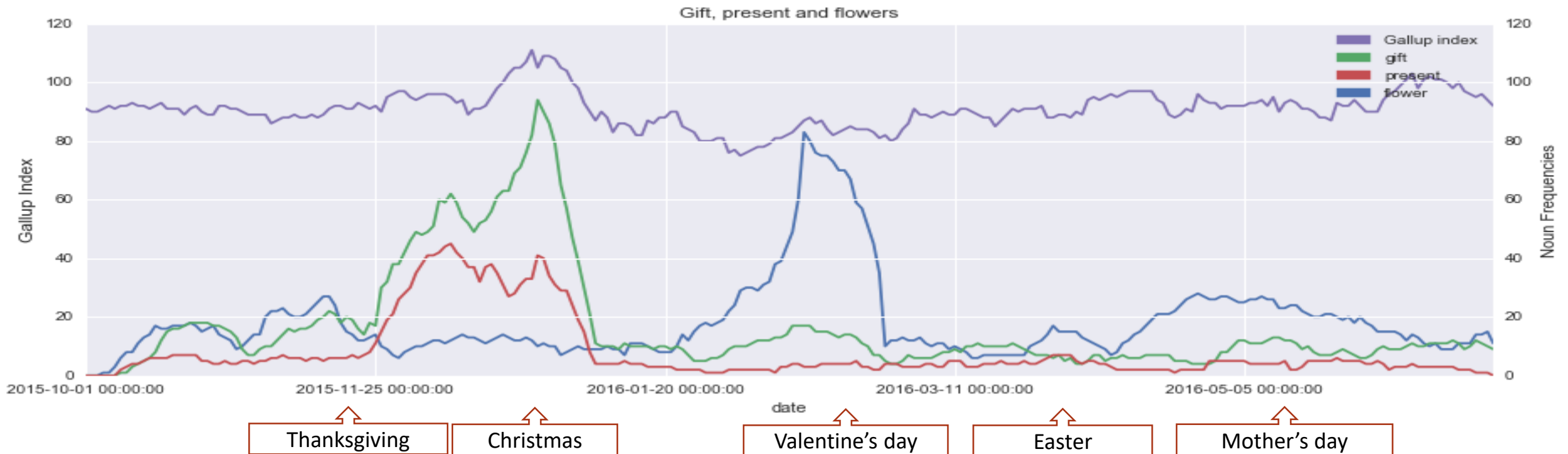
Household Final Consumption Expenditure Index (UK):

- On a quarterly basis
- 1,500 households
- Via telephone interviews
- Interviewees are asked to reflect on their recent purchases and describe expected purchases in the future
- Drawbacks:
 - Based on low-frequency observations
 - Delayed publication
 - Costly to organize

Estimating consumer spending from social media and search engine queries

- User-generated content (USG) - social media data, search engine queries – can provide an alternative to traditional surveys
- Responses obtained from USG can potentially overcome the drawbacks of traditional surveys:
 - they can be obtained near-real time,
 - high-frequency,
 - they contain unsolicited hence unbiased opinions
- Our aim is to forecast consumer spending based on detecting purchase intentions expressed on social media and in the form of search engine queries

Buying “present”, “gift”, “flowers”



Related work: search engine data

- A search engine query is often assumed to be an indicator of a purchase intention.
- Vosen and Schmidt (2011) use search results for a broad set of Google Trends (GT) categories to model private consumption in USA, and find they are better predictors than the Michigan Consumer Sentiment Index and Conference Board Consumer Confidence Index.
- Carrière-Swallow and Labbe (2013) build a ARMA model for automobile sales in Chile and show that the introduction of an exogenous GT variable improves both in- and out-of-sample forecasts.
- Wu and Brynjolfsson (2015) predict real-estate sales by incorporating GT data into an AR model, along with other exogenous variables such as housing price index.
- The GT categories are compiled in an application-independent manner, and thus may be unlikely to reflect specifics of a particular economic phenomenon being modelled.

