Proceedings of Machine Learning Research 58 (2016) 31-40

Forecasting Spot Oil Price Using Google Probabilities

Krzysztof Drachal KDRACHAL@WNE.UW.EDU.PL
Faculty of Economic Sciences
University of Warsaw
Warsaw, Poland

Editor: Tatiana V. Guy, Miroslav Kárný, David Rios-Insua, David H. Wolpert

Abstract

In this paper DMA (Dynamic Averaging Model) is expanded by adding certain probabilities based on Google Trends. Such a method is applied to forecasting spot oil prices (WTI). In particular it is checked whether a dynamic model including stock prices in developed markets, stock prices in China, stock prices volatility, exchange rates, global economic activity, interest rates, production, consumption, import and level of inventories as independent variables might be improved by including a certain measure of Google searches. Monthly data between 2004 and 2015 were analysed. It was found that such a modification leads to slightly better forecast. However, the weight ascribed to Google searches should be quite small. Except that it was found that even unmodified DMA produced better forecast than that based on futures contracts or naive forecast.

Keywords: Dynamic Model Averaging, Dynamic Model Selection, Google Trends, Internet Search Data, Spot Oil Price

1. Introduction

The oil market is quite a complex one. As a result, for practitioners there is no commonly accepted model of forecasting spot oil prices. Usually, forecasts based on futures contracts are used (Yang et al., 2002). Despite that obstacles, well-performing forecasts are crucial for the whole energy market, oil-importing and oil-exporting countries, macroeconomic forecasters, etc. (Alquist et al., 2013).

Generally, various forecasting methods on the oil market can be classified as: time series models, financial models, structural models, artificial neural networks based models, support vector machines models and qualitative method. The review of them can be found, for example, in papers of Drachal (2016a) or Behmiri and Pires Manso (2013).

In this paper quite a novel method, called DMA (Dynamic Model Averaging) is explored (Raftery et al., 2010). However, following Koop and Onorante (2014) the original method is slightly modified to include Internet searches. Recently, Internet search data were extensively applied in economics (Pavlicek and Kristoufek, 2015; Bangwayo-Skeete and Skeete, 2015; Scott and Varian, 2015; Choi and Varian, 2012; Dangl and Halling, 2012; Wu and Brynjolfsson, 2010; Choi and Varian, 2009; Schmidt and Vosen, 2009).

Actually, also for the oil market Google searches were applied (Li et al., 2015; Fantazzini and Fomichev, 2014), but not within the context of DMA. In particular, herein Google variables are not used as independent variables in the regression equation, but they are used to construct certain probabilities, which are used in computation of posteriori probabilities in DMA recursive estimation. This is motivated by the assumption that a surge in searches about certain determinant might indicate the relevance of this variable in the model.

Indeed, many researches indicated that the impact of various determinants of the oil price might change in time (Aastveit and Bjornland, 2015; Zhang and Wu, 2014; Baumeister and Peersman, 2013). DMA seems to be a very good method in such a case. Actually, the interest in this method grows rapidly in finance (Naser, 2016; Aye et al., 2015; Bork and Moller, 2015; Baur et al., 2014; Koop and Korobilis, 2012; Nicoletti and Passaro, 2012).

Herein, in particular the following problem was addressed: Do Google Trends might somehow improve DMA forecast of the spot oil price? This question is positively answered further in the text.

2. Data

Now, a short review of potential oil price determinants is presented. This serves as an argument for the further data selection to the model.

According to Hotelling (1931) the price of non-renewable commodity should depend on the interest rate. Up to 1980s it was commonly agreed that the most important oil price determinant is OPEC decisions. Later attention was shifted to gross domestic product, stock market activity and exchange rates (Bernabe et al., 2004; Yousefi and Wirjanto, 2004), as well as, emerging markets (Basher et al., 2012).

Indeed, emerging markets were suggested to significantly impact the oil market (Li and Leung, 2011). There is some evidence that oil price changes between 2007 and 2008 might have happened due to the halt in the Chinese demand and supply (Kaufmann, 2011).

Within this context the global economic activity was also analysed. One of the approaches, which allows to use monthly frequency data is by the Kilian index (He et al., 2010; Kilian, 2009).

Moreover, Du and He (2015), as well as many other researchers focused on the impact of stock market volatility. Recently, much attention was put on speculative pressures (Carmona, 2015; Kilian and Murphy, 2014; Fattouh et al., 2013). Usually, they were measured by the level of inventories (Hamilton, 2009).

Much more extensive literature review on the oil price determinants within the context of a time-varying framework can be found, for example, in a paper by Drachal (2016a).

According to the presented literature review 10 potential oil price determinants were identified (Tab. 1). Strategic Petroleum Reserves were excluded from the level of inventories (Bu, 2014). Monthly data beginning on 2004 and ending on 2015 were taken. (Only for IMP the weekly data were aggregated to monthly one by taking the daily mean for the corresponding month.) This results in relatively short (in the context of DMA approach) data set. However, this is an obstacle that cannot be overcome, because Google Trends date back to 2004 only. (This is a common problem for all research with Google Trends statistics.)

If there is much easily available data for U.S., there is a lack of global time series. However, following, for example, Hamilton (2009) and Kilian and Murphy (2014), U.S. data can be taken as proxies.

The estimated forecast was compared to the one obtained by NFP (Alquist et al., 2013), i.e., 1-month NYMEX WTI futures prices (in USD).

Table 1: Data description

name	description
WTI	WTI spot price (in USD)
MSCI	MSCI World Index
TB3MS	U.S. 3-month treasury bill secondary market rate (in percentages)
KEI	Kilian index of global economy activity (Kilian, 2009)
TWEXM	trade weighted U.S. dollar index (Mar, 1973 = 100)
PROD	U.S. crude oil production (in 1'000 barrels)
IMP	daily average of U.S. crude oil import (in 1'000 barrels / day)
INV	U.S. total ending stocks of commercial crude oil (in 1'000 barrels)
VIX	implied volatility of S&P 500
CONS	total consumption of petroleum products in OECD (in quad BTU)
CHI	Shanghai Composite Index

It should be noticed that DMA does not require data to be stationary. On the other hand, it is desirable to normalize data, as this might significantly improve the outcomes (Drachal, 2016b).

In case of search terms the following ones were taken. For MSCI: "stocks", "developed markets", "msci index", "stock prices", "stock market", "stock quotes", "equity performance". For TB3MS: "market rates", "interest rates", "cpi", "inflation", "bond rates", "treasury bill", "fed", "libor". For KEI: "world economic activity", "gdp growth", "economic activity", "economy", "economic growth", "business cycle", "industrial production". For TWEXM: "exchange rates", "USD". For PROD: "oil production", "energy production", "oil supply", "opec". For IMP: "oil import". For INV: "oil inventories", "oil speculation". For VIX: "stock market volatility", "market stress", "market risk", "implied volatility", "vix", "volatility index". For CONS: "oil consumption", "energy consumption", "oil demand", "opec". For CHI: "china", "chinese economy", "china market", "shanghai composite index". For WTI: "wti", "oil price", "crude oil price". Then for each variable the mean of corresponding Google Trends was computed. This choice of search terms is quite general and arbitrary. It might be desirable to include more search terms. However, it was left for the further and much more extensive study, as even the choice of search terms is quite a challenging task (Stephens-Davidowitz and Varian, 2015).

All calculations were done in R (2015) software.

3. Methodology

Herein, just a brief sketch of DMA (Dynamic Model Averaging) and DMS (Dynamic Model Selection) is presented in order to explain the proposed modification with Google probabilities. The detailed explanation can be found in the original paper (Raftery et al., 2010).

Let there be m determinants. Then, $K = 2^m$ different models can be constructed, including the one with constant solely. Let us index time by $_t$, and let the dependent variable (the oil price) be y_t. Let x_t^k denote determinants in the kth model ($k = \{1, \dots, K\}$). Notice, that for independent variables 1st lags were taken in all estimated models. Then, the state space model is given by

$$y_t = x_t^k \theta_t^k + \epsilon_t^k \quad , \tag{1}$$

$$\theta_t^k = \theta_{t-1}^k + \delta_t^k \quad . \tag{2}$$

The regression parameters are denoted by θ_t^k and $\epsilon_t^k \sim N(0, V_t^k)$ and $\delta^k \sim N(0, W_t^k)$. Initially, $V_0^k := I$ and W_0^k is set according to the algorithm given by Raftery et al. (2010). I denotes the unit matrix. (Setting $V_0^k := I$ is reasonable, because data were normalized. In particular, let Y_t denote the core data. Then, the normalization is done with the formula $y_t := \frac{Y_t - \min(Y_0,\ldots,Y_t,\ldots)}{\max(Y_0,\ldots,Y_t,\ldots)-\min(Y_0,\ldots,Y_t,\ldots)}$.) Further, V_t^k is estimated by a recursive method of moments estimator, and W_t^k by the Kalman filter updating. This needs a certain forgetting factor $\lambda \in (0, 1]$ to be specified (Raftery et al., 2010; Dedecius et al., 2012). (Notice, that if $\lambda = 1$ there is no forgetting.)

The estimation is done recursively. First, it is set $\pi_{0|0,k} := \frac{1}{K}$. Then, it is proceeded with

$$\pi_{t|t-1,k} = \frac{(\pi_{t-1|t-1,k})^\alpha + c}{\sum_{i=1}^K (\pi_{t-1|t-1,i})^\alpha + c} \quad , \tag{3}$$

$$\pi_{t|t,k} = \frac{\pi_{t|t-1,k} f_k(y_t|y_0, y_1, \ldots, y_{t-1})}{\sum_{i=1}^K \pi_{t|t-1,i} f_i(y_t|y_0, y_1, \ldots, y_{t-1})} \quad . \tag{4}$$

The above equations contain the second forgetting factor $\alpha \in (0, 1]$ and $f_k(y_t|y_0, y_1, \ldots, y_{t-1})$ denotes the predictive density of the kth model at y_t given the data from the previous periods. $\pi_{t|t,k}$ are called posteriori inclusion probabilities. Also, a small constant is specified, for example, $c := K \cdot 10^{-3}$ in order to avoid reducing the probabilities to zero due to numerical approximations during computations.

Then, the DMA forecast is formulated in the following way

$$\widehat{y_t} = \sum_{k=1}^K \pi_{t|t-1,k} \widehat{y_t^k} \tag{5}$$

where $\widehat{y_t^k}$ is the forecast produced by the kth model.

According to Raftery et al. (2010), if $\alpha = 1 = \lambda$, there is no forgetting, and the method is still recursive but not dynamic.

However, let $\widehat{\pi_{t|t-1,k}} := \max_{i=\{1,\ldots,K\}} \{\pi_{t|t-1,i}\}$, where $\pi_{t|t-1,i}$ are computed as in Eq. 3. In DMS the Eq. 5 is modified in the following way

$$\widehat{y_t} = \widehat{\pi_{t|t-1,k}} \widehat{y_t^k} \quad . \tag{6}$$

Now, let g_i be the Google Trends variable corresponding to Internet search of ith term. This variable is between 0 and 100. So, it can be easily rescaled to fit between 0 and 1 and interpreted as Google probability (Koop and Onorante, 2014). (Notice that Google Trends correspond not to the absolute volume of Internet search terms, but to the relative one, i.e., relative to all Internet searches.) Now let

$$p_{t,k} := \prod_{i \in IN} g_i \cdot \prod_{j \in OUT} (1 - g_j) \quad , \tag{7}$$

where IN correspond to variables included in the kth model at time t, and OUT correspond to variables not included in the kth model at time t.

Then, Eq. 3 can be modified in the following way

$$\pi_{t|t-1,k} = \omega \cdot \frac{(\pi_{t-1|t-1,k})^{\alpha} + c}{\sum_{i=1}^{K}(\pi_{t-1|t-1,i})^{\alpha} + c} + (1-\omega) \cdot p_{t,k} \quad , \tag{8}$$

where ω is a parameter from $[0,1]$. In this way DMA with Google probabilities and DMS with Google probabilities are obtained, by repeating the rest of procedures since Eq. 3 unmodified. If $\omega = 1$ then the standard DMA (DMS) is obtained. If $\omega = 0$ the procedure is highly changed and Google Trends play the whole role.

4. Results

First of all, it was imposed in all tested models that $\alpha = \lambda$, because these parameters correspond to the weight of information from the past that is put in the present. Therefore, this restriction can be nicely interpreted.

First, DMA models were estimated with variables as that in Tab. 1. Initially, models with $\omega = \{1, 0.75, 0.50, 0.25, 0\}$ and $\alpha = \{1, 0.99, \ldots, 0.90\} = \lambda$ were estimated. Unfortunately, in all cases the modification of DMA with Google probabilities lead to worse forecast (with respect to MSE), i.e., for all $\alpha = \lambda$ the smallest MSE was given by the model with $\omega = 1$, and decline in ω resulted in higher MSE for all $\alpha = \lambda$ fixed. The model with the best forecast (with respect to minimise MSE) was that with $\alpha = 0.93 = \lambda$. In such a case MSE was 0.0072153. Unfortunately, it is higher than that of the naive forecast, i.e., 0.0043214.

Therefore, having fixed $\alpha = 0.93 = \lambda$ more detailed examination was performed, i.e., with $\omega = \{0.99, 0.98, 0.97, 0.96, 0.95, 0.90, 0.85, 0.80\}$. It was found that for high ω slight forecast improvements are present. In particular, if $\omega = 0.99$ then the forecast of DMA with Google probabilities is approximately 2% better than the forecast of a standard DMA, i.e., it is 0.0070549. However, it is better than the forecast based on futures contracts (its MSE = 0.0118336).

Notice that, if $\mathcal{E}_t$ are residuals for Y_t, and ϵ_t are residuals for y_t, where $Y_t = a \cdot y_t + b$ (which corresponds to normalization), then $\mathcal{E}_t = a \cdot \epsilon_t$. So, the futures based forecast can be computed for the initial data and then its errors can be rescaled to be comparable with the estimated DMA/DMS models.

Still, this is a very small improvement. Moreover, the weight put to Google probabilities in Eq. 8 is marginal.

As a result, in this case it cannot be said that Google internet searches lead to significant improvement of DMA forecast performance. However, in case of relatively small $\alpha = \lambda$ even other (but still high ones) values of $\omega \neq 1$ lead to outperforming the forecast based on futures contracts. (Due to the limited space the details are not reported.)

Following Koop and Onorante (2014) it was also checked whether switching to DMS with Google probabilities would give better forecast. Unfortunately, for $\alpha = 0.93 = \lambda$ and $\omega = 0.99$ the forecast is slightly worse than in the corresponding case of DMA with Google probabilities (MSE = 0.0070922).

Some attempts were taken to improve the above model, for example, by reducing the set of oil price determinants up to those variables which met the following criterion. Notice

Table 2: MSE for DMA

$\omega \setminus \alpha = \lambda$	1	0.99	0.98	0.97	0.96	0.95
1	0.0043768	0.0042989	0.0042445	**0.0042226**	0.0042234	0.0042408
0.75	0.0059041	0.0055031	0.0051488	0.0048673	0.0046572	0.0045124
0.50	0.0089551	0.0079541	0.0070400	0.0062966	0.0057389	0.0053560
0.25	0.0121312	0.0104178	0.0088705	0.0076298	0.0067211	0.0061154
0	0.0148107	0.0125550	0.0105091	0.0088525	0.0076452	0.0068500

that, posteriori inclusion probabilities for every model which contains a given variable can be summed. Now, let reduce the set of oil price determinants up to the variables which posteriori inclusion probabilities are over 50% for most of the time since 2010. (Before 2010 it could have been observed that there is much variation of these probabilities, because the model "learns".) As a result, the set of determinants was reduced to MSCI, TB3MS, TWEXM and CHI. Unfortunately, it did not lead to any significant improvement of forecast.

Therefore, following Drachal (2016b) the following oil price determinants were considered: 1st and 2nd lags of WTI, MSCI, CHI, and 2nd lag of VIX. It can be seen that in such a case, if $\alpha = \{1, 0.99, \ldots, 0.90\} = \lambda$ and $\omega = \{1, 0.75, 0.50, 0\}$ the naive forecast was outperformed, but only if $\omega = 1$. (See Tab. 2.) The best forecast was obtained for $\alpha = 0.97 = \lambda$ and $\omega = 1$. (Due to the limited space herein only the most important part of results is presented.)

Switching to DMS method did not help, as for the estimations with the same spread of parameters α, λ and ω the minimal MSE was 0.0044905.

Further, $\omega = \{0.99, 0.98, \ldots, 0.90\}$ for fixed $\alpha = 0.97 = \lambda$ was examined (see Tab. 3). It can be seen that the best forecast was given by DMA with $\omega = 0.96$. However, this lead to only 2% improvement with respect to MSE in comparison with DMA without Google probabilities. Moreover, the weight put to Google probabilities is very small.

Finally, it was also checked if DMS model with $\alpha = 0.97 = \lambda$ and $\omega = 0.96$ would give better forecast. It was not so, as its MSE = 0.0046273.

5. Conclusions

It can be concluded that, even if Google probabilities can slightly improve DMA performance (by approx. 2%) the current findings are not much amazing, as this improvement is marginal. However, it can be expected that further investigation will provide better results. Due to the limited space herein, they will be presented elsewhere.

For further research it seems that it should be rather DMA framework explored than DMS. It would be also interesting to consider more indices from financial markets as oil price determinants, and omitting macroeconomics factors and supply-demand forces indicators. Such data are obtainable in higher frequencies, which allows to prepare a better data sample. Secondly, Google Trends statistics should be examined more carefully, and some other variations of Eq. 7 could be proposed.

Table 3: MSE for DMA with $\alpha = 0.97 = \lambda$

ω	
0.99	0.0041645
0.98	0.0041381
0.97	0.0041266
0.96	**0.0041249**
0.95	0.0041303
0.94	0.0041414
0.93	0.0041572
0.92	0.0041771
0.91	0.0042005
0.90	0.0042271
naive	0.0043214
futures	0.0118336

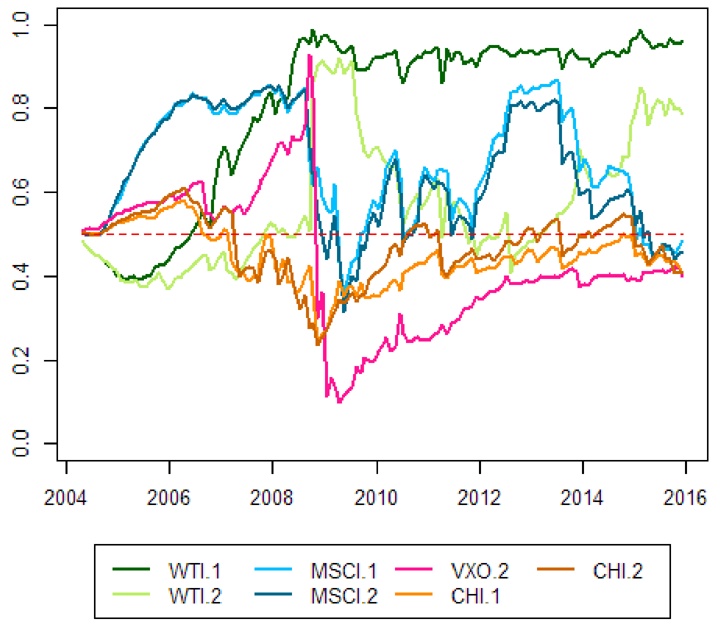

Figure 1: Posteriori probabilities for DMA with $\alpha = 0.97 = \lambda$ and $\omega = 0.96$

Acknowledgments

Research funded by the Polish National Science Centre grant under the contract number DEC-2015/19/N/HS4/00205.

Data sources

WTI, PROD, IMP, INV, CONS and NFP (U.S. Energy Information Administration)
http://www.eia.gov/dnav/pet/pet_crd_crpdn_adc_mbbl_m.htm
http://www.eia.gov/dnav/pet/pet_move_wkly_dc_NUS-Z00_mbblpd_4.htm
http://www.eia.gov/dnav/pet/pet_stoc_wstk_dcu_nus_m.htm
http://www.eia.gov/dnav/pet/pet_pri_spt_s1_m.htm
http://www.eia.gov/countries
http://www.eia.gov/dnav/pet/pet_pri_fut_s1_m.htm
MSCI (MSCI World)
http://www.msci.com/end-of-day-data-search
TB3MS and TWEXM (Federal Reserve Bank of St. Louis)

http://research.stlouisfed.org/fred2/series/TB3MS
http://research.stlouisfed.org/fred2/series/TWEXBMTH
KEI (Kilian, 2009)
http://www-personal.umich.edu/∼lkilian/paperlinks.html
VIX (Chicago Board Options Exchange)
http://www.cboe.com/micro/buywrite/monthendpricehistory.xls
CHI (Stooq)
http://stooq.com
GOOGLE SERACH
http://www.google.com/trends

References

K.A. Aastveit and H.C. Bjornland. What drives oil prices? Emerging versus developed economies. *Journal of Applied Econometrics*, 30:1013–1028, 2015.

R. Alquist, L. Kilian, and R.J. Vigfusson. *Handbook of Economic Forecasting*, chapter Forecasting the price of oil. Elsevier, 2013.

G. Aye, R. Gupta, S. Hammoudeh, and W.J. Kim. Forecasting the price of gold using dynamic model averaging. *International Review of Financial Analysis*, 41:257–266, 2015.

P.F. Bangwayo-Skeete and R.W. Skeete. Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. *Tourism Management*, 46: 454–464, 2015.

S.A. Basher, A.A. Haug, and P. Sadorsky. Oil prices, exchange rates and emerging stock markets. *Energy Economics*, 34:227–240, 2012.

C. Baumeister and G. Peersman. Time-varying effects of oil supply shocks on the US economy. *American Economic Journal: Macroeconomics*, 5:1–28, 2013.

D.G. Baur, J. Beckmann, and Czudaj R. Gold price forecasts in a dynamic model averaging framework - Have the determinants changed over time? *Ruhr Economic Papers*, 506, 2014.

N.B. Behmiri and J.R. Pires Manso. Crude oil price forecasting techniques: a comprehensive review of literature. *CAIA Alternative Investment Analyst Review*, 2:30–48, 2013.

A. Bernabe, E. Martina, J. Alvarez-Ramirez, and C. Ibarra-Valdez. A multi-model approach for describing crude oil price dynamics. *Physica A*, 338:567–584, 2004.

L. Bork and S.V. Moller. Forecasting house prices in the 50 states using dynamic model averaging and dynamic model selection. *International Journal of Forecasting*, 31:63–78, 2015.

H. Bu. Effect of inventory announcements on crude oil price volatility. *Energy Economics*, 46:485–494, 2014.

R. Carmona. *Commodities, Energy and Environmental Finance*, chapter Financialization of the commodities markets: A non-technical introduction. Springer, 2015.

H. Choi and H. Varian. Predicting initial claims for unemployment insurance using Google Trends. http://ssrn.com/abstract=1659307, 2009.

H. Choi and H. Varian. Predicting the present with Google Trends. *Economic Record*, 88: 2–9, 2012.

T. Dangl and M. Halling. Predictive regressions with time varying coefficients. *Journal of Financial Economics*, 106:157–181, 2012.

K. Dedecius, I. Nagy, and M. Karny. Parameter tracking with partial forgetting method. *International Journal of Adaptive Control and Signal Processing*, 26:1–12, 2012.

K. Drachal. Forecasting spot oil price in a dynamic model averaging framework - Have the determinants changed over time? *Energy Economics*, 60:35–46, 2016a.

K. Drachal. Determining time-varying drivers of spot oil price in a dynamic model averaging framework. Working Paper, 2016b.

L. Du and Y. He. Extreme risk spillovers between crude oil and stock markets. *Energy Economics*, 51:455–465, 2015.

D. Fantazzini and N. Fomichev. Forecasting the real price of oil using online search data. *International Journal of Computational Economics and Econometrics*, 4:4–31, 2014.

B. Fattouh, L. Kilian, and L. Mahadeva. The role of speculation in oil markets: What have we learned so far? *Energy Journal*, 34:20–30, 2013.

J.D. Hamilton. Causes and consequences of the oil shock of 2007-08. *Brookings Papers on Economic Activity*, 40:215–259, 2009.

Y. He, S. Wang, and K.K. Lai. Global economic activity and crude oil prices: A cointegration analysis. *Energy Economics*, 32:868–876, 2010.

H. Hotelling. The economics of exhaustible resources. *Journal of Political Economy*, 39: 137–175, 1931.

R.K. Kaufmann. The role of market fundamentals and speculation in recent price changes for crude oil. *Energy Policy*, 39:105–115, 2011.

L. Kilian. Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review*, 99:1053–1069, 2009.

L. Kilian and D.P. Murphy. The role of inventories and speculative trading in the global market for crude oil. *Journal of Applied Econometrics*, 29:454–478, 2014.

G. Koop and D. Korobilis. Forecasting inflation using dynamic model averaging. *International Economic Review*, 53:867–886, 2012.

G. Koop and L. Onorante. Macroeconomic nowcasting using Google probabilities. http://www.ecb.europa.eu/events/pdf/conferences/140407/OnoranteKoop_MacroeconomicNowcastingUsingGoogleProbabilities.pdf, 2014.

R. Li and G.C.K. Leung. The integration of china into the world crude oil market since 1998. *Energy Policy*, 39:5159–5166, 2011.

X. Li, J. Ma, S. Wang, and X. Zhang. How does Google search affect trader positions and crude oil prices? *Economic Modelling*, 49:162–171, 2015.

H. Naser. Estimating and forecasting the real prices of crude oil: A data rich model using a dynamic model averaging (DMA) approach. *Energy Economics*, 56:75–87, 2016.

G. Nicoletti and R. Passaro. Sometimes it helps: Evolving predictive power of spreads on GDP. *European Central Bank*, 1447, 2012.

J. Pavlicek and L. Kristoufek. Nowcasting unemployment rates with Google searches: Evidence from the Visegrad Group countries. *PLoS ONE*, 10, 2015. http://dx.doi.org/10.1371/journal.pone.0127084.

A.E. Raftery, M. Karny, and P. Ettler. Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. *Technometrics*, 52:52–66, 2010.

T. Schmidt and S. Vosen. Nowcasting private consumption: Survey-based indicators vs. Google trends. *Ruhr Economic Papers*, 155, 2009.

S. Scott and H. Varian. *Economic Analysis of the Digital Economy*, chapter Bayesian variable selection for nowcasting economic time series. National Bureau of Economic Research, 2015.

S. Stephens-Davidowitz and H. Varian. A hands-on guide to Google data. http://people.ischool.berkeley.edu/~hal/Papers/2015/primer.pdf, 2015.

L. Wu and E. Brynjolfsson. The future of prediction: How Google searches foreshadow housing prices and sales. Technical report, MIT, 2010.

C.W. Yang, M.J. Hwang, and B.N. Huang. An analysis of factors affecting price volatility of the us oil market. *Energy Economics*, 24:107–119, 2002.

A. Yousefi and T.S. Wirjanto. The empirical role of the exchange rate on the crude-oil price information. *Energy Economics*, 26:783–799, 2004.

Y.-J. Zhang and G. Wu. Does China factor matter? An econometric analysis of international crude oil prices. *Energy Policy*, 72:78–86, 2014.