The Journal of Economic Asymmetries 21 (2020) e00154

Contents lists available at ScienceDirect

The Journal of Economic Asymmetries

journal homepage: www.elsevier.com/locate/jeca

Crude oil price changes and the United Kingdom real gross domestic product growth rate: An out-of-sample investigation[☆]

Nima Nonejad[a,b,*]

[a] Danske Bank, Denmark
[b] CREATES, Denmark

ARTICLE INFO

JEL classification:
C22
C53
E32

Keywords:
Crude oil price
Macroeconomy forecasting
Structural change
Real GDP growth rate

ABSTRACT

We evaluate the impact of changes in the price of crude oil on the United Kingdom (U.K.) real gross domestic product (GDP) growth rate by way of an out-of-sample forecasting analysis. We compare the performance of several nonlinear models and determine, which aspects of nonlinearities are most useful for obtaining forecast improvements. Likewise, our approach takes into account the possibility that relative predictive performance can vary over the out-of-sample period. Results based on quarterly data from 1974q1 through 2018q4 illustrate that our conclusions depend on the definition of forecast improvement and whether we rely on pairwise or multiple forecast comparison. For instance, it is very difficult to find evidence that point forecasts exploiting crude oil price variables are statistically significant more accurate than point forecasts produced under the benchmark. On the other hand, the null hypothesis of no population-level predictability is borderline rejected for certain nonlinear crude oil price variables. We also observe notable differences between using real-time and ex-post revised GDP data with regards to local out-of-sample performance. The predictive power associated with the more successful crude oil price measures appears to concentrate in the early 1990s and around the onset of the Great Recession.

1. Introduction

The aim of this study is to evaluate the predictive impact of changes in the price of crude oil on the United Kingdom (U.K.) real gross domestic product (GDP) growth rate via an out-of-sample analysis. We are motivated to do so based on several interesting results reported in widely cited studies, such as Bachmeier, Li, and Liu (2008), Carlton (2010), Kilian and Vigfusson (2013), Ravazzolo and Rothman (2013) and Ravazzolo and Rothman (2016) .[1] The central question running through these studies is whether changes in the price of crude oil have a predictive impact on the real U.S. GDP growth rate out-of-sample.[2] However, in the context of other countries

☆ The views expressed in this paper are our own and do not in any way reflect those of Danske Bank.

* Dalbergstrøget 17, Høje Taastrup, 2630, Taastrup, Denmark
E-mail address: nimanonejad@gmail.com.

[1] Bachmeier et al. (2008) find that the price of crude oil does not have a predictive impact on the real U.S GDP growth rate. However, Ravazzolo and Rothman (2013) find statistical evidence of population-level predictability at long forecast horizons.

[2] The initial studies on U.S. data document that crude oil price increases are relevant when they exceed the maximum price over the last three years. The recent study by Kilian and Vigfusson (2013) finds that preserving symmetry between price increases and decreases over the last three years further improves out-of-sample GDP growth rate forecasts relative to the benchmark. We can also refer the reader to interesting studies, such as Kilian and Vigfusson (2011a), Serletis and Istiak (2013) and Romero-Meza, Coronado, and Serletis (2014) for more details on the nature of nonlinearities between the price of crude oil and the level of economic activity.

https://doi.org/10.1016/j.jeca.2020.e00154
Received 3 December 2019; Received in revised form 7 January 2020; Accepted 15 January 2020

relatively less is known about whether (a): Changes in the price of crude oil help forecast the country specific real GDP growth rate, (b): If the predictive impact of crude oil price changes on the real GDP growth rate is nonlinear and (c): If so what types of nonlinearities. In fact, in their seminal study, Kilian and Vigfusson (2013) argue that understanding the nonlinear/asymmetric predictive impact of changes in the price of crude oil on the real GDP growth rate deserves further study on non-U.S. data.

Arguably, the U.K. can be considered as an interesting case because it is a developed, open economy, which has transitioned from being a net crude oil importer in the 1970s to a net exporter in the 1980s and early 1990s before returning to being a net importer again since the mid 2000s.[3] This in turn, provides a very good opportunity to evaluate (a), (b) and (c). Likewise, studies such as Jiménez-Rodríguez and Sánchez (2006), Harrison, Ryland, and de Weymarn (2011), Millard (2011) and Millard and Shakir (2013) focus on the causes and effects of crude oil price changes on the U.K. economy via examination of impulse response functions. However, they do not account for nonlinearities in the relationship between the price of crude oil and the growth rate. Similar to simulations provided by the National Institute for Economic and Social Research (NIESR) and the Bank of England they conclude that a fall in the price of crude oil leads to an increase in the U.K. GDP growth rate. To date, a comprehensive analysis on the predictive impact of crude oil price changes on the U.K. GDP real growth rate has not been conducted. The purpose of this study is to fill this gap. Finally, the increasing financialization of the crude oil market and the important role of crude oil in the context of a developed, open economy, such as the U.K. makes our analysis interesting to a broad audience.[4]

Following well-known studies, such as Kilian and Vigfusson (2013), Ravazzolo and Rothman (2013) and Ravazzolo and Rothman (2016), besides the percentage change in the real (nominal) price of crude oil, we also consider a number of nonlinear crude oil price variables in our analysis. These include well-known measures, such as the three-year net crude oil price increase as well as more recent measures, such as the three-year crude oil price gap and large crude oil price change (increase). This in turn, enables us to shed light on which aspects of nonlinearities are most useful for achieving forecast improvements. Following the mentioned studies, we start by constructing the crude oil price measure of interest. Then, lagged values of the measure enter into the model as regressors to have their predictive performance evaluated.

We distinguish between two different definitions of forecast improvement, namely, population-level predictability and finite-sample forecast accuracy. The former evaluates whether at the population value of the model parameters, forecasts produced under the predictive model and the benchmark are equally accurate. The latter tests whether employing the crude oil price-based predictor of interest improves the accuracy of point forecasts relative to the benchmark given the data at hand. Although these notions sound similar they are in fact different. For example, suppose that the real GDP growth rate depends on lagged changes in the price of crude oil, such that there is evidence of population-level predictability. However, out-of-sample point forecasts employing lagged values of this variable could nevertheless be less accurate than point forecasts produced under the (misspecified) because there is a bias-variance trade-off at play.[5]

To evaluate the predictive performance of models augmented with the suggested crude oil price variables relative to the benchmark, we rely on two widely applied tests, namely, Diebold and Mariano (1995) and Clark and West (2007). Both tests are based on comparing the mean square error (MSE) produced under the predictive model of interest with the MSE produced under the benchmark. However, contrary to Diebold and Mariano (1995), the Clark and West (2007) test adds an adjustment term to the MSE difference that accounts for parameter estimation noise. The Clark and West (2007) test is used for determining the evidence of population-level predictability, whereas the Diebold and Mariano (1995) test is designed to examine whether point forecasts produced under one specification are statistically significant more accurate than another.

We start by applying the mentioned tests to select specifications that forecast best on average over the out-of-sample period, which we henceforth refer to as global predictive evaluation. Accordingly, our framework outlined above relies on comparing forecasts produced under the alternative models with forecasts produced under the benchmark one by one, i.e., pairwise comparison. However, given data snooping concerns raised in a seminal study by White (2000), we also perform multiple forecast comparison tests. We use the original framework suggested in White (2000) as well as a more recent method, namely, the model confidence set (MCS) suggested in Hansen, Lunde, and Nason (2011). The null hypothesis of White (2000) is that that all alternative forecasts are as good as forecasts produced under the benchmark. The objective of the MCS procedure is to determine the set of "best'' models that forecast better that other models in the set. We can view MCS as the set of models that includes the best models with a given level of confidence.

Recently, Giacomini and Rossi (2010) argue that solely relying on global predictive evaluation can provide misleading results. This is because relative predictive performance can vary substantially over the out-of-sample period and averaging such movements can result in a loss of information. For example, the practitioner may ignore the fact that the specification with the higher MSE computed over the entire out-of-sample period actually produces more accurate forecasts when considering recent data. Therefore, to bring completeness to our empirical analysis, we also apply the local forecast evaluation framework suggested by the authors, which allows us to explore how

[3] According to the Department of Energy & Climate Change (DECC), the U.K. is the largest producer of crude oil and second largest producer of natural gas in the EU. Production from U.K. oil and natural gas fields in the North Sea peaked around the late 1990s and has declined steadily since then.

[4] Financialization refers to the increasing presence of financial investors in the crude oil markets, see Gorton and Rouwenhorst (2006), Zhang, Fan, Tsai, and Wei (2008) an Reboredo et al. (2013) among others. Recent studies, such as Tang and Xiong (2012), Cheng and Xiong (2014), Kolodziej, Kaufmann, Kulatilaka, Bicchetti, and Maystre (2014) and Kartsakli and Adams (2017) argue that as a result of the financialization, variations in the price of crude oil are increasingly being influenced by what occurs in equity markets rather than demand/supply factors.

[5] With bias-variance trade-off, we refer to the fact that the additional forecast accuracy afforded by including lagged values of the variable of interest in the model (even when evidence of population-level predictability is strong) might not offset increased forecast variance related to estimation of the additional model parameters.

relative predictive performance changes over the out-of-sample.

Briefly foreshadowing the main results our findings are as follows: To begin with, our findings survive a number of robustness checks, namely, examining for omitted variable, using real-time instead of ex-post revised GDP data and relying on the real as well as the nominal price of crude oil. In the context of pairwise comparison, for both ex-post revised and real-time GDP data, the null hypothesis of no population-level predictability is rejected for models employing certain nonlinear crude oil price variables at forecast horizons beyond one quarter. Among them, the three-year net crude oil price change and the three-year crude oil price gap are the top performers. Results are also positive for the three-year net crude oil price increase in some instances. Conversely, nonlinearities embodied in the three-year crude oil price decrease and large crude oil price change cause notable deterioration in forecast performance relative to the benchmark. It must also be mentioned that (somewhat similar to results using U.S. data) even for the most successful crude oil price measures, the null hypothesis of no population-level predictability is mainly rejected at the 10% significance level and less so at the 5% level.[6] Furthermore, except for rare instances, none of our forecasts outperform forecasts produced under the benchmark at the 1% significance level. Evidence of population-level predictability weakens further when focusing on multiple forecast comparison tests. Except for four and five-quarters ahead at the 10% level, the benchmark consistently belongs to 10% and 10% MCS.