Related work: sentiment analysis on social media

- Modelling Consumer Economic Confidence via sentiment expressed on social media, which itself is a predictor of consumer expenditure
- Overall approach is to automatically assess sentiment in posts containing keywords referring to economy (“economy”, “job”, etc)
- A daily sentiment index is then compiled and used as a variable in a regression model, on its own or along with commonly used variables
- O’Connor et al. (2010) predicts consumer confidence from sentiment found in economy-related Twitter posts: a daily sentiment index is used as a predictor in an OLS model of a consumer sentiment index.
- Daas and Putz (2014) use a commercial sentiment analysis system on tweets containing economy-related keywords, and show that co-integration exists between the sentiment index and an official index of consumer confidence

Purchase intentions on Google Trends

- 18 subcategories of the top-level “Shopping” category, e.g. "Apparel", "Consumer Electronics", "Luxury Goods", "Ticket Sales"
- Restricted to queries originating from the US
- Daily search volume for the period between 01.01.2014 and 31.07.2017 (1,310 days)
- The search volume on each category is a separate exogenous variable in the model

Purchase intentions on Twitter

Detection of purchase intentions

- Twitter Search API, restricting search to messages originating in the US
- Generated queries following the template: <I,we> + <will, plan, intend, ...> + <buy, purchase, ...>
- The same period as for GT data, between 01.01.2014 and 31.07.2017
- Removed negated statements, checking for presence of negation words between the subject and the main verb
- Heads of noun phrases following the “buy” verbs are extracted and their counts are recorded for each day
- E.g., "headphones" in "I am looking to buy new headphones"

Purchase intentions on Twitter

Semantic vectors

- Kept 1,000 most common nouns in the data to reduce noise resulting from parsing errors, non-grammatical text, etc.
- Word2vec (Mikolov et al 2013) model of words, trained on a large Twitter corpus, 100 dimensions
- A word is represented via a fixed-length vector
- Each extracted noun is mapped to a word2vec vector
- Vectors of nouns observed on a particular day are averaged, so each day's purchase intentions are represented by a single 100d vector
- The step captures semantic similarity of the nouns and helps to reduce model complexity
- The semantic vector constitutes exogenous variables in the model of consumer spending

Intention lag

- We investigate the optimal 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

Experiment design

- Target variable: Gallup's daily Consumer Spending Index (CSI) for US for 01.01.2014-31.07.2017
- Train-validation-test split: 60%-20%-20%
- Regression method: Support Vector Regression, a popular machine learning method for supervised learning
- During evaluation, we experimentally determine free parameters of SVR (the cost parameter, the gamma parameter and the kernel type) on a validation dataset using grid search

Experiment design

- Baseline: SVR trained only on autoregressive variables of CSI
- Because CSI displays weekly seasonality we use 7 lags of CSI as autoregressive variables
- Dynamic forecasting ($h = 7, 14, \text{ and } 28$ days)
- Evaluation metrics: Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE)
- Modelling strategies: level, differenced, detrended, season-adjusted, detrended+season-adjusted
- Detrending and seasonal adjustment was performed using STL (Cleveland et al 1990)

Results: Modelling strategies

	h=7		h=14		h=28	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
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	10.5	7.96	11.48	8.64	10.96	8.23

Results: Effect of GT variables

	h=7		h=14		h=28	
Intention lag	RMSE	MAE	RMSE	MAE	RMSE	MAE
0	10.49	8.03	10.9	8.19	10.57	8
1	10.21	7.69	10.81	8.06	10.52	7.85
2	10.4	7.64	10.78	7.97	10.7	7.98
3	10.56	8	10.78	8.02	10.68	8.02
4	10.84	8.3	11.15	8.34	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	10.9	8.24
7	11.19	8.58	11.28	8.39	11.09	8.33

Error reduction at intention lag = 1

	-2.8%	-3.4%	-5.8%	-6.7%	-4.0%	-4.6%
--	--------------	--------------	--------------	--------------	--------------	--------------

Results: Effect of Twitter variables

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

Error reduction at intention lag = 0

	-1.5%	-2.5%	-4.9%	-5.0%	-0.7%	-1.0%
--	--------------	--------------	--------------	--------------	--------------	--------------

Discussion and conclusions

- Presented a study comparing indicators of purchase intentions obtained from Google Trends to those obtained from Twitter using NLP analysis of the messages
- 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
- GT variables seem to be more useful predictors than the semantic variables extracted from Twitter posts, although the differences in performance are not very large

Future work

- Analysis of more fine-grained Google Trends subcategories
- Introduction of GT and Twitter variables into more complex models of consumer spending
- Longer-term forecasts: weeks and months, instead of days

Thank you!
Questions? Comments?