New Findings Regarding the Out-of-Sample Predictive Impact of the Price of Crude Oil on the United States Industrial Production

Nima Nonejad\textsuperscript{1,2}

Received: 9 September 2021 / Accepted: 3 March 2022 / Published online: 25 March 2022
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

Abstract
Contrary to the extensive literature pioneered by James Hamilton in the early 1980s that focuses on analyzing the relationship between changes in the price of crude oil and the U.S. real gross domestic product growth (GDP) rate, Herrera et al. (2011) is essentially the first study that explores the \textit{in-sample} predictive impact of the price of crude oil on the U.S. industrial production index. To date, almost nothing is known about the nature and degree of the \textit{out-of-sample} predictive impact of the price of crude oil on the U.S. industrial production index. This study fills the gap. Using various nonlinear transformations of the price crude oil widely employed in the crude oil price/GDP predictability literature as well as crude oil price volatility measures, we document (rather surprisingly) that the form of nonlinearity that delivers the most consistent pattern of out-of-sample population-level predictability gains relative to the benchmark when forecasting ex-post revised as well as real-time U.S. industrial production has to do with crude oil price decreases below the minimum price in recent memory. In contrast to the GDP predictability literature, crude oil price increases beyond the maximum in recent memory do not afford any predictive power. On the contrary, they deteriorate relative forecast performance. These results go directly against a distinct sense of déjà vu that one would expect given the degree of affinity between industrial production and GDP. The predictive power afforded by crude oil price net decreases also translate into economic gains.

Keywords Crude oil price · Nonlinearity · Out-of-sample population-level predictability · Realized volatility · U.S. industrial production

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

\* Nima Nonejad
nimanonejad@gmail.com

1 Danske Bank and CREATEs, Copenhagen, Denmark
2 København K, Denmark
Introduction

Contrary to the extensive literature that explores the relationship between the price of crude oil and real gross domestic product (GDP) growth rate going back to well-known studies, such as Hamilton (1983), Mork (1989) and Hamilton (1996), Herrera et al. (2011) published in an issue of Macroeconomic Dynamics is essentially the first study that comprehensively focuses on analyzing the in-sample predictive impact of the price of crude oil on the U.S. industrial production index\(^1\). The authors argue that doing so is important because any eventual predictive impact is likely to be more prevalent on industrial production than on GDP. One reason is that contrary to GDP, industrial production is more closely linked to fluctuations in the demand for industrial commodities. Furthermore, over the years, the weight of the service sector in the real U.S. GDP has greatly increased at the expense of the industrial sector. Likewise, as mentioned in studies, such as Bresnahan and Ramey (1993) and Davis and Haltiwanger (2001), among others, if crude oil price innovations involve a costly reallocation of capital and labor, then concentrating on real aggregate GDP may obscure the nature of the reallocative effect.

The aim of Herrera et al. (2011)’s analysis can also be motivated by examining simple plots of movements in the price of crude oil and the U.S. industrial production index. For example, in panels (a) and (b) of Fig. 1, we display the first difference of the logarithm of the monthly U.S. industrial production index and the nominal West Texas Intermediate (WTI) price of crude oil from 1983\(^{m1}\) through 2020\(^{m12}\). We refer the reader to Section 3 for details on where the respective time-series are extracted from, and the choice of the time-span. Furthermore, to facilitate graphical comparison, the series are standardized prior to plotting. Note, the standardization is only performed in the context of making Fig. 1, and nowhere else in this study do we standardize any series. In both panels, the gray bars display the the U.S. business cycle using National Bureau of Economic Research (NBER) recession dates. Accordingly, both series co-move with the business cycle. In panel (a), the U.S. industrial production index tends to decrease on the onset of recessions and starts increasing as the U.S. economy recovers from the recession. In panel (b), we observe that variations in the price of crude oil tend to be of higher magnitude during recessions. At the same time, the degree of co-movement between the series and the business cycle is heterogeneous. For instance, both series respond more aggressively to the Great Recession and the recession in the early 2000s. Another interesting observation is that by comparing panel (a) with (b), we tend to observe changes