The evidence of superior predictive ability is not positive. Reductions in MSE relative to the benchmark are at best modest (around 5%) and the augmented models fail to generate statistically significant improvements in forecast accuracy relative to the benchmark across the considered forecast horizons. Therefore, to better understand the pattern of our out-of-sample results, we conduct a simple Monte Carlo experiment in a similar fashion as Paye (2012). The simulated data are calibrated to match the empirical features of the observed data as much as possible. In one case, the simulated variable predicts the dependent variable at the population level. In the second, we impose the null hypothesis of no population-level predictability. Simulation results illustrate a pattern of divergence between the Diebold and Mariano (1995) and Clark and West (2007) tests similar to results observed in our empirical application.

When evaluating global out-of-sample performance, results are qualitatively similar across the use of ex-post revised and real-time GDP data. However, when we focus on local relative out-of-sample performance, we observe that imposition of real-time data constraint does indeed lead to notable differences in out-of-sample results. Finally, time-series plots of re-scaled loss differences over the out-of-sample period display that predictive gains are short-lived and concentrate in specific periods. Furthermore, different predictors exhibit different patterns of improvements and a uniform pattern of gains cannot be detected.

The rest of this study is as follows: In Section 2, we provide a motivation for our analysis, justifying why it should not be considered as pure forecasting exercise and make the case that it is interesting for a broad audience. Section 3 presents the econometric framework. Crude oil price variables used in the analysis are presented in Section 4. Results are reported and discussed in Section 5. Finally, in Section 6, we conclude.

2. Motivation

Why would one want to quantify the predictive impact of changes in the price of crude oil on the real U.K. GDP growth rate? In the previous section, we mentioned some interesting facts about U.K.'s role as a crude oil exporter/importer over time, which is rather unique among OECD countries. Likewise, we mentioned studies that explore the impact of crude oil price changes on the U.K. economy via impulse response analyses while at the same time failing to account for nonlinearties regarding this relationship. Finally, we also mentioned that almost all studies focusing on out-of-sample predictability rely exclusively on U.S. real GDP data, which makes the case for conducting our analysis on one of the largest economies across the pond. In this section, we provide further justifications for our out-of-sample analysis.

In panel (a) of Fig. 1, we display the real GDP growth rate and the first difference of log-real price of crude oil lagged one quarter, $\triangle oil_{t-1}$. The former is extracted from the Office for National Statistics' website. For the latter, we use Brent Blend crude oil U.S. dollar spot prices extracted from the Global Financial Database deflated by the U.S. CPI index.[7] To make comparison easier, we standardize the series and "smooth'' them using yearly moving averages. It is important to note that this is done only with regards to panels (a) to (c) in Fig. 1 and nowhere else do we smooth any series. Overall, the series co-move with each other. We also observe a degree of time-variation regarding the affinity between lagged crude oil price changes and the GDP growth rate. Even more interesting, decreases in the GDP growth rate tend to be followed by rising crude oil prices. In panels (b) and (c) of Fig. 1, we look more carefully at the periods 1974q1-2000q1 and 2000q1-2018q4, respectively. Around 2007, we observe that the GDP growth rate starts decreasing after the price of crude oil reaches a high level compared to previous quarters. A similar pattern is observed in panel (b) during the mid 1980s and early 1990s. Based on these figures, it is plausible to assume that one could (notably) improve real GDP growth forecasts by conditioning on the price of crude oil.

Next, we consider a simple autoregression of order two augmented with $\triangle oil_{t-1}$. Here, we conclude that the autoregressive coefficients are statistically different than zero, whereas we fail to reject the null hypothesis that the coefficient associated with $\triangle oil_{t-1}$, namely, β_1 is zero. In panel (d) of Fig. 1, we report β_1 estimated using a rolling window approach, where the window size corresponds to fifteen years of historical data, see Section 5.1 for details on why this window length is chosen. As indicated by the 95% confidence

[6] We must also mention that certain highly cited studies, such as Kilian and Vigfusson (2013) do not conduct out-of-sample predictability tests. Instead, the authors report MSE ratios relative to the benchmark across different forecast horizons.

[7] There are two major crude oil markets: West Texas Intermediate (WTI) and Brent Blend. WTI is a blend of several U.S. domestic streams of crude oil. Brent Blend is a combination of crude oil from fifteen different oil fields in the North Sea. Similar to WTI, it is priced in U.S. dollars and is usually considered as the benchmark for the price of crude oil in Europe.

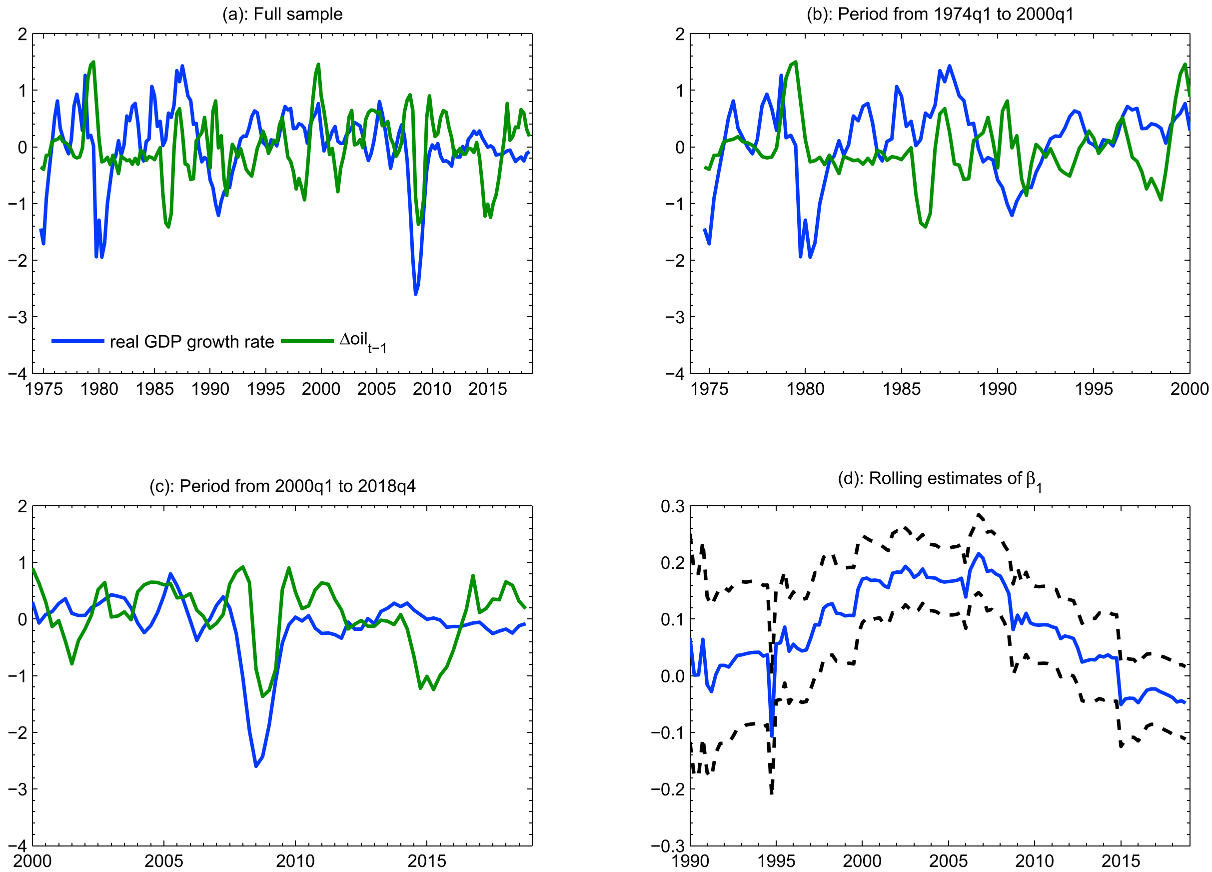

Fig. 1. Quarterly U.K. real GDP growth rate and changes in log-real price of crude oil.
In panel (d), the blue line displays rolling OLS estimates of β_1 for the model: $y_t = \varphi_0 + \varphi_1 y_{t-1} + \varphi_1 y_{t-2} + \beta_1 \Delta oil_{t-1} + \varepsilon_t$, where y_t is the ex-post revised real GDP growth rate at time t. The dotted lines display the corresponding confidence bands. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

bands, the evidence of predictability is strong in certain instances whilst weak in others.

What to make from the above analysis? Given the affinity between crude oil price changes and the GDP growth rate displayed in panel (a) of Fig. 1, the evidence of out-of-sample predictability ought to be strong. However, our very simple in-sample analysis suggests the opposite. Furthermore, several unresolved questions still need answers. First, it is well-known that in-sample evidence does need not translate into out-of-sample gains, which is the ultimate question of interest to policy makers and forecasters. Second, our analysis thus far has been based on ex-post revised GDP data. It would be very interesting to explore whether out-of-sample results change drastically if we were to rely on real-time GDP data. Third, we should also consider nonlinear crude oil price variables to evaluate how/if results change. Finally, as discussed in the previous section, the definition of forecast improvement is important as it can lead to opposite conclusions. Overall, these points justify conducting our out-of-sample analysis.

3. Econometric framework

Following Kilian and Vigfusson (2013) and Ravazzolo and Rothman (2013), we consider an autoregressive (AR) model of order p as the benchmark model, namely,

$$y_t = \varphi_0 + \sum_{i=1} \varphi_i y_{t-i} + \varepsilon_t, \quad \varepsilon_t \tilde{} N(0, \sigma^2), \tag{1}$$

where y_t denotes the real GDP growth rate at quarter t. Next, we extend the benchmark as:

$$y_t = \varphi_0 + \sum_{i=1} \varphi_i y_{t-i} + \sum_{i=1} \beta_i X_{t-i} + \varepsilon_t, \quad \varepsilon_t \tilde{} N(0, \sigma^2), \tag{2}$$

where X_{t-i} is the possibly vector-valued crude oil price measure of interest at time $t-i$ for $i=1,\ldots,q$. Throughout this study, the lag orders in (3.1) and (3.2) are chosen using the BIC criterion.[8] Specifically, following Rossi and Sekhposyan (2010), p is selected based on full-sample estimation of (3.1) across different lags. We denote the lag order associated with the lowest BIC value as p^*. Thereafter, q^* is chosen is a similar fashion, where $.p=p^*$[9] Results are similar if we were to choose p and q simultaneously.

