**University of Pretoria**
***Department of Economics Working Paper Series***

# Oil-Price Uncertainty and the U.K. Unemployment Rate: A Forecasting Experiment with Random Forests Using 150 Years of Data

Rangan Gupta
University of Pretoria
Christian Pierdzioch
Helmut Schmidt University
Afees A. Salisu
University of Ibadan
Working Paper: 2020-95
October 2020

Department of Economics
University of Pretoria
0002, Pretoria
South Africa
Tel: +27 12 420 2413

# Oil-Price Uncertainty and the U.K. Unemployment Rate: A Forecasting Experiment with Random Forests Using 150 Years of Data

August 2020

Rangan Gupta[*], Christian Pierdzioch[†], Afees A. Salisu[‡]

**Abstract**

We analyze the predictive role of oil-price uncertainty for changes in the UK unemployment rate using more than a century of monthly data covering the period from 1859 to 2020. To this end, we use a machine-learning technique known as random forests. Random forests render it possible to model the potentially nonlinear link between oil-price uncertainty and subsequent changes in the unemployment rate in an entirely data-driven way, where it is possible to control for the impact of several other macroeconomic variables and other macroeconomic and financial uncertainties. Upon estimating random forests on rolling-estimation windows, we find evidence that oil-price uncertainty predicts out-of-sample changes in the unemployment rate, where the relative importance of oil-price uncertainty has undergone substantial swings during the history of the modern petroleum industry that started with the drilling of the first oil well at Titusville (Pennsylvania, United States) in 1859.

*JEL Classifications*: C22, C53, E24, E43, F31, G10, Q02.

*Keywords*: Machine learning; Random forests; Oil uncertainty, Macroeconomic and financial uncertainties; Unemployment rate; United Kingdom

*Conflicts of interest*: The authors declare no conflict of interest.

[*]Department of Economics, University of Pretoria, Pretoria, 0002, South Africa; Email address: rangan.gupta@up.ac.za

[†]Department of Economics, Helmut Schmidt University, Holstenhofweg 85, P.O.B. 700822, 22008 Hamburg, Germany; Email address: macroeconomics@hsu-hh.de. The research of C.P. was supported by the German Science Foundation (Project: Exploring the experience-expectation nexus in macroeconomic forecasting using computational text analysis and machine learning; Project number: 275693836).

[‡]Centre for Econometric & Allied Research, University of Ibadan, Ibadan, Nigeria. Email: adebare1@yahoo.com

# 1 Introduction

The (adverse) impact of oil-price volatility (uncertainty) on economic activity has received considerable attention after the 1973 and 1979 oil-price shocks (following the Yom Kippur war and the Iranian Revolution), and continues to be so in the wake of the outbreak of the global pandemic of COVID-19 since the beginning of 2020. Intuitively, the negative effect of oil-price uncertainty (volatility) on economic activity is generally explained by the real option theory (see for example, Bernanke 1983, Pindyck 1991, Dixit 1992, Dixit and Pindyck 1994, and more recently, Bloom 2009), which suggests that decision making is affected by (oil-price) uncertainty because it raises the option value of waiting. In other words, given that the cost associated with wrong investment decisions is very high due to irreversibility, (oil-price) uncertainty makes firms and, in the case of durable goods, also consumers more cautious. As a result, economic agents postpone investment, hiring, and consumption decisions to periods of lower (oil-price) uncertainty, which results in cyclical fluctuations in macroeconomic aggregates. Given this theoretical channel suggesting a decline in measures of real economic activity following a hike in oil-price uncertainty, a large international empirical literature has evolved trying to validate these claims (see for example, Elder and Serletis 2009, 2010, 2011, Rahman and Serletis 2010, 2011, 2012, Bredin et al. 2011, Kilian and Vigfusson 2011, Bashar et al. 2013, Pinno and Serletis 2013, Jo 2014, Elder 2018, van Eyden et al. 2019, and the references cited therein, for earlier studies on this topic). In general, these studies find evidence in favor of the negative impact of oil-price uncertainty on metrics of economic activity.

These studies are indeed insightful, but as has been pointed out by Campbell (2008), the ultimate test of any predictive model (in terms of econometric methodologies and the predictors being used) is in its out-of-sample performance. Because existence of in-sample impact (predictability) does not necessarily ensure out-of-sample forecasting gains (Rapach and Zhou 2013), our paper aims to provide a robust extension of the literature on oil-price uncertainty and its impact on economic activity by conducting an out-of-sample forecasting analysis of the predictive value of oil-price volatility for changes in the unemployment rate of the United Kingdom (UK) over the historical monthly period from 1859:10 to 2020:05. While our focus is oil-price uncertainty, to prevent omitted-variable bias, we incorporate a host of other macroeconomic and financial uncertainties (as well as the first-moments of these variables), the importance of which in forecasting macroeconomic aggregates has gained tremendous prominence in the wake of the "Great Recession" and the Global Financial Crisis (GFC) that followed thereafter (see for example, Karnizova and Li 2014, Balcilar et al., 2016, Junttila and Vataja 2018, Aye et al. 2019a, 2019b, Pierdzioch and Gupta 2020). Note that the choice of the UK as our case study is driven by the availability of a long span of data involving economic activity (as captured by the unemployment rate in our case) and a wide-array of predictors. Besides the data-availability issue, the UK is an important player in the oil market, with

a standing of the 10th and 20th as an oil importer and exporter, respectively (Central Intelligence Agency (CIA) World Factbook 2019). Hence, evidence of a predictive role of oil-price uncertainty for the development of the real economy would be a matter of valid concern for policymakers. It is important to understand that the need to look at a long historical period, which essentially covers the entire history of the modern petroleum industry that started with the drilling of the first oil well at Titusville (Pennsylvania, United States) in 1859, ensures that our empirical results do not suffer from any sample-selection bias.

As far as the econometric approach is concerned, we rely on a machine-learning approach, known as random forests (Breiman, 2001), which in turn has two main advantages. First, random forests can accurately analyze the links between changes in the UK unemployment rate and a large number of predictors in a full-fledged data-driven manner. Second, random forests automatically capture potential nonlinear links between the unemployment rate and its predictors, including uncertainties (Christou et al. 2019, Kandemir Kocaaslan 2019, Kocaarslan et al. 2020), as well as any interaction effects between the predictors. To the best of our knowledge, this is the first paper to analyze the role of oil-price uncertainty, over and above a host of other macroeconomic and financial uncertainties, in forecasting change in the unemployment rate of the UK spanning over 150 years of monthly data using a machine-learning approach.

We organize the remainder of our research as follows. In Section 2, we outline the basics of random forests. In Section 3, we briefly discuss our data. In Section 4, we present the results from our forecasting experiment. In Section 5, we conclude.

## 2 Random Forests

Random forests are a machine-learning technique that makes it possible to inspect the predictive value of oil-price uncertainty for changes in the unemployment rate in the presence of several other predictors (including, e.g., stock-market and exchange-rate uncertainty). Moreover, random forests capture in a natural data-driven way potential interaction effects between these predictors. In addition, random forests capture in a data-driven way any nonlinearities present in the data. Accounting for such nonlinearities is potentially important because economic theory suggests that the effect of uncertainty on economic variables is likely to be nonlinear (that is, the effect of high uncertainty is likely to differ from the effect of low or moderate uncertainty). In addition, an intuitive algorithm governs the computation of random forests, which clearly is an advantage given that random forests have not yet been extensively studied by the economics profession.

A random forest consists of an ensemble of individual regression trees. We, therefore, start with a description of how a regression tree, $T$, is computed. The core idea underlying the computation of a regression tree is that the branches of the tree partition the space of predictors, $\mathbf{x} = (x_1, x_2, ...)$, into

$l$ non-overlapping regions, $R_l$ (for an in-depth introduction, see Hastie et al. 2009; we largely use their notation). The branches and the resulting regions are formed in a top-down way starting at the root of the tree and then moving to its leaves by means of a recursive search-and-split algorithm. The easiest way to describe how this algorithm works is to start at the root of a regression tree.

At the root of a tree, the search-and-split algorithm loops over the array of predictors. For every predictor, $s$, the algorithm then loops over its realizations. Every realization of a predictor is a candidate for a splitting point, $p$. For every combination of a predictor and a splitting point, the search-and-split algorithm forms two half-planes, $R_1(s,p) = \{x_s | x_s \leq p\}$ and $R_2(s,p) = \{x_s | x_s > p\}$. The algorithm identifies the optimal half-planes and, thereby, the optimal combination of a predictor and a splitting point, by minimizing the following squared-error loss criterion:

$$\min_{s,p} \left\{ \min_{\bar{f}_1} \sum_{x_s \in R_1(s,p)} (f_i - \bar{f}_1)^2 + \min_{\bar{f}_2} \sum_{x_s \in R_2(s,p)} (f_i - \bar{f}_2)^2 \right\}, \tag{1}$$

where the index $i$ identifies those data of the predictand, $f$ (in the context of our analysis, the change in the unemployment rate at some forecast horizon), that belong to a half-plane, and $\bar{f}_k = \text{mean}\{f_i \,|x_s \in R_k(s,p)\}, k = 1, 2$ denotes the half-plane-specific mean of the predictand (for notational convenience, we do not use a time index).

The intuition behind Equation (1) is straightforward: The outer minimization loops over all pairs of $s$ and $p$, while the inner minimization identifies, for a given pair of $s$ and $p$, the half-plane-specific means of the predictand by minimizing the half-plane-specific squared error loss. At this point, the simple regression tree basically resembles a conventional least-squares regression model with one predictor variable, where two dummy variables capture whether the predictor variable takes on a value above or below the splitting point.

When one repeatedly applies the search-and-split algorithm to grow a more complex regression tree, this amounts to replacing the two dummy variables with several additional dummy variables that partition the predictor space into finer and finer intervals. For example, the solution of the minimization problem given in Equation (1) gives the first optimal splitting predictor, the first optimal splitting point, and the two region-specific means of the predictand. One then solves again the minimization problem given in Equation (1), but now separately for the two optimal top-level half-planes, $R_1(s,p)$ and $R_2(s,p)$, resulting in up to two second-level optimal splitting predictors and their optimal splitting points, and four second-level region-specific means of the predictand. Further application of the search-and-split algorithm then renders it possible to form a regression tree whose hierarchical structure becomes increasingly complex.

Once the search-partition algorithm stops, the data on the predictors are sent from the root to the terminal nodes down a regression tree along its nodes and branches, and a forecast of the

predictand can then be computed by means of the region-specific means at the terminal nodes of a regression tree. When the regression tree has $L$ regions, a forecast is computed as follows (1 denotes the indicator function):

$$T\left(\mathbf{x}_i, \{R_l\}_1^L\right) = \sum_{l=1}^{L} \bar{f}_l \mathbf{1}(\mathbf{x}_i \in R_l). \tag{2}$$

While repeated application of the search-and-split algorithm results in fine granular forecasts of the predictand, growing an increasingly complex regression tree does not always improve forecast accuracy. Again, the intuition is straightforward when one consider the analogy to a least-squares regression model that features several dummy variables to partition the predictor space into various subintervals. No empirical researcher would use such a complex model for forecasting purposes because it obviously suffers from a severe overfitting problem. In a similar vein, the increasing complexity of the hierarchical structure of a large regression tree naturally results in an overfitting and data-sensitivity problem.

Random forests provide a modeling platform to overcome this data-sensitivity problem. To this end, random forests combine a large number of (random) regression trees. A random forest is grown in two steps: 1) A large umber of bootstrap samples is generated from the data. Sampling is done with replacement. 2) A random regression tree is estimated on every bootstrap sample. A random regression tree selects for every splitting decision a randomly chosen subset of the predictors. This random sampling of predictors curbs the influence of influential predictors on tree growing. Growing a large number of random regression trees using this two-step recipe decorrelates the forecasts obtained from the individual (random) regression trees that form a random forest, while averaging the forecasts across the random regression trees stabilizes the forecasts computed by means of a random forest.

# 3 Data

Our raw data set covers the monthly period from 1859:09 to 2020:05, with the start date driven by the West Texas Intermediate (WTI) oil price, and the end date governed by the unemployment rate at the time of writing of this paper. Data on the unemployment rate, the consumer price index (CPI), the pound-dollar exchange rate, the short-term monetary policy rate, the long-term government bond yield (used to compute the term-spread as the difference between the long-term and short-term interest rates), and a dummy defining recession dates until 2016 are derived from the "A Millennium of Macroeconomic Data", which is maintained by the Bank of England,[1] and updated until the recent dates using the Main Economic Indicators (MEI) database of the Organisation for

[1]The data is available at: https://www.bankofengland.co.uk/statistics/research-datasets.

Economic Co-operation and Development (OECD).[2] It must be pointed out that, in order to account for the zero lower bound (ZLB) situation and unconventional monetary policy decisions during the GFC, European sovereign debt crisis, and COVID-19 outbreak, we use the Shadow Short Rate (SSR) of Wu and Xia (2016),[3], derived by modeling the term structure of the yield curve, to capture the monetary policy stance during these episodes, instead of the Bank Rate. WTI oil price and the All Share Stock Index (ALSI) are derived from the Global Financial data.[4] The WTI data is in US dollars and is converted to British pound by dividing with the dollar-pound nominal exchange rate. To ensure stationarity, the unemployment rate and the short-term interest rate are first-differenced, while we use the first-difference of the natural logarithmic values of the CPI, the ALSI, the exchange rate, and the WTI oil price in local currency. Finally, following Sadorsky (1999), and the extant literature on oil-price uncertainty, as well as macroeconomic and financial uncertainty, we fit a Generalized Autoregressive Conditional Heteroskedasticity (GARCH(1,1)) model to obtain the conditional volatilities of the first-difference of the unemployment rate and interest rate, inflation rate (i.e., the growth rate of the CPI), and (log-returns of) the ALSI, exchange rate and WTI oil price. Given the data transformation undertaken for the sake of mean-reversion, our effective sample covers the period 1859:10 to 2020:05. The dependent variable is the change in the unemployment rate, which we aim to forecast, while the (13) predictors are the recession dummy, inflation rate, change in the monetary policy rate, term-spread, stock returns, oil returns, exchange-rate returns, and the conditional volatilities of changes in the unemployment rate and the interest rate, inflation rate, oil, stock, and exchange-rate returns.

# 4 Empirical Results

Given the length of our sample period, we account for a potentially time-varying predictive value of oil-price uncertainty for changes in the UK unemployment rate by estimating random forests on rolling-estimation windows of length 120, 240, and 480 months. We forecast changes in the UK unemployment rate at four forecast horizons ($h$): 1 month, 3 months, 6 months, and 12 months, where we forecast the sum of the change in the unemployment rate (that is, the accumulated changes) when we study the latter three forecast horizons. We use the R language for statistical computing (R Core Team 2019) to carry out our forecasting experiments, where we estimate random forests by means of the add-on package "grf" (Tibshirani et al. 2020). When we move the rolling-estimation windows across our data, we use cross validation to optimize the number of predictors randomly selected for splitting, the minimum node size of a tree, and the parameter that governs

[2] https://www.oecd.org/sdd/oecdmaineconomicindicatorsmei.htm.
[3] The data is available for download from: https://sites.google.com/view/jingcynthiawu/shadow-rates?authuser=0.
[4] http://www.globalfinancialdata.com/.

the maximum imbalance of a node.[5] We use 2,000 random regression trees to grow a random forest.

In order to set the stage for our analysis, we report baseline forecasting statistics in Table 1. Specifically, we report in Panel A the ratio of the root-mean-squared forecasting error (RMSFE) of a model that neglects oil-price uncertainty and the RMSFE of a model that includes oil-price uncertainty in the array of predictors. A RMSFE ratio that exceeds unity, thus, indicates that oil-price uncertainty improves forecast accuracy under squared-error (L2) loss. In Panel B, we replace the L2 loss function with an L1 loss function and report the ratio of the mean-absolute forecast errors (MAFE) implied by the models with/without oil-price uncertainty in the array of predictors. Again, a value of the MAFE ratio above unity shows that oil-price uncertainty helps to improve forecast accuracy.

– Please include Table 1 about here. –

While the predictive value of oil-price uncertainty at the two shorter forecast horizons (that is, $h = 1, 3$) depends on the length of the rolling-estimation window, the RMSFE and MAFE ratios are consistently larger than unity for the two longer forecast horizons ($h = 6, 12$). In fact, the RMSFE and MAFE ratios attain their maximum in four out of the six considered cases for $h = 12$. Hence, the RMSFE and MAFE ratios provide evidence that oil-price uncertainty has predictive value for subsequent changes in the unemployment rate, especially when we consider the longer forecast horizons.

– Please include Table 2 about here. –

The results of Diebold and Mariano (1995) tests corroborate the predictive value of oil-price uncertainty at the longer forecast horizon. We report the test results (p-values) under the L1 and L2 loss functions for the well-known modified Diebold-Mariano test proposed by Harvey et al. (1997) in Table 2.[6] The significant test results (assuming a marginal significance level of 10%) mainly can be observed for the two long forecast horizons, $h$ = 6 and $h$ = 12.

– Please include Table 3 about here. –

In Table 3, we summarize the results for the test proposed by Clark and West (2007). Four out of six test results are significant for the two longer forecast horizons, where we use robust standard errors to assess the significance of the test results. In addition, we find three significant test results for the two shorter forecast horizons. These three significant test results have significant counterparts in Table 2, albeit the level of significance is weaker for the Diebold-Mariano test than for the Clark-West test and depends on the assumed loss function.

[5] The "grf" package also allows different subsamples to be used for constructing a tree and for making predictions. We deactivate this option, as in a classic random forest.

[6] We report p-values computed using the R package "forecast" (Hyndman 2017, Hyndman and Khandakar 2008).

– Please include Figure 1 about here. –

Figure 1 plots the relative importance of both the oil price and oil-price uncertainty over time, where we focus on the long rolling-estimation window (480 months) and the long forecast horizon ($h = 12$). Relative importance is defined as the weighted sum of how often a predictor is used for splitting. Oil-price uncertainty apparently was relatively more important than the oil price itself for most of the sample period. Moreover, our results show that, from a historical perspective, oil-price uncertainty was relatively important before 1900 and then again during two periods of time after the World War II. The first roughly covers the 1970s while the second period covers the 1980s, that is, in the periods of time during which the first and second oil-price shocks hit Western industrialized countries.

# 5 Conclusion

We have used a long-span macroeconomic and financial data set and random forests to shed light on the predictive value of oil-price uncertainty for changes in the UK unemployment rate at various forecast horizons. Random forests have several advantages: They are a data-driven approach, they account in a natural way for a potential nonlinear predictive link between oil-price uncertainty and changes in the unemployment rate, and the make it possible to control, in a unified framework, for the impact of several other macroeconomic and financial data. Our results show that oil-price uncertainty does have predictive value for movements in the unemployment rate, especially at the longer forecast horizons of six and twelve months. We also have documented the time-varying relative importance of oil-price uncertainty for changes in the unemployment rate.

In terms of future research, it would be interesting to apply the method we have used in this research to study uncertainty of oil and other commodity prices, and their impact on other developed and emerging market economies, but of course this would not involve the entire history of the petroleum industry, but a much shorter post World War II sample of data available for these economies and commodity markets. From a policy perspective, our empirical results highlight that policymakers in general, and central banks in particular, need to monitor oil-price uncertainty, over and above macroeconomic and financial uncertainties, when forecasting developments in the real economy while making their policy decisions.

## References


Aye, G.C., Christou, C., Gil-Alana, L.A., and Gupta, R. (2019a). Forecasting the probability of recessions in South Africa: the role of decomposed term spread and economic policy uncertainty. Journal of International Development, 31: 101–116.

Aye, G.C., Gupta, R., Lau, C.K.M., and Sheng X. (2019b). Is there a role for uncertainty in forecasting output growth in OECD countries? evidence from a time-varying parameter-panel vector autoregressive model. Applied Economics, 51: 3624–3631.

Balcilar M, Gupta R, Segnon M. 2016. The role of economic policy uncertainty in predicting US recessions: A mixed-frequency Markov-switching vector autoregressive approach. Economics: The Open-Access, Open-Assessment E-Journal, 10: 1–20.

Bashar, O.H., Wadud, I.M., and Ahmed, H.J.A. (2013). Oil price uncertainty, monetary policy and the macroeconomy: The Canadian perspective. Economic Modelling, 35: 249–259.

Bernanke, B. (1983). Irreversibility, uncertainty, and cyclical investment. Quarterly Journal of Economics, 98: 85–106.

Bloom, N. (2009). The impact of uncertainty shocks. Econometrica, 77(3): 623–685.

Bredin, D., Elder, J., and Fountas, S. (2011). Oil volatility and the option value of waiting: an analysis of the G-7. Journal of Futures Markets, 31: 679–702.

Breiman, L. (2001). Random forests. Machine Learning, 45: 5–32.

Campbell, J.Y., (2008) Viewpoint: estimating the equity premium, Canadian Journal of Economics, 41: 1–21.

Clark, T.D., and West, K.D. (2007). Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics, 138: 291–311.

Christou, C., Gabauer, D., and Gupta, R. (2019). Time-Varying impact of uncertainty shocks on macroeconomic variables of the united kingdom: Evidence from over 150 years of monthly data. Finance Research letters. DOI: https://doi.org/10.1016/j.frl.2019.101363.

Central Intelligence Agency (CIA) World Factbook (2019). URL: https://www.cia.gov/library/publications/the-world-factbook/.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. Journal of Business and Economic Statistic, 13: 253–263.

Dixit, A. (1992). Investment and hysteresis. The Journal of Economic Perspectives, 6(1): 107–132.

Dixit, A.K., and Pindyck R.S. (1994). Investment Under Uncertainty, Princeton, NJ: Princeton University Press.

Elder, J. (2018). Oil price volatility: Industrial production and special aggregates. Macroeconomic Dynamics, 22(S3): 640–653.

Elder, J., and Serletis, A. (2009). Oil price uncertainty in Canada. Energy Economics, 31: 852–856.

Elder, J., and Serletis, A. (2010). Oil price uncertainty. Journal of Money, Credit and Banking, 42: 1137–1159.

Elder, J., and Serletis, A. (2011). Volatility in oil prices and manufacturing activity: an investigation of real options. Macroeconomic Dynamics, 15: 379–395.

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. International Journal of forecasting, 13: 281–291.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). New York, NY: Springer.

Hyndman, R.J. (2017). forecast: Forecasting functions for time series and linear models. R package version 8.0, URL: http://github.com/robjhyndman/forecast.

Hyndman, R.J., and Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. Journal of Statistical Software, 26: 1–22.

Jo, S. (2014). The effects of oil price uncertainty on global real economic activity. Journal of Money, Credit and Banking, 46(6): 1113–1135.

Junttila J., and Vataja, J. (2018). Economic policy uncertainty effects for forecasting future real economic activity. Economic Systems, 42: 569–583.

Karnizova L., and Li J.C. (2014). Economic policy uncertainty, financial markets and probability of US recessions. Economics Letters, 125: 261–265.

Kandemir Kocaaslan, O. (2019). Oil price uncertainty and unemployment. Energy Economics, 81: 577–583.

Kilian, L., and Vigfusson, R.J. (2011). Nonlinearities in the oil price-output relationship. Macroeconomic Dynamics, 15: 337–363.

Kocaarslan, B., Soytas, M.A., and Soytas, U. (2020). The asymmetric impact of oil prices, interest rates and oil price uncertainty on unemployment in US. Energy Economics, 86: 104625.

Pierdzioch, C.; Gupta, R. Uncertainty and Forecasts of U.S. Recessions. Studies in Nonlinear Dynamics & Econometrics 2020, 24(4): 1–20.

Pindyck, R.S. (1991). Irreversibility, Uncertainty, and Investment. Journal of Economic Literature, 24: 1110–1148.

Pinno, K., and Serletis, A. (2013). Oil Price Uncertainty and Industrial Production. The Energy Journal, 34: 191–216.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Rahman, S., and Serletis, A. (2010). The asymmetric effects of oil price and monetary policy shocks: a nonlinear VAR approach. Energy Economics, 32: 1460–1466.

Rahman, S., and Serletis, A. (2011). The asymmetric effects of oil price shocks. Macroeconomic Dynamics, 15: 437–471.

Rahman, S., and Serletis, A. (2012). Oil price uncertainty and the Canadian economy: evidence from a VARMA, GARCH-in-Mean, asymmetric BEKK model. Energy Economics, 34: 603–610.

Rapach, D., and Zhou, G. (2013). Forecasting Stock Returns. Handbook of Economic Forecasting, Volume 2A, Graham Elliott and Allan Timmermann (Eds.) Amsterdam: Elsevier: 328–383.

Sadorsky, P. (1999). Oil price shocks and stock market activity. Energy Economics, 2: 449-469.

Tibshirani,J., Athey, S., and Wager, S. (2020). grf: Generalized Random Forests. R package version 1.1.0. https://CRAN.R-project.org/package=grf.

Wu, J.C., and Xia, F.D. (2016). Measuring the macroeconomic impact of monetary policy at the zero lower bound. Journal of Money, Credit and Banking, 48 (2-3): 253–291.

van Eyden R., Difeto, M., Gupta, R., and Wohar, M.E. (2019). Oil price volatility and economic growth: Evidence from advanced economies using more than a century's data. Applied Energy, 233: 612-621.

Table 1: Baseline Forecasting Statistics

Panel A: RMSFE ratios

| Window | $h = 1$ | $h = 3$ | $h = 6$ | $h = 12$ |
|---|---|---|---|---|
| 120 | 1.0023 | 1.0086 | 1.0230 | 1.0189 |
| 240 | 1.0104 | 1.0155 | 1.0028 | 1.0112 |
| 480 | 0.9903 | 0.9939 | 1.0210 | 1.0231 |

Panel B: MAFE ratios

| Window | $h = 1$ | $h = 3$ | $h = 6$ | $h = 12$ |
|---|---|---|---|---|
| 120 | 0.9945 | 1.0126 | 1.0108 | 1.0181 |
| 240 | 1.0110 | 1.0085 | 1.0020 | 1.0187 |
| 480 | 0.9920 | 0.9931 | 1.0132 | 1.0128 |

Note: This table reports results of RMSFE and the MAFE ratio obtained by dividing the RMSFE of a model that does not feature oil-price uncertainty in the array of predictors by the RMSFE of a model that can use oil-price uncertainty to build random forests. The parameter $h$ denotes the forecast horizon (in months).

Table 2: Diebold-Mariano Test Results

Panel A: L1 loss

| Window | $h = 1$ | $h = 3$ | $h = 6$ | $h = 12$ |
|---|---|---|---|---|
| 120 | 0.7649 | 0.0720 | 0.1090 | 0.0062 |
| 240 | 0.0591 | 0.1343 | 0.3787 | 0.0427 |
| 480 | 0.8248 | 0.8486 | 0.0305 | 0.1251 |

Panel B: L2 loss

| Window | $h = 1$ | $h = 3$ | $h = 6$ | $h = 12$ |
|---|---|---|---|---|
| 120 | 0.4471 | 0.1279 | 0.0509 | 0.0009 |
| 240 | 0.1253 | 0.0923 | 0.4134 | 0.1055 |
| 480 | 0.9183 | 0.8562 | 0.0395 | 0.0989 |

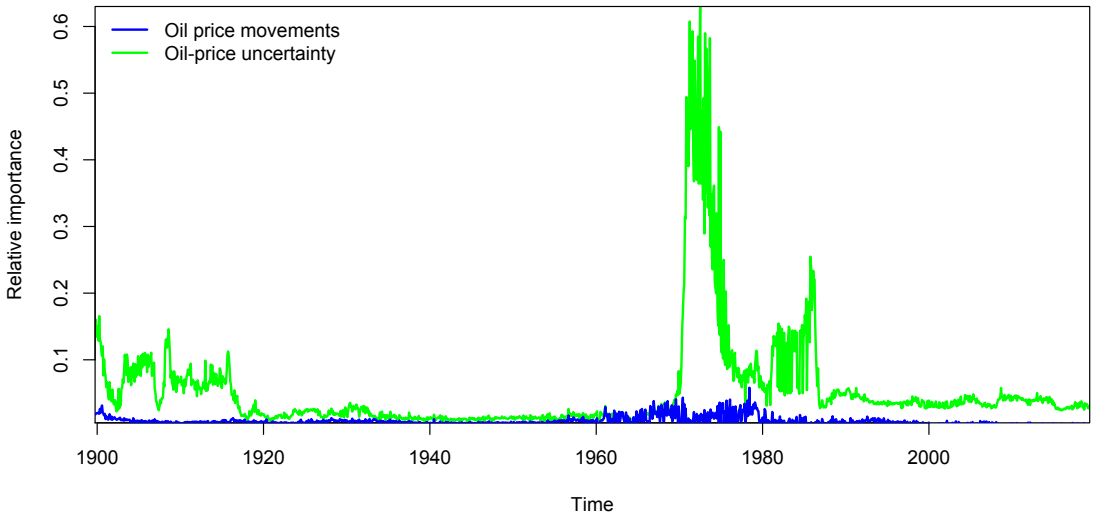
Note: This table reports results (p-values) of Diebold-Mariano tests under L1 (loss depends on the absolute forecast error) and L2 (loss depends on the squared forecast error) loss. The null hypothesis is that the accuracy of forecasts extracted from of a model that does not feature oil-price uncertainty in the array of predictors is equal to the accuracy of forecast computed by means a model that can use oil-price uncertainty to build random forests. The alternative hypothesis is that the latter forecasts are more accurate than the former. The parameter $h$ denotes the forecast horizon (in months).

Table 3: Clark-West Test Results

| Window | $h = 1$ | $h = 3$ | $h = 6$ | $h = 12$ |
|---|---|---|---|---|
| 120 | 0.1866 | 0.0033 | 0.0221 | 0.0001 |
| 240 | 0.0282 | 0.0213 | 0.1109 | 0.0059 |
| 480 | 0.4444 | 0.3964 | 0.0260 | 0.0348 |

Note: This table reports results (p-values) of Clark-West tests loss. t-statistics of the Clark-West tests are based on Newey-West standard errors. The parameter $h$ denotes the forecast horizon (in months). The parameter $h$ denotes the forecast horizon (in months).

Figure 1: Relative Importance of Oil-Related Predictors




Window length 480 months. Predictor importance is computed for $h = 12$. Predictor importance is defined as the weighted sum of how often a predictor is used for splitting.