\(^1\) With regards to analyzing the nature of the in-sample predictive impact of the price of crude oil on the U.S. real GDP growth rate, we can refer the reader to studies, such as Mork (1989), Hamilton (1996), Hamilton (2003), Kilian (2009) and Hamilton (2011b). With regards to out-of-sample analysis, we can refer the reader to studies, such as Carlton (2010), Kilian and Vigfusson (2011), Kilian and Vigfusson (2013), Ravazzolo and Rothman (2013), Ravazzolo and Rothman (2016) and Nonejad (2020a).
in the U.S. industrial production index following movements in the price of crude oil. Therefore, to better assess the relationship between the series, we compute the cross-correlations as follows:

\[ r_{xy}(l) = \frac{\rho_{xy}(l)}{\sqrt{\rho_{xx}(0)\rho_{yy}(0)}} \]

where the subscript \( x \) denotes the first difference of the logarithm of the price of crude oil, \( y \) denotes the first difference of the logarithm of the U.S. industrial production index, and \( \rho_{xy}(l) \) is given as:

\[ \rho_{xy}(l) = \begin{cases} T^{-1} \sum_{t=1}^{T-l} (x_t - \bar{x})(y_{t+l} - \bar{y}) & \text{for } l = 0, 1, 2, \ldots \\ T^{-1} \sum_{t=1}^{T+l} (x_{t-l} - \bar{x})(y_t - \bar{y}) & \text{for } l = 0, -1, -2, \ldots \end{cases} \]

In the above formula, the bar denotes the mean of the variable of interest and \( T \) the sample size. If changes in the price of crude oil lead changes in the U.S. industrial production index, then we would expect the cross-correlations to be positive for positive \( l \)s. As evidenced from panel (c) of Fig. 1, we observe that this indeed appears to be the case. The correlation obtains its highest value at \( l = 1 \). We also conduct the cross-correlation analysis over time to explore whether \( r_{xy}(l) \) is subject to time-variation. Here, we set the window size equal to five years of monthly data, and estimate \( r_{xy}(l) \) over time. We generally find that the highest correlation occurs for positive \( l \)s, especially \( l = 1 \) and \( l = 2 \). In panel (d), we report the rolling window estimates of \( r_{xy}(l) \) with \( l = 1 \). As evidenced from panel (d), there is evidence of time-variation in the estimates of \( r_{xy}(l) \). Specifically, we observe an upward trend from the late 1990s until the mid 2000s. Likewise, we observe notable shifts in the estimate of \( r_{xy}(l) \) in 2009 and 2020. Overall, the simple analysis in Fig. 1 seems to support the notion that there should be a (time-varying) predictive impact from the price of crude oil on the U.S. industrial production.

The questions raised in Herrera et al. (2011) are (i): Does the price of crude oil have an in-sample predictive impact on the U.S. industrial production index? and (ii): Is the predictive impact nonlinear? Using monthly data from 1947m1 through 2009m12, the authors provide several interesting results. To begin with, similar to the literature that explores the in-sample predictive impact of the price of crude oil on the U.S. real GDP growth rate, slope-based tests a la Hamilton (2011b) and Kilian and Vigfusson (2011) provide evidence in favor of a nonlinear in-sample predictive impact. Here, the net crude oil price increase measure suggested in Hamilton (1996) is successful\(^2\). Hence, in a similar fashion as the GDP predictability literature, the authors establish that crude oil price increases beyond peaks in recent memory afford in-sample predictive power. Furthermore, the impulse response function-based tests of symmetry suggest evidence against the null hypothesis of

\(^2\) Hamilton (1996)’s reason for constructing this measure is the claim that what matters is not crude oil price increases themselves, but whether the crude oil price increases are big enough to reverse (offset) crude oil price decreases observed in the recent past.
symmetric responses of industrial production to crude oil price innovations. Not everything is positive though. Due to the distinct nature of crude oil price changes from 1947 through 1972, Herrera et al. (2011) find that results are highly sensitive to whether pre-1973 data are included in the sample or not. In fact, for the post-1973 period, the authors find no evidence against the null hypothesis that industrial production responds symmetrically to crude oil price innovations.

More than ten years after publication of Herrera et al. (2011), we believe that it is at its place to complement the study of Herrera et al. (2011) by quantifying the predictive impact of the price of crude oil on the U.S. industrial production index from an out-of-sample perspective. We believe that this analysis is important for several reasons.

---

3 As mentioned in studies, such as Hamilton (2011a), after the Great Depression and until the early 1970s, the price of crude oil was to a great degree determined by the Texas Railroad Commission and the Oklahoma Corporation Commission. This meant essentially that the price of crude oil during this period was largely not subject to sizable changes, and at most behaved like a step-function, see for example, the figures displayed in Narayan and Gupta (2015) and Nonejad (2019).

4 In a subsequent study, Herrera et al. (2015) conduct a similar analysis using industrial production index data for selected OECD countries containing both crude oil exporting and crude oil importing countries. Using post-1973 data, the authors find very little support for the hypothesis that the response of industrial production to crude oil price changes is asymmetric.