It is worth mentioning that we can easily allow for time-variation in the regression coefficients and conditional volatility in a similar fashion as Groen, Richard, and Ravazzolo (2013) or Bauwens, Koop, Korobilis, and Rombout (2015). However, the time-varying parameter models in these studies are applied in the context of an expanding window forecasting approach (i.e., the estimation sample grows as we move through the out-of-sample period). However, we rely on a rolling window forecasting approach (i.e., the estimation sample is kept at the same size as it moves over time) because under an expanding window approach test statistics, such as Diebold and Mariano (1995) and Giacomini and Rossi (2010) have a non-standard limiting distribution, see for example, Clark and McCracken (2012). Likewise, Stock and Watson (1996) and Swanson (1998) demonstrate that a rolling window forecasting approach is able to account for parameter instability in the data generating process. In fact, point forecasts produced under (3.2) and the time-varying parameter specification suggested in Groen et al. (2013) both generated using the rolling window approach are fairly similar[10].

Finally, it is important to note that the purpose of this study is not to perform forecast comparison across specifications that allow for model uncertainty, time-variation in the model parameters, or both in one way or another. Conversely, we aim to answer an empirical question, namely, whether it is possible to improve GDP growth rate forecasts by conditioning on the price of crude oil.

4. Crude oil price variables

An obvious choice for a crude oil price-based predictor is the percentage change in the price of crude oil at quarter t defined as: $\triangle \mathit{oil}_t = \mathit{oil}_t - \mathit{oil}_{t-1}$, where oil_t is the log-crude oil price at quarter t. However, as argued in Kilian and Vigfusson (2013) among others, it is possible that this linear variable (in the sense that an increase in the price of crude oil at quarter t has the same impact on the real GDP growth rate at quarter $t+1$ as a decrease in the price of crude oil) might only lead to small forecast improvements relative to the benchmark. Therefore, we follow the authors and consider several nonlinear transformations of the price of crude oil as predictors in our analysis. It is worth mentioning that these nonlinearities are motivated on the basis of behavioral arguments rather than pure economic theory. They also provide an excellent platform to formally test, which aspects of nonlinearities are most useful for achieving forecast improvement. Kilian and Vigfusson (2011b) and Kilian and Vigfusson (2013) provide rigorous justifications for using these variables. Therefore, we briefly define them in Table 1 and refer the reader to the mentioned studies for more details. It is important to mention that similar to Hamilton (2003), Kilian and Vigfusson (2013), Ravazzolo and Rothman (2013) and many others, our nonlinear crude oil price variables are based on price movements relative to recent highs and lows. However, they are agnostic about whether these movements are due to demand or supply shocks. This is because this topic is related mainly to in-sample analyses. In the context of the U.K., it has been studies in Millard and Shakir (2013) by way of impulse response analyses.

The three-year net crude oil price increase suggested by James Hamilton in the early 1990s, namely, $\mathit{net}_t^+ = \max(0, \mathit{oil}_t - \dot{\mathit{oil}}_t)$, where $\dot{\mathit{oil}}_t = \max(\mathit{oil}_{t-1}, \ldots, \mathit{oil}_{t-12})$ denotes the highest price of crude oil over the last twelve quarters embodies two distinct types of nonlinearities. The first assumes that consumers perceive the current price of crude oil differently, depending on how much it differs from recent historical experience. This nonlinearity is symmetric, i.e., it also applies to net crude oil price decreases. The second nonlinearity arises from the assumption that consumers do not respond to crude oil price decreases, thereby allowing us to omit this component. This assumption imposes an asymmetry. The obvious counterpart to net^+ is the three-year net crude oil price decrease, $\mathit{net}_t^- = \min(0, \mathit{oil}_t - \ddot{\mathit{oil}}_t)$, where $\ddot{\mathit{oil}}_t = \min(\mathit{oil}_{t-1}, \ldots, \mathit{oil}_{t-12})$.

It is not immediately evident whether the second form of nonlinearity contained in net^+ is more important than the first. We can easily address this issue by comparing net^+ with a specification that does not involve asymmetry but accounts for net deviations, namely, the three-year net change in the price of crude oil, $\mathit{net} = \mathit{net}^+ + \mathit{net}^-$. Here, the predictive impact of the price of crude oil is symmetric in net price increases and decreases. The asymmetric net crude oil price change, which compared to net contains a weaker form of asymmetry can be defined as $\mathit{anet}_t = [\max(0, \mathit{oil}_t - \dot{\mathit{oil}}_t), \min(0, \mathit{oil}_t - \ddot{\mathit{oil}}_t)]$. Contrary to net, net crude oil price increase and net crude oil price decrease enter anet with different weights.

It is also possible that crude oil consumers might not respond to net price changes but simply the deviation of the price of oil from the highest price in recent memory. This behavioral aspect can be captured by replacing net^+ by the uncensored variable, $\mathit{gap}_t = \mathit{oil}_t - \dot{\mathit{oil}}_t$, which we henceforth refer to as the three-year crude oil price gap. As argued in Kilian and Vigfusson (2013), this variable allows us to relax the assumption that the economy only responds to the price of crude oil exceeding recent peaks. However, we still maintain the notion that there is state dependence in the feedback from the price of crude oil to the economy.

The final form of nonlinearity considered in this study focuses on large crude oil price changes. The fact that small crude oil price

[8] We choose BIC over AIC because the former penalizes model complexity more heavily. However, out-of-sample results do not change substantially if we were to use AIC as opposed to BIC.

[9] Following Rossi and Sekhposyan (2010), we use this approach to minimize the effect of lag orders across different specifications on out-of-sample performance. We also experiment with re-estimating the lag orders in each window over the out-of-sample period and observe that results are largely similar to those reported in this study.

[10] Furthermore, (3.2) using the rolling window approach outperforms the change-point specification of Bauwens et al. (2015).

Table 1
Crude oil price variables.

Label	description
$\triangle oil$	first difference of log-crude oil price, $oil_t - oil_{t-1}$, where oil_t is log-crude oil price at quarter t.
net^{+}	three-year net crude oil price increase, $\max(0, oil_t - \dot{oil}_t)$, where $\dot{oil}_t = \max(oil_{t-1}, \ldots, oil_{t-12})$.
net^{-}	three-year net crude oil price decrease, $\min(0, oil_t - \ddot{oil}_t)$, where $\ddot{oil}_t = \min(oil_{t-1}, \ldots, oil_{t-12})$.
net	three-year net crude oil price change, $net_t^{+} + net_t^{-}$.
$anet$	three-year asymmetric net crude oil price change, $[\max(0, oil_t - \dot{oil}_t), \min(0, oil_t - \ddot{oil}_t)]$.
gap	three-year crude oil price gap, $oil_t - \dot{oil}_t$, where as before $\dot{oil}_t = \max(oil_{t-1}, \ldots, oil_{t-12})$.
$large$	three-year large crude oil price change, $\triangle oil_t \cdot I(\mid \triangle oil_t \mid > \mathrm{std}(\triangle oil_t))$, where $I(\cdot)$ is an indicator function and "std'' is the
$large^{+}$	three-year large crude oil price increase, $\triangle oil_t \cdot I(\triangle oil_t > \mathrm{std}(\triangle oil_t))$. standard deviation of $\triangle oil_t$ over the last 12 quarters.

In the case of the asymmetric net price change, *anet*, the three-year net crude oil price increase and decrease, namely, net^{+} and net^{-}, constitute a 2-element vector.

changes might go unnoticed by consumers could be rationalized by the costs of constantly monitoring crude oil price changes and adjusting consumption patterns, see for example, Goldberg (1998). We define a large crude oil price change as: $\triangle oil_t \cdot I(\mid \triangle oil_t \mid > \mathrm{std}(\triangle oil_t))$, where "std" refers to the sample standard deviation of $\triangle oil$ over the last twelve quarters and $I(\cdot)$ is an indicator function equal to 1 if the inequality is satisfied, otherwise 0. An asymmetric version of this variable in which only a large crude oil price increase matters is defined as: $\triangle oil_t \cdot I(\triangle oil_t > \mathrm{std}(\triangle oil_t))$.

Before proceeding further, one important issue must be addressed, namely, whether we should rely on the nominal or real price of crude oil. Kilian and Vigfusson (2011b) argue that one should rely on the real price of crude oil, whereas Hamilton (2011) favors the nominal price of crude oil by arguing that real price induces measurement errors, which can ultimately result in deterioration in forecasting performance.[11] Likewise, Hamilton (2011) argues that the argument for using nonlinear crude oil price variables, such as the three-year net crude oil price increase is mainly behavioral. Hence, a nominal specification appears just as reasonable as a real specification. Another argument for use of the nominal price of crude oil is that the nominal price can be considered exogenous with respect to U.K. economy and hence unpredictable on the basis of lagged GDP values. In light of this debate and for the sake of completeness, we follow Kilian and Vigfusson (2013) and conduct our analysis using the real as well as the nominal price of crude oil.

5. Out-of-sample analysis

We rely on seasonally adjusted quarterly data from 1974q1 through 2018q4. The choice of the start date is standard in the literature and is primarily motivated by the discussion in Alquist, Kilian, and Vigfusson (2013). The authors advise against using crude oil price data before 1974 because the price of crude oil varies very little and tends to exhibit a pattern resembling a step-function. This property also implies problems for the real price of crude oil. Finally, they argue that it is inappropriate to combine data before 1973 with data after 1973.

Following Ravazzolo and Rothman (2013) and Ravazzolo and Rothman (2016), we rely on ex-post revised as well as real-time GDP data. Both are extracted from the Office for National Statistics' website. As a measure for the price of crude oil, we follow Jiménez-Rodríguez and Sánchez (2006), Millard and Shakir (2013), Akram and Mumtaz (2019) and use Brent Blend crude oil spot prices extracted from the Global Financial Database. The Brent Blend price of crude is of course not subject to revisions. When necessary, we follow Akram and Mumtaz (2019) and deflate the price of crude oil by the U.S. CPI index.[12]

5.1. *Design specifics*

The data is divided into an in-sample portion composed of the first r observations and an out-of-sample period based on the final n observations. The forecasts are produced using a rolling window estimation approach: The one-quarter ahead forecast for time $t+1$ is generated conditional on data from time 1 to t. The estimation window is then rolled forward one quarter and forecast for $t+2$ is produced using data from time 2 to $t+1$ and so on. The window size corresponds to fifteen years of data.[13] To produce forecasts at horizons greater than one, $h > 1$, we use the direct method of forecasting. This way, we avoid using multiequation systems. Following the literature, the out-of-sample runs from 1990q1 through 2018q4.

[11] Kilian and Vigfusson (2011b) state "The focus on real oil price innovations … makes sense because theoretical models that imply asymmetries in the transmission of oil price shocks are expressed in terms of real price of oil." On the other hand, Hamilton (2011) states "… deflating by a particular number, such as the CPI, introduces a new source of measurement error, which could lead to a deterioration in the forecasting performance. In any case, it is again quite possible that there are differences in the functional form of forecasts based on nominal instead of real prices.".

[12] As an experiment, we convert the U.S. dollar based Brent Blend price of crude oil to Pound Sterling and then deflate the corresponding series by the U.K. CPI. However, out-of-sample results are qualitatively similar to those reported in Section 5.2.

[13] This window size ensures that we have enough observations to conduct estimation, accounts for the possibility of parameter instability in the data generating process and allows us to start the out-of-sample as early as 1990q1, such that results are qualitatively comparable to the mainstream literature. Furthermore, nowhere in this study do we attempt to optimize the performance of our models by modifying these settings, in recognition of the concern over data mining discussed in Rossi and Inoue (2012).

As previously mentioned, we use the test suggested in Clark and West (2007) to evaluate the evidence of population-level predictability. The idea of the test is as follows: Under the null hypothesis, the benchmark model generates the data. Therefore, the population mean square error (MSE) produced under the benchmark is smaller than the population MSE produced under the augmented model. Given the nested structure of the models, Clark and West (2007) adjust the estimated MSE difference to account for additional noise associated with the larger model's forecast. More precisely, let $\widehat{y}^b_{t+1}$ and $\widehat{y}^l_{t+1}$ denote the one-quarter ahead point forecast of y_{t+1} produced under the benchmark and model l, respectively. The Clark and West (2007) test statistic, henceforth labeled as CW is given as:

$$CW = n^{-1}\sum_{j=t+1}\left(y_j - \widehat{y}^b_j\right)^2 - n^{-1}\sum_{j=t+1}\left(y_j - \widehat{y}^l_j\right)^2 + n^{-1}\sum_{j=t+1}\left(\widehat{y}^b_j - \widehat{y}^l_j\right)^2. \tag{3}$$

The first two terms in (5.1), $n^{-1}\Sigma^T_{j=t+1}(y_j - \widehat{y}^b_j)^2$ and $n^{-1}\Sigma^T_{j=t+1}(y_j - \widehat{y}^l_j)^2$, are the MSEs produced under the benchmark and model l, respectively. The last term in (5.1), $n^{-1}\Sigma^T_{j=t+1}(\widehat{y}^b_j - \widehat{y}^l_j)^2$, captures the adjustment for additional noise associated with model l's forecasts. It involves the average squared difference between point forecasts generated by the benchmark and model l. Accordingly, when point forecasts produced under model l are highly volatile compared to point forecasts produced under the benchmark, the additional noise associated with parameter estimation is large and the adjustment term in (5.1) is therefore also large. Following Clark and West (2007), once constructed, we compare (5.1) with one-sided critical values from a standard Normal distribution.[14] This is because under the alternative hypothesis, the population MSE produced under model l is less than the population MSE produced under the benchmark model.

To determine whether models that employ the crude oil price variables displayed in Table 1 one by one produce statistically significant more accurate point forecasts than the benchmark, we use the Harvey, Leybourne, and Newbold (1997) small sample correction of the Diebold and Mariano (1995) test given as:

$$DM = \Lambda\cdot\left(\frac{MSE^b - MSE^l}{\widehat{\sigma}_{HAC}/\sqrt{n}}\right), \tag{4}$$

where Λ is the correction term suggested in Harvey et al. (1997) and $\widehat{\sigma}_{HAC}$ is a HAC estimator of the asymptotic variance, $\sigma^2\sqrt{n}(MSE^b - MSE^l)$. Under the Diebold and Mariano (1995) null hypothesis, the benchmark and model l possess equal predictive ability, i.e., $E[MSE^b - MSE^l] = 0$.

5.2. Results

We consider (3.2) with the variables in Table 1 one by one, (3.1) and generate $h = 1, \ldots, 5$-quarters ahead forecasts using both ex-post revised and real-time GDP data. Tables 2 and 3 report out-of-sample results using the real and the nominal price of crude oil, respectively. In these tables, columns labeled CW p-values and DM p-values report p-values associated with (5.1) and (5.2). We also report Theil's U (TU) defined as the MSE produced under (3.2) with the variable of interest relative to the MSE produced under (3.1).

Overall, our results are very interesting. To begin with, we observe minor changes when we compare results using the real price of crude oil with results using the nominal price of crude oil. Furthermore, although we rely on U.K. real GDP data, our out-of-sample results are qualitatively similar to studies, such as Kilian and Vigfusson (2013) and Ravazzolo and Rothman (2013) who rely on real GDP U.S. data, namely, at the one-quarter ahead forecast horizon, our predictive models fail to produce lower MSEs than the benchmark. Likewise, all CW and DM p-values are consistently above 10%. Conversely, as we increase the forecast horizon, we tend to observe predictive gains for certain specifications.

What explains the better performance of certain models at the longer forecast horizons? This finding can be explained by the gradual information diffusion hypothesis suggested Hong and Stein (1999) and Hong, Torous, and Valkanov (2007), where the authors argue that (rather surprisingly) information tends to disperse slowly and as a consequence it takes time before information about the price of crude oil becomes fully reflected in the economy. Essentially, as argued in these studies, the total market response to news/information that has a substantial impact on economic activity involves the aggregation of private signals. In situations when economic agents have difficulty assessing the impact of the news or when they simply react to information at different points in time, then such underreaction can occur.

As displayed in Tables 2 and 3, augmenting (3.1) with the percentage change in the price of crude oil does not generally lead to forecast improvements relative to the benchmark. The TUs are often above one and the respective no predictability null hypotheses cannot be rejected. The only exemption is $h = 3$, where the Clark and West (2007) null hypothesis is rejected at the 10% significance level for ex-post revised data and at the 5% significance level for real-time GDP data. For the latter, gains in forecast accuracy relative to (3.1) are around 3%. However, as indicated by the DM p-values, they are not statistically significant.

[14] To compute the Clark and West (2007) test statistic, we regress the adjusted out-of-sample squared error differences on a constant and examine the associated t-statistic. The t-test is performed using a heteroscedasticity and autocorrelation-consistent (HAC) variance, $\widehat{\sigma}_{HAC}$ computed with the pre-whitened quadratic spectral estimator of Andrews and Monahan (1992).

Table 2
Out-of-sample forecast evaluation produced under (3.2) with the predictor of interest relative to (3.1) using the real price of crude oil.

Forecast horizon	Ex-post revised data			real-time data			ex-post revised data			real-time data		
	CW p − value	DM p − value	TU	CW p − value	DM p − value	TU	CW p − value	DM p − value	TU	CW p − value	DM p − value	TU
$\triangle oil$							*anet*					
1	0.36	0.76	1.02	0.49	0.84	1.03	0.62	0.87	1.09	0.91	0.96	1.15
2	0.59	0.94	1.03	0.40	0.78	1.03	0.17	0.90	1.08	0.10	0.97	1.13
3	0.08	0.62	1.01	0.04	0.30	0.97	0.05	0.28	0.96	0.03	0.35	0.98
4	0.34	0.87	1.08	0.52	0.87	1.14	0.17	0.93	1.10	0.18	0.88	1.13
5	0.38	0.90	1.07	0.03	0.66	1.01	0.27	0.95	1.09	0.14	0.52	1.00
net^{+}							*gap*					
1	0.58	0.73	1.02	0.74	0.86	1.05	0.48	0.91	1.07	0.52	0.85	1.11
2	0.09	0.79	1.05	0.14	0.88	1.06	0.04	0.58	1.01	0.08	0.76	1.06
3	0.07	0.30	0.96	0.04	0.27	0.95	0.03	0.59	1.02	0.07	0.85	1.12
4	0.11	0.82	1.06	0.10	0.83	1.12	0.04	0.64	1.03	0.05	0.92	1.12
5	0.19	0.87	1.07	0.13	0.48	1.00	0.01	0.54	1.01	0.04	0.87	1.08
net^{-}							*large*					
1	0.42	0.81	1.05	0.82	0.87	1.09	0.64	0.81	1.04	0.65	0.85	1.05
2	0.24	0.80	1.02	0.16	0.85	1.05	0.44	0.95	1.04	0.83	0.98	1.09
3	0.42	0.21	0.98	0.32	0.78	1.03	0.21	0.81	1.03	0.10	0.59	1.01
4	0.58	0.81	1.03	0.70	0.89	1.03	0.38	0.89	1.08	0.70	0.89	1.12
5	0.64	0.80	1.02	0.33	0.63	1.01	0.74	0.93	1.13	0.22	0.92	1.05
net							$large^{+}$					
1	0.28	0.72	1.03	0.21	0.51	1.00	0.95	0.96	1.13	0.85	0.86	1.21
2	0.05	0.66	1.02	0.03	0.67	1.03	0.57	0.97	1.06	0.59	0.98	1.08
3	0.02	0.36	0.97	0.01	0.62	1.03	0.08	0.46	1.00	0.08	0.39	0.99
4	0.05	0.51	1.00	0.04	0.93	1.08	0.07	0.64	1.02	0.24	0.92	1.16
5	0.20	0.95	1.05	0.08	0.52	1.00	0.04	0.83	1.05	0.19	0.65	1.01

CW p-value columns report p-values associated with the null hypothesis of no population-level predictability as specified in Clark and West (2007). DM p-value columns report p-values associated with the null hypothesis of no finite-sample predictability as specified in Diebold and Mariano (1995). TU reports the MSE produced under the specification of interest relative to the MSE produced the benchmark.

Table 3
Out-of-sample forecast evaluation produced under (3.2) with the predictor of interest relative to (3.1) using the nominal price of crude oil.

Forecast horizon	Ex-post revised data			real-time data			ex-post revised data			real-time data		
	CW p − value	DM p − value	TU	CW p − value	DM p − value	TU	CW p − value	DM p − value	TU	CW p − value	DM p − value	TU
$\triangle oil$							*anet*					
1	0.39	0.79	1.02	0.47	0.79	1.03	0.74	0.89	1.12	0.95	0.95	1.19
2	0.54	0.93	1.03	0.39	0.78	1.03	0.14	0.88	1.06	0.08	0.95	1.09
3	0.08	0.64	1.02	0.04	0.35	0.98	0.04	0.21	0.94	0.02	0.27	0.96
4	0.35	0.88	1.08	0.54	0.87	1.15	0.14	0.84	1.06	0.11	0.86	1.09
5	0.38	0.90	1.08	0.03	0.73	1.02	0.17	0.93	1.07	0.12	0.58	1.01
net^{+}							*gap*					
1	0.73	0.83	1.03	0.81	0.90	1.05	0.60	0.93	1.07	0.68	0.91	1.14
2	0.13	0.74	1.03	0.10	0.82	1.04	0.04	0.57	1.01	0.10	0.76	1.06
3	0.05	0.21	0.94	0.02	0.21	0.93	0.03	0.59	1.02	0.10	0.85	1.13
4	0.10	0.65	1.02	0.05	0.77	1.07	0.05	0.66	1.04	0.09	0.91	1.12
5	0.16	0.87	1.06	0.12	0.50	1.00	0.01	0.57	1.01	0.05	0.92	1.09
net^{-}							*large*					
1	0.41	0.81	1.07	0.87	0.87	1.11	0.73	0.85	1.05	0.69	0.84	1.05
2	0.13	0.78	1.01	0.17	0.85	1.04	0.37	0.96	1.04	0.73	0.97	1.08
3	0.24	0.42	1.00	0.36	0.82	1.03	0.21	0.80	1.04	0.13	0.69	1.01
4	0.62	0.84	1.04	0.81	0.93	1.04	0.35	0.88	1.08	0.68	0.89	1.13
5	0.37	0.69	1.01	0.31	0.67	1.02	0.71	0.93	1.13	0.13	0.87	1.04
net							$large^{+}$					
1	0.36	0.79	1.03	0.23	0.55	1.01	0.95	0.96	1.12	0.84	0.86	1.20
2	0.05	0.65	1.02	0.03	0.66	1.02	0.56	0.97	1.05	0.59	0.98	1.07
3	0.02	0.34	0.96	0.00	0.56	1.02	0.09	0.45	1.00	0.15	0.53	1.00
4	0.04	0.49	1.00	0.02	0.90	1.08	0.07	0.62	1.02	0.24	0.92	1.16
5	0.15	0.95	1.06	0.07	0.55	1.01	0.02	0.79	1.05	0.10	0.50	1.00

CW p-value columns report p-values associated with the null hypothesis of no population-level predictability as specified in Clark and West (2007). DM p-value columns report p-values associated with the null hypothesis of no finite-sample predictability as specified in Diebold and Mariano (1995). TU reports the MSE produced under the specification of interest relative to the MSE produced the benchmark.

Among the nonlinear models, forecasts employing the three-year net crude oil price change and the three-year crude oil price gap tend to perform best. Regardless of whether we use ex-post revised instead of real-time GDP data or likewise whether we rely on the nominal versus the real price of crude oil, the Clark and West (2007) null hypothesis is often rejected for $h = 2$, 3 and 4 at the 5% significance level. With regards to point forecast accuracy, reductions in MSE produced under (3.2) with *net* relative to the benchmark are around 4% at $h = 3$ for ex-post revised data. However, as indicated by the DM p-values, these MSE reductions are not statistically significant. In certain instances, results are also positive for net^+, *anet* and $large^+$. Compared to *net* and *gap*, forecasts produced under (3.2) with net^+ and (3.2) with $large^+$ tend to display some sensitivity depending on whether we use ex-post revised or real-time GDP data. Conversely, we do not observe any statistically significant forecast improvements from employing net^- and *large* as predictors in (3.1). Here, the no predictability null hypotheses are rejected across all forecast horizons. Likewise, the TUs are consistently above one.

Accordingly, it is possible that using a different crude oil price series might change our conclusions. Therefore, we repeat our out-of-sample analysis using the WTI price of crude oil. However, we observe that our out-of-sample results are very similar to those reported in Tables 2 and 3 This is not surprising since Brent Blend and WTI prices are highly correlated. Likewise, we evaluate the predictive content from the U.K. real GDP growth rate to the price of crude oil to test whether there is feedback from the former to the latter, which would mean that our predictive regressions in Section 3 are misspecified. However, we observe that the real GDP growth rate does not have a predictive impact on the price of crude oil, meaning that we do not need to worry about such misspecification issues.

The above analysis is conducted by comparing forecasts produced under the predictive models relative to forecasts produced under the benchmark one by one. However, as previously mentioned, to address potential data snooping concerns, we also perform multiple forecast comparison tests. We do so using the framework suggested in White (2000) as well as the model confidence set (MCS) suggested Hansen et al. (2011). Contrary to pairwise comparison, the null hypothesis of White (2000) is that that all alternative forecasts are as good as forecasts produced under the benchmark. The MCS procedure determines the set of "best" models that forecast better that other models in the set. With regards to finite-sample predictive ability, the null hypothesis of White (2000) cannot be rejected across all h, regardless of whether we rely on ex-post revised or real-time GDP data. Likewise, the benchmark model consistently belongs to MCS at the 1%, 5% and 10% significance levels, respectively. A similar pattern is observed with regards to population-level predictability. Here, we only observe White (2000) p-values close to 10% at $h = 5$. Results from MCS document that the benchmark model belongs to MCS across all forecast horizons, except for $h = 4$ and $h = 5$ at the 10% significance level. Here, MCS is composed of the most successful models, such as the three-year crude oil price gap, net crude oil price change and large crude oil price change.

Overall, out-of-sample results illustrate that the nature and magnitude of the predictive impact of crude oil price changes on the U.K. GDP growth rate is very complex. It depends on the nature of nonlinearity and the definition of forecast improvement. The observed gains are limited to long forecast horizons and exclusively at the population level. Furthermore, neglecting data snooping concerns can result in misleading conclusions. For example, MCS results indicate that only few specifications end up as statistically significant at the population level. Even setting aside concerns regarding statistical significance, it is rather doubtful that the forecast gains are economically meaningful as the largest MSE reduction is around 6%.

Accordingly, there is the possibility that our predictive regressions are misspecified and addressing potentially harmful misspecifications can change the current picture. Likewise, it is possible that results based on local predictive performance can provide insights into how the relative predictive power afforded by our crude oil price variables changes over the out-of-sample period and help understand the relative weak predictive power afforded by them at the global level. We will address these issues in the next sections.

5.3. Omitted variable bias

There is of course the possibility that the failure of the crude oil price-based models to deliver forecast accuracy gains or likewise the relative success of certain nonlinear models with respect to population level-predictability is due to an omitted variable in (3.1) and (3.2). Particularly, when specifying our predictive models as (3.2), we risk leaving information possibly correlated with the crude oil price measure, X_{t-i}, in the residuals. This leads to biased OLS estimates due to violation of the exogeneity assumption, which in turn, can have severe consequences on the quality of the generated forecasts. Therefore, to examine this possibility, we follow Ravazzolo and Rothman (2013) and consider the following benchmark model:

$$y_t = \varphi_0 + \sum_{i=1} \varphi_i y_{t-i} + \sum_{i=1}^{\tilde{q}} \rho_i Z_{t-i} + \varepsilon_t, \quad \varepsilon_t \tilde{} N\left(0, \sigma^2\right), \tag{5}$$

where Z_{t-i} is a non-crude oil price variable at time $t - i$. As an alternative to (5.3), we consider

$$y_t = \varphi_0 + \sum_{i=1} \varphi_i y_{t-i} + \sum_{i=1} \beta_i X_{t-i} + \sum_{i=1}^{\tilde{q}} \rho_i Z_{t-i} + \varepsilon_t, \quad \varepsilon_t \tilde{} N\left(0, \sigma^2\right). \tag{6}$$

The variables to include in (5.3) and (5.4) can be large. Therefore, following Ravazzolo and Rothman (2013), we identify variables that are considered as indicators of the state of the U.K. economy, international business cycle and are arguably also correlated with the price of crude oil. They are: The percentage change in global crude oil production, first difference of Kilian's real global activity, Moody's Baa-Aaa spread, the U.S. term spread, the first difference of the three-month and ten-year U.K. government bond yields, respectively.[15] The percentage change in global crude oil production represents changes in global crude oil supply. Kilian's global activity measure is constructed using data published by Drewry Shipping Consultants Ltd based on various bulk dry cargoes shipping freight rates measured in dollars per metric ton. This indicator has become a popular choice to represent global real economic activity. The Baa-Aaa spread and the U.S. term spread are generally considered as an important indicator of the international business cycle. Finally, changes in the U.K. government bond yields contain information on the state of the U.K. economy.

For each of these variables, we produce forecasts under (5.3) and compare results to (3.1). We obtain improvements relative to the benchmark mainly for the Baa-Aaa spread. We then use this variable to evaluate the performance (5.4) relative to (5.3), where as before X_{t-i} corresponds to the variables in Table 1.

As illustrated in Tables 4 and 5, our conclusions remain largely the same as in Section 5.2 and we do not find any notable changes worth mentioning. For example, similar to previous results, the CW null hypothesis is consistently rejected at longer forecast horizons for (5.4) with *net* and (5.4) with *gap* using ex-post revised and real-time GDP data. Furthermore, contrary to results in Section 5.2, the CW test p-values produced under (5.4) with net^{+} decrease at $h=2$ and $h=4$ in Tables 4 and 5, suggesting that omission of the Baa-Aaa spread may be a factor behind rejections of the CW null hypothesis for this model in Table 2. A similar pattern is observed for *anet*. With respect to forecast accuracy, we obtain gains around 5% to 6% from employing net^{+} and *net* one by one at $h=3$ for ex-post revised data. However, similar to previous results, we fail to reject the null hypothesis of equal predictive ability in favor of the augmented specifications. This shows that DM results reported in Tables 2 and 3 are not due to omission of the Baa-Aaa spread from our models. We also repeat the multiple comparison tests using the MCS procedure. Here, results are similar to results reported in the previous section and we do not find any changes worth mentioning.

5.4. Discussion

Thus far, we have established several interesting results. However, we have not yet explained the notable differences between CW and DM-based results. The aim of this section is to provide an intuitive answer to this question. Essentially, the contrast between population-level predictability and finite-sample forecast accuracy results illustrate the distinction between the definition of the two out-of-sample tests. The relative positive Clark and West (2007) test results is mainly attributable to a large upward adjustment term in (5.1), namely, $n^{-1}\Sigma_{j=t+1}^{T}\left(\widehat{y}_j^{b}-\widehat{y}_j^{l}\right)^2$. Clark and West (2007) incorporate this term to account for the additional noise related to estimating the larger model. Contrary to DM, when point forecasts produced under the augmented model are more volatile than those produced under the benchmark, the adjustment term in (5.1) becomes correspondingly large and increases the test statistic. Hence, although our crude oil price-based specifications underperform the benchmark (as indicated by the TUs), they do not do so by enough to be consistent with the null hypothesis of no population-level predictability. As a consequence, the Clark and West (2007) null hypothesis is rejected.

To better explain the pattern of results obtained in our analysis, we conduct a Monte Carlo analysis in a similar fashion as Paye (2012). Data are simulated from the following data generating process (DGP):

$$y_t = 0.24 + 0.16y_{t-1} + 0.24y_{t-2} + \beta_1 X_{t-1} + \varepsilon_t, \quad \varepsilon_t \tilde{} N\left(0,\sigma^2\right),$$

$$X_t = -0.04 + 0.85X_{t-1} + \eta_t, \quad \eta_t \tilde{} N\left(0,\omega^2\right),$$

where $\sigma^2 = 0.57$, $\omega^2 = 0.02$ and $E[\varepsilon_t\eta_t] = 0.01$. These DGP values are based on estimates of a restricted VAR(2) using the full sample ex-post revised GDP data, where X_t corresponds to the three-year crude oil price gap. In one set of simulations, β_1 is set to the corresponding in-sample estimate, namely, -0.25. A second set of simulations imposes the null hypothesis of no population predictability by setting $\beta_1 = 0$.

The simulation analysis focuses on a rolling window estimation scheme, where the window size is identical to that in our empirical analysis. We consider out-of-sample sizes ranging from 50 to 250. For each set of simulated data, one-step ahead forecasts are computed over the out-of-sample period for both the benchmark and the predictive model. The resulting MSEs are used to construct the CW and DM test statistics. We approximate the probabilities that these tests reject their respective null hypothesis at the 10% level using the proportion of rejections over thousand simulated samples. Results are reported in Table 6.

Columns two to four of the table address the case in which there is population-level predictability, i.e., $\beta_1 \neq 0$. Hence, the null hypothesis of the CW test is false. As indicated by the rejection probabilities, the power of the CW test is strictly increasing in out-of-sample size. Let use consider $n = 100$, which corresponds roughly to the out-of-sample sample size in our study. Here, the power of the CW test is approximately 20%. On the other hand, the DM test is not at all able to distinguish between the generated forecasts at the same rate. This is because the difference in the generated MSEs are on average very small, less that 0.1%. As we increase the out-of-sample size, the percentages of rejections for the DM test increase. However, they remain below those for the CW test. We also

[15] Kilian's basic idea is that changes in global economic activity drive demand for shipping. In the short-run, this higher demand shows up as an increase in the real cost of shipping. Kilian and Zhou (2018) compare several measures of global real economic activity. They find that traditional proxies for global real GDP and global industrial production are less suited for modeling industrial commodity prices than the global real economic activity index constructed by Kilian (2009).

Table 4
Out-of-sample forecast evaluation produced under (5.4) with the predictor of interest relative to (5.3), where Z_{t-i} corresponds to the Baa-Aaa spread using the real price of crude oil.

Forecast horizon	Ex-post revised data			real-time data			ex-post revised data			real-time data		
	CW p − value	DM p − value	TU	CW p − value	DM p − value	TU	CW p − value	DM p − value	TU	CW p − value	DM p − value	TU
$\triangle oil$							*anet*					
1	0.69	0.86	1.07	0.67	0.92	1.05	0.55	0.85	1.15	0.89	0.92	1.23
2	0.59	0.94	1.03	0.52	0.82	1.03	0.12	0.88	1.14	0.09	0.93	1.20
3	0.06	0.23	0.97	0.11	0.29	0.97	0.02	0.28	0.94	0.06	0.75	1.06
4	0.17	0.67	1.03	0.43	0.78	1.08	0.06	0.68	1.04	0.27	0.82	1.10
5	0.29	0.77	1.04	0.07	0.17	0.96	0.24	0.84	1.07	0.04	0.27	0.97
net^{+}							*gap*					
1	0.43	0.63	1.01	0.63	0.83	1.04	0.45	0.95	1.10	0.52	0.91	1.16
2	0.05	0.54	1.01	0.11	0.82	1.06	0.04	0.45	0.99	0.09	0.74	1.07
3	0.03	0.24	0.93	0.08	0.51	1.00	0.02	0.55	1.01	0.06	0.70	1.08
4	0.06	0.69	1.04	0.09	0.82	1.10	0.02	0.55	1.02	0.05	0.55	1.01
5	0.17	0.75	1.06	0.16	0.48	1.00	0.02	0.38	0.97	0.04	0.54	1.01
net^{-}							*large*					
1	0.46	0.82	1.11	0.84	0.86	1.15	0.81	0.90	1.08	0.78	0.92	1.07
2	0.64	0.88	1.09	0.15	0.84	1.10	0.31	0.68	1.02	0.97	0.99	1.11
3	0.27	0.23	0.98	0.25	0.63	1.02	0.25	0.32	0.98	0.20	0.43	0.99
4	0.31	0.20	0.98	0.29	0.39	0.99	0.20	0.66	1.02	0.64	0.87	1.09
5	0.55	0.82	1.02	0.18	0.28	0.98	0.65	0.88	1.10	0.13	0.48	1.00
net							$large^{+}$					
1	0.13	0.65	1.02	0.15	0.53	1.01	0.96	0.97	1.13	0.83	0.85	1.21
2	0.02	0.50	1.00	0.04	0.66	1.03	0.65	0.93	1.04	0.73	0.97	1.07
3	0.00	0.33	0.95	0.04	0.56	1.03	0.05	0.37	0.99	0.11	0.44	1.00
4	0.01	0.35	0.97	0.06	0.68	1.04	0.06	0.60	1.01	0.38	0.86	1.11
5	0.15	0.73	1.03	0.06	0.28	0.97	0.08	0.75	1.04	0.07	0.26	0.97

CW p-value columns report p-values associated with the null hypothesis of no population-level predictability as specified in Clark and West (2007). DM p-value columns report p-values associated with the null hypothesis of no finite-sample predictability as specified in Diebold and Mariano (1995). TU reports the MSE produced under the specification of interest relative to the MSE produced the benchmark.

Table 5

Out-of-sample forecast evaluation produced under (5.4) with the predictor of interest relative to (5.3), where Z_{t-i} corresponds to the Baa-Aaa spread using the nominal price of crude oil.

Forecast horizon	Ex-post revised data			real-time data			ex-post revised data			real-time data		
	CW $p-$ value	DM $p-$ value	TU	CW $p-$ value	DM $p-$ value	TU	CW $p-$ value	DM $p-$ value	TU	CW $p-$ value	DM $p-$ value	TU
$\triangle oil$							*anet*					
1	0.72	0.88	1.07	0.64	0.88	1.05	0.69	0.87	1.18	0.93	0.92	1.26
2	0.44	0.89	1.02	0.49	0.80	1.04	0.10	0.85	1.09	0.06	0.92	1.14
3	0.06	0.23	0.97	0.10	0.30	0.97	0.01	0.20	0.92	0.03	0.61	1.02
4	0.18	0.69	1.03	0.43	0.78	1.08	0.04	0.55	1.01	0.14	0.72	1.05
5	0.28	0.77	1.04	0.07	0.19	0.97	0.15	0.77	1.06	0.04	0.29	0.97
net^{+}							*gap*					
1	0.63	0.78	1.02	0.75	0.88	1.05	0.48	0.95	1.09	0.61	0.93	1.16
2	0.03	0.48	1.00	0.07	0.77	1.04	0.04	0.44	0.99	0.10	0.72	1.06
3	0.02	0.17	0.91	0.05	0.39	0.98	0.02	0.53	1.01	0.07	0.69	1.07
4	0.05	0.57	1.01	0.16	0.74	1.06	0.02	0.51	1.00	0.05	0.50	1.00
5	0.18	0.72	1.05	0.08	0.45	0.99	0.02	0.36	0.97	0.04	0.52	1.00
net^{-}							*large*					
1	0.43	0.81	1.12	0.82	0.86	1.17	0.86	0.93	1.08	0.80	0.90	1.07
2	0.62	0.88	1.06	0.16	0.81	1.06	0.27	0.61	1.01	0.94	0.99	1.11
3	0.13	0.16	0.99	0.25	0.54	1.01	0.13	0.33	0.98	0.23	0.46	0.99
4	0.14	0.23	0.99	0.25	0.36	0.99	0.19	0.66	1.02	0.63	0.87	1.09
5	0.28	0.57	1.00	0.18	0.29	0.97	0.63	0.88	1.10	0.11	0.39	0.99
net							$large^{+}$					
1	0.19	0.70	1.03	0.17	0.55	1.01	0.96	0.98	1.12	0.83	0.85	1.20
2	0.02	0.46	0.99	0.02	0.65	1.03	0.63	0.93	1.04	0.73	0.97	1.07
3	0.00	0.31	0.94	0.02	0.53	1.01	0.07	0.37	0.99	0.25	0.59	1.00
4	0.01	0.37	0.98	0.06	0.65	1.04	0.06	0.59	1.01	0.37	0.85	1.10
5	0.12	0.74	1.04	0.05	0.33	0.97	0.06	0.72	1.04	0.03	0.19	0.96

CW p-value columns report p-values associated with the null hypothesis of no population-level predictability as specified in Clark and West (2007). DM p-value columns report p-values associated with the null hypothesis of no finite-sample predictability as specified in Diebold and Mariano (1995). TU reports the MSE produced under the specification of interest relative to the MSE produced the benchmark.

Table 6
Monte Carlo analysis illustrating the difference between Clark and West (2007) and Diebold and Mariano (1995).

Simulation under	population-level predictability			no population-level predictability		
Out-of-sample size	CW	DM	CW ¬ DM^C	CW	DM	CW ¬ DM^C
50	18.20	7.20	17.00	8.20	8.30	7.40
100	20.10	10.70	19.30	6.90	16.40	6.60
150	26.30	11.70	25.60	8.60	21.60	8.60
200	27.40	14.80	27.40	7.80	29.90	7.80
250	28.40	18.50	28.10	8.30	34.60	8.30

CW and DM report the percentage of simulations under each out-of-sample size in which the Clark and West (2007) and Diebold and Mariano (1995) tests reject their respective null hypothesis at the 10% significance level. Columns labeled CW ¬ DM^C denote the percentage of simulations in which the Clark and West (2007) test rejects the null hypothesis of no population-level predictability, but the Diebold and Mariano (1995) test fails to reject the null hypothesis of equal predictive ability.

report the number of times over the thousand replications, where CW rejects but DM fails to reject the null hypothesis. As evidenced by the percentages, when the CW test rejects, it is typically the case that the DM test fails to reject. For example, the CW test rejection rate is 27% when $n = 200$. The event CW¬DM^C, i.e., CW rejects but DM fails to reject occurs nearly at the same rate.

In the last three columns of Table 6, $\beta_1 = 0$ in the DGP, i.e. there is no predictive impact. Under the null hypothesis of no population-level predictability, the rejection proportions reported for the CW test characterize size. At the same time, the null hypothesis of equal predictive ability is also false. Hence, the rejection probabilities for the DM test indicate power. Here, the CW test appears to be relatively well-sized. Consistent with simulation results reported in Clark and West (2007), the test is slightly undersized. The power of the DM test is also increasing in out-of-sample size. This means that fact that as n increases, the deviation from the null hypothesis of equal predictive ability increases. On average, forecasts produced under the predictive model underperform the benchmark by 1%. Finally, as illustrated by the last column when the CW test rejects, the DM test also tends to reject the null hypothesis nearly at the same rate.

To sum up, the Monte Carlo results are consistent with our findings reported in Section 5.2. The power of the Diebold and Mariano (1995) test to distinguish between forecasts when it is actually the case (given our DGP values) is very weak. When it does contain power, it is concentrated in the direction of rejecting the null in favor of forecasts produced under the benchmark. On the other hand, the Clark and West (2007) test has reasonable power to detect population-level predictability under our setup.

5.5. *Local relative predictive performance*

As our final analysis, we evaluate the local predictive performance of our crude oil price-based models relative to the benchmark. Therefore, we utilize the fluctuation test suggested in Giacomini and Rossi (2010). Essentially, the fluctuation test compares the predictive performance of model l with model b over rolling windows of out-of-sample data. Following Giacomini and Rossi (2010), the re-scaled loss difference over a rolling window of m observations centered at time s is computed as follows:

$$F_{s,m} = \widehat{\sigma}_{HAC}^{-1} m^{-1/2}\Big(\sum_{j=s-m/2} \triangle L_j\Big), \tag{7}$$

where $\triangle L_j$ is the forecast loss difference between the benchmark and model l over $.s = t+1+m/2, \ldots, T-m/2$[16] In (5.5), $\triangle L_j$ corresponds either to the error difference in Diebold and Mariano (1995) or Clark and West (2007). For example, with regards to the latter, we have that $\triangle L_j = \widehat{\varepsilon}_j^{b,2} - [\widehat{\varepsilon}_j^{l,2} - (\widehat{y}_j^b - \widehat{y}_j^l)^2]$, where $\widehat{\varepsilon}_j^{b,2} = (y_j - \widehat{y}_j^b)^2$ and $\widehat{\varepsilon}_j^{l,2} = (y_j - \widehat{y}_j^l)^2$, see also Clark and West (2007). Accordingly, if $\max\{F_s\}$ is higher than the critical value, k_α, provided in Giacomini and Rossi (2010), then the null hypothesis of no predictability for the m-period window centered at s is rejected in favor of the alternative. The fluctuation test is implemented by plotting (5.5) over n for each specification together with the critical values.

Since results are qualitatively similar for the Clark and West (2007) and Diebold and Mariano (1995) tests, we focus on discussing results for the former. Figs. 2 and 3 display (5.5) with $m = 20$, where $\triangle L_j$ is equal to the Clark and West (2007) error difference at $h = 1$ and $h = 4$, respectively. Panels (a) and (c) report results using ex-post revised GDP data, whereas real-time GDP-based results are reported in panels (b) and (d). The dotted lines indicate the 10% one-sided critical values of the test (recall the CW test is one-sided).

Overall, we observe two very interesting results. To begin with, Figs. 2 and 3 demonstrate ample evidence of time-variation in predictive performance relative to the benchmark. Often the resulting forecasts vary, sometimes performing well and sometimes bad relative to the benchmark. This in turn, explains why our alternatives to the benchmark perform poorly at $h = 1$ because all predictive gains are more than offset by losses. For instance, as displayed in panels (a) and (b) of Fig. 2, the specification with *net* performs very well at the start of the out-of-sample. Here, (5.5) even briefly exceeds the 10% critical value. However, *net* quickly losses its predictive power and as a consequence, (5.5) remains essentially flat until the onset of the Great Recession. A similar pattern is observed in the context of

[16] The asymptotic distribution of the fluctuation test under the null hypothesis can be approximated by functionals of Brownian motions. Critical values for various significance levels are provided in Giacomini and Rossi (2010).

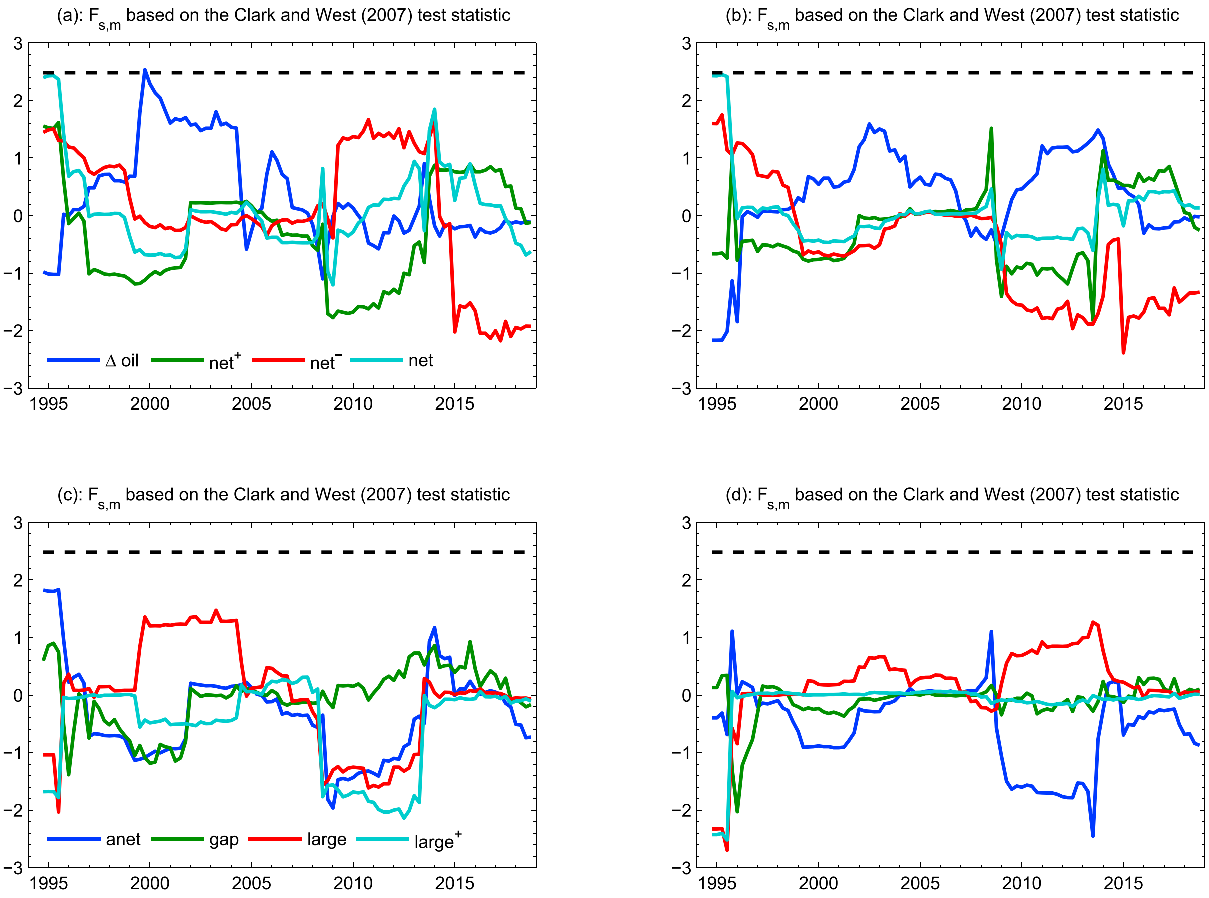

Fig. 2. One-quarter ahead fluctuation test results for population-level predictability.
This figure displays Giacomini and Rossi (2010) fluctuation test results based on sequences of the Clark and West (2007) error differences. The size of the rolling window used to compute (5.5), namely, m, is fixed at 20 quarters. Fluctuation test critical values at the 10% significance level are shown by dotted horizontal lines.

forecasts employing Δoil, net^+ and *large* in panels (a) and (c) of Fig. 2, where any short-lived predictive gains are offset by subsequent losses.

We also observe that in contrast to results based on average out-of-sample performance, local-level predictive performance tends to differ strongly across the use of ex-post revised and real-time GDP data. We invite the reader to examine each figure more carefully. Here, we provide some illustrative examples. For instance, consider the performance of (3.2) with *net* at $h = 4$. As illustrated in Table 2, for this specification, the null hypothesis of no population-level predictability is rejected for both ex-post revised and real-time data: The former at the 10% significance level and the latter at the 5% level. Likewise, the MSEs relative to the benchmark are similar. However, results differ considerably when we focus on local predictive performance. As demonstrated in panel (a) of Fig. 2, for ex-post revised real GDP data, (3.2) with *net* performs very well at the start of the out-of-sample before losing its predictive power. Thereafter, (5.5) remains flat until the onset of the Great Recession. During this period, we observe a short-lived increase in (5.5). However, the population-level predictability gains gradually vanish towards the end of the sample. With regards to real-time GDP data reported in panel (b) of Fig. 3, we observe the same pattern at the start of the out-of-sample period. However, contrary to results reported in panel (a), (3.2) with *net* responds much more aggressively to the Great Recession and delivers notable improvements following the Great Recession. As evidenced from panel (b), the local-level null hypothesis of no population-level predictability is rejected for the window centered at 2014. The three-year crude oil price gap displays a similar pattern. Another illustrative example is the performance of (3.2) with net^- at $h = 1$ and $h = 4$. As illustrated in panels (a) and (b) of Figs. 2 and 3, the fluctuation test results differ strongly across the use of ex-post revised and real-time GDP data from 2008 until 2015. For example, at $h = 4$, (5.5) remains largely unchanged when we use ex-post revised data before shifting downwards around 2015. On the other hand, for real-time GDP data, (5.5) shifts downwards from the onset of the Great Recession until 2015.

Fig. 3. Four-quarters ahead fluctuation test results for population-level predictability.
This figure displays Giacomini and Rossi (2010) fluctuation test results based on sequences of the Clark and West (2007) error differences. The size of the rolling window used to compute (5.5), namely, *m*, is fixed at 20 quarters. Fluctuation test critical values at the 10% significance level are shown by dotted horizontal lines.

6. Conclusion

This study examines predictive regressions for the real U.K. GDP growth rate that attempt to exploit information contained in the price of crude oil. Motivated by past studies, we identify a set of crude oil price variables and test their ability to improve real GDP growth rate forecasts. We observe that our conclusions depend on the definition of forecast improvement and whether we rely on pairwise or multiple comparison. Pairwise and multiple forecast comparison test results document that there are no statistically significant predictive gains afforded by our crude oil price variables at the one-quarter ahead forecast horizon. However, at horizons beyond one, the null hypothesis of no population-level predictability is often rejected when individually comparing certain nonlinear models with the benchmark. Not everything is positive, though. When conducting multiple forecast comparison tests, we observe that the evidence of population-level predictability weakens and is limited only to few instances at four and five-quarters ahead forecast horizons.

We also explore whether our out-of-sample results are affected by omitted variable bias and conclude that they are not. Likewise, our Monte Carlo analysis illustrates the pattern of divergence between the Diebold and Mariano (1995) and Clark and West (2007) tests consistent with that observed in the data. Finally, when focusing on global relative predictive performance, we do not find strong differences between using real-time and ex-post revised data. However, when focusing on local relative predictive performance, we observe that results tend to differ substantially across the use of ex-post revised and real-time GDP data. Time-series plots of re-scaled loss differences over the out-of-sample period display that predictive gains are short-lived and concentrate in specific periods. Furthermore, our predictors exhibit different patterns of improvements and a uniform pattern of gains cannot be detected.

Our analysis adds to a growing number of studies that explore the role of nonlinearities between the price of crude oil and the level of economic activity, with focus mainly on U.S. GDP data. As one of the first, we have analyzed this topic using U.K. GDP data. Given the link between crude oil price changes and the GDP growth rate established in Section 2, a tendency exists to assume that using information contained in the price of crude oil would greatly improve real GDP forecasts. Our relatively comprehensive analysis shows that only modest predictive gains are possible. Evidently, given that the GDP growth rate is persistent and co-moves with business conditions implies that lagged GDP growth rates provide a very good indicator of the economic state. Therefore, while certain crude oil price variables help predict the GDP growth rate in certain instance, leveraging these variables into large out-of-sample forecast improvements is very difficult.

Clearly, the question of how important these nonlinearites are deserve further study on other European countries. Finally, the findings in this study also have implications for economists interested in modeling the transmission of crude oil price shocks to the U.K economy. For example, a theoretical framework could be developed to rationalize the type of nonlinearity embodied in the three-year net crude oil price change or the three-year crude oil price gap.

Declaration of competing interest

None.

References

Akram, F. A., & Mumtaz, H. (2019). Time-varying dynamics of the Norwegian economy. *The Scandinavian Journal of Economics, 121*, 407–434.
Alquist, R., Kilian, L., & Vigfusson, R. J. (2013). Forecasting the price of oil. In A. Timmermann, & G. Elliott (Eds.), *Handbook of economic forecasting*. Amsterdam: Elsevier.
Andrews, D. W. K., & Monahan, J. C. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica, 60*, 953–966.
Bachmeier, L., Li, Q., & Liu, D. (2008). Should oil prices receive so much attention? An evaluation of the predictive power of oil prices for the U.S. economy. *Economic Inquiry, 46*, 528–539.
Bauwens, L., Koop, G., Korobilis, D., & Rombout, J. (2015). A comparison of forecasting models for macroeconomics series: The contribution of structural break models. *Journal of Applied Econometrics, 30*, 596–620.
Carlton, A. B. (2010). *Oil prices and real-time output growth*. University of Houston. Working Paper.
Cheng, I. H., & Xiong, W. (2014). Financialization of commodity markets. *Annual Review of Financial Economics, 6*, 419–441.
Clark, T. E., & McCracken, M. W. (2012). Advances in forecast evaluation. In A. Timmermann, & G. Elliott (Eds.), *Handbook of economic forecasting*. Amsterdam: Elsevier.
Clark, T. E., & West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics, 138*, 291–311.
Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics, 13*, 253–263.
Giacomini, R., & Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics, 25*, 595–620.
Goldberg, P. (1998). The effects of the corporate average fuel economy standards in the U.S. *The Journal of Industrial Economics, 46*, 1–33.
Gorton, G., & Rouwenhorst, K. (2006). Facts and fantasies about commodity futures. *Financial Analysts Journal, 62*, 47–68.
Groen, J. J. J., Richard, P., & Ravazzolo, F. (2013). Real-time inflation forecasting in a changing world. *Journal of Business & Economic Statistics, 31*, 29–44.
Hamilton, J. D. (2003). What is an oil shock? *Journal of Econometrics, 113*, 363–398.
Hamilton, J. D. (2011). Nonlinearities and the macroeconomic effects of oil prices. *Macroeconomic Dynamics, 15*, 472–497.
Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica, 79*, 453–497.
Harrison, R., Ryland, T., & de Weymarn, I. (2011). *The impact of permanent energy price shocks on the UK economy*. Bank of England. working paper 433.
Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting, 13*, 281–291.
Hong, H. G., & Stein, J. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of Finance, 54*, 2143–2148.
Hong, H. G., Torous, W., & Valkanov, R. (2007). Do industries lead stock markets? *Journal of Financial Economics, 83*, 367–396.
Jiménez-Rodríguez, R., & Sánchez, M. (2006). Oil price shocks and real GDP growth: Empirical evidence for some OECD countries. *Applied Economics, 37*, 201–228.
Kartsakli, M., & Adams, Z. (2017). *Has crude oil become a financial asset? Evidence from ten years of financialization*. University of St. Gallen. Working paper.
Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *The American Economic Review, 99*, 1053–1069.
Kilian, L., & Vigfusson, R. J. (2011a). Are the responses of the U.S. economy asymmetric in energy price increases and decreases? *Quantitative Economics, 2*, 419–453.
Kilian, L., & Vigfusson, R. J. (2011b). Nonlinearities in the oil price-output relationship. *Macroeconomic Dynamics, 15*, 337–363.
Kilian, L., & Vigfusson, R. J. (2013). Do oil prices help forecast U.S. Real GDP? The role of nonlinearities and asymmetries. *Journal of Business & Economic Statistics, 31*, 78–93.
Kilian, L., & Zhou, X. (2018). Modeling fluctuations in the global demand for commodities. *Journal of International Money and Finance, 88*, 54–78.
Kolodziej, M., Kaufmann, R. K., Kulatilaka, N., Bicchetti, D., & Maystre, N. (2014). Crude oil: Commodity or financial asset? *Energy Economics, 46*, 46216–46223.
Millard, S. (2011). *An estimated DSGE model of energy, costs and inflation in the United Kingdom*. Bank of England. working paper 432.
Millard, S., & Shakir, T. (2013). *Oil shocks and the UK economy: The changing nature of shocks and impact over time*. Bank of England. working paper 476.
Paye, B. S. (2012). Déja vol: Predictive regressions for aggregate stock market volatility using macroeconomic variables. *Journal of Financial Economics, 106*, 527–546.
Ravazzolo, F., & Rothman, P. (2013). Oil and U.S. GDP: A real-time out-of-sample examination. *Journal of Money, Credit, and Banking, 45*, 449–463.
Ravazzolo, F., & Rothman, P. (2016). Oil-price density forecasts of US GDP. *Studies in Nonlinear Dynamics & Econometrics, 20*, 441–453.
Reboredo, J. C., & Rivera-Castro, M. A. (2013). A wavelet decomposition approach to crude oil price and exchange rate dependence. *Economic Modelling, 32*, 3242–3257.
Romero-Meza, R., Coronado, S., & Serletis, A. (2014). Oil and the economy: A cross bicorrelation perspective. *The Journal of Economic Asymmetries, 11*, 91–95.
Rossi, B., & Inoue, A. (2012). Out-of-Sample forecast tests robust to the choice of window size. *Journal of Business & Economic Statistics, 30*, 432–453.
Rossi, B., & Sekhposyan, T. (2010). Have economic models' forecasting performance for US output growth and inflation changed over time, and when? *International Journal of Forecasting, 26*, 808–835.
Serletis, A., & Istiak, K. (2013). Is the oil price-output relation asymmetric? *The Journal of Economic Asymmetries, 10*, 10–20.
Stock, J. H., & Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics, 14*, 11–30.
Swanson, N. R. (1998). Money and output viewed through a rolling window. *Journal of Monetary Economics, 41*, 455–474.
Tang, K., & Xiong, W. (2012). Index investment and the. *Financialization of Commodities- Financial Analysts Journal, 68*, 54–74.
White, H. (2000). Reality check for data snooping. *Econometrica, 68*, 1097–1126.
Zhang, Y., Fan, Y., Tsai, H., & Wei, Y. (2008). Spillover effect of U.S. dollar exchange rate on oil prices. *Journal of Policy Modeling, 30*, 973–991.