5 An out-of-sample framework recursively utilizes a subset of the available data to forecast values outside of the estimation period, and compares point (density) forecasts to the corresponding outcomes of the variable being predicted. An in-sample approach relies on all available data to perform estimation.
reasons: First, to our knowledge, Bachmeier et al. (2008) and Nonejad (2020b) are the only studies that (briefly) touch on this subject. Indeed, both studies consider industrial production as a variable among many others (such as inflation and interest rates) to be forecasted out-of-sample, and both studies focus exclusively on evaluating the evidence of predictability from a statistical viewpoint. Furthermore, Bachmeier et al. (2008) is more than ten years old, and naturally fails to incorporate the more recent suggested nonlinear crude oil price (crude oil price volatility) measures. Lastly, in a recent study, Nonejad (2021) evaluates the out-of-sample predictive impact of the price of crude oil on the world industrial production index favored by Baumeister and Hamilton (2019). However, as we shall demonstrate, the degree and nature of the predictive impact of the price of crude oil on the U.S. industrial production index is different, and one can argue rather unique. For example, contrary to Nonejad (2021), there is evidence in favor of notable (unconditional) predictive impact from certain nonlinear crude oil price measures on the U.S. industrial production index growth rate one-month ahead.

Second, as argued in Kilian and Vigfusson (2011) among others, outcomes of in-sample tests, such as the slope-based tests performed in Herrera et al. (2011) and Herrera et al. (2015) or likewise impulse response analysis do not necessarily need to translate into out-of-sample relative forecast improvements, which is the ultimate question of interest to policy makers and applied forecasters. In other words, these tests cannot answer whether by conditioning on the price of crude oil today one can more accurately forecast the variable of interest relative to the benchmark.

Third, in their study, Herrera et al. (2011) rely on only two nonlinear transformations of the price of crude oil, namely, Mork’s (1989) crude oil price increase and the net crude oil price increase suggested in Hamilton (1996). The former is defined as a vector containing positive crude oil price changes, otherwise zero. The latter is defined as the positive gap between the price today and the maximum price over the last \( m \) periods. However, since Herrera et al. (2011), several new nonlinear crude oil price measures have been introduced in the literature, see Kilian and Vigfusson (2013). Hence, it is now possible to go beyond the basic nonlinear measures. Indeed, it is probable that similar to results reported in Kilian and Vigfusson (2013), the predictive power afforded by these new measures, such as the symmetric net crude oil price change or crude oil price gap are stronger than Mork’s (1989) crude oil price increase or the net crude oil price increase. Therefore, it is important to perform the analysis using these new nonlinear crude oil price-based measures.

Fourth, recent studies, such as Nonejad (2020a) and Nonejad (2020b) find evidence in favor of using crude oil price volatility as opposed to the price of crude

---

6 We also have to mention that Nonejad (2021) relies on evaluating the evidence of conditional predictability, whereas in this study, we evaluate the evidence of unconditional population-level predictability.

7 In their study that focuses on forecasting the U.S. real GDP growth rate out-of-sample by conditioning on the price of crude oil, Kilian and Vigfusson (2013) find that the net crude oil price change measure, which is symmetric in net crude oil price increases and decreases outperforms the net crude oil price increase measure suggested in Hamilton (1996).
oil to forecast macroeconomic variables out-of-sample. For example, using ex-post revised as well as real-time data, Nonejad (2020a) finds that crude oil price volatility affords more predictive power than the price of crude oil when forecasting the U.S. GDP growth rate one-quarter ahead. This occurs in terms of population-level predictability as well as finite-sample accuracy. The large out-of-sample forecast evaluation conducted in Nonejad (2020b) finds that for many macroeconomic series, the null hypothesis of no population-level (finite-sample) predictability is more often rejected for models employing crude oil price volatility than the price of crude oil. In both studies, out-of-sample results are particularly encouraging for crude oil price semivolatilities, which distinguish between volatility increases due to crude oil price increases and volatility increases due to crude oil price decreases. Therefore, in our analysis, we also include crude oil price volatility measures. Doing so, we can better explore whether the out-of-sample predictive impact is nonlinear in crude oil price changes but linear in crude oil price volatility, vice versa or maybe even linear in both.

To sum up, the contribution of this study is to conduct a comprehensive out-of-sample forecasting analysis, and evaluate the predictive power afforded by the price of crude oil on ex-post revised as well as real-time U.S. industrial production data using statistical as well as economic criteria. Doing so, we establish several interesting results that are important not only to forecasters, but also those who are interested in understanding the nature of the transmission of crude oil price shocks on industrial production.

The most interesting result established in this study is that contrary to intuition and results documented in the crude oil/GDP predictability literature, the form of nonlinearity that affords the most consistent pattern of out-of-sample population-level predictive gains has to do not with crude oil price increases beyond recent peaks or net crude oil price changes, but crude oil price decreases below recent lows. At the one-month ahead forecast horizon, the relative reduction in mean square error (MSE) produced under the model with the one-year net crude oil price decrease relative to the benchmark can be as high as 9%. We also observe statistically significant MSE reductions as we increase the forecast horizon even though at forecast horizons beyond one month results are sensitive to whether the COVID 19 pandemic is included in the out-of-sample period or not. In contrast, linear models fail to deliver consistent relative forecast improvements. Compared to the linear model with the first difference of the logarithm of the price of crude oil, the predictive model with crude oil price log-realized volatility performs better over the 1990m1–2019m12 out-of-sample period, producing accurate point forecasts as the model with the one-year net crude oil price decrease at longer forecast horizons. For the other nonlinear models, we find mixed evidence of relative forecast improvements at the population level. At the one-month ahead horizon, these nonlinear measures perform better than the benchmark. However, this is not the case at longer

---

8 Herrera et al. (2011) do not consider the in-sample predictive impact of crude oil price volatility on industrial production.

9 Interestingly, in their study, Herrera et al. (2011) did not consider this form of nonlinearity.
forecast horizons. Lastly, we observe that the evidence of predictability from a statistical viewpoint also translates into economic gains. Compared to the benchmark and other competitors, point forecasts produced under the predictive model with the one-year net crude oil price decrease help produce an earlier and stronger signal of a downturn in industrial production on the onset of the Great Recession and COVID-19.

The main findings of this study survive an array of robustness checks. These include using alternative measures of the price of crude oil, experimenting with increasing the truncation lag order of the crude oil price measures, such as the net crude oil price increase (decrease), and comparing the accuracy of point forecasts produced under the nonlinear crude oil price model with point forecasts produced under the Markov-switching regression suggested in Hamilton (1989), which is considered as a “traditional” way of accounting for nonlinearities.

The extensive literature on predicting the U.S. real GDP growth rate out-of-sample by conditioning on the price of crude oil generally agrees that crude oil price net increases and crude oil price net changes tend to afford the most consistent pattern of out-of-sample forecast gains relative to the benchmark, see Kilian and Vigfusson (2013) and Ravazzolo and Rothman (2013) among others. However, these studies have almost never explored the predictive power afforded by crude oil price net decreases. The same holds true for Herrera et al. (2011) and Herrera et al. (2015), where the authors focus only on evaluating the in-sample predictive power of the price of crude oil on industrial production.

Given the degree of affinity between the real U.S. GDP growth rate and U.S. industrial production index growth rate (see footnote 12), one would expect that the same nonlinear crude oil price measures that help forecast the U.S. real GDP growth rate out-of-sample also help forecast the U.S. industrial production index growth rate out-of-sample. The contribution of this study is to show that this is not the case, and contrary to what one is inclined to assume a priori, more robust and accurate industrial production index growth rate point forecasts are obtained by conditioning on net crude oil price decreases. Thus, our results go directly against a distinct sense of déjà vu that one would expect given the degree of affinity between industrial production and real GDP.

The rest of this study is organized as follows: The econometric framework is detailed in Section 2. Section 3 describes the data used in the analysis. Out-of-sample results are presented and discussed in Section 4. In Section 5, we present several robustness checks. Finally, Section 6 concludes.

10 Nonejad (2020c) finds similar conclusions when using U.K. real GDP data.
11 In studies, such as Ravazzolo and Rothman (2016), Nonejad (2020a), Nonejad (2020c) and Nonejad (2021), where the net crude oil price decrease has been considered, the relative predictive power afforded by this measure has been somewhat weak.
12 The correlation between the series is very high, around 0.8.
2 Econometric Framework

Similar to the mainstream literature that uses the price of crude oil to predict macro-economic variables out-of-sample, our economic framework is simple, and relies on the ordinary least squares (OLS) technique for estimation. Essentially, we consider predictive regressions, where the variable being forecasted is the percentage change in the U.S. industrial production index, and the explanatory variables include the lagged values of the dependent variable as well as lagged values of of the log-price of crude oil, possibly with restrictions imposed on them. Following studies, such as Bachmeier et al. (2008), Hamilton (2011b), Ravazzolo and Rothman (2013), Ravazzolo and Rothman (2016), Nonejad (2020b) and Nonejad (2021) among others, our crude oil price (crude oil price volatility)-based predictive regression model is given as:

\[ y_t = \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} X_{t-i} \beta_i + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2), \]  

(2.1)

where \( y_t \) denotes the first difference of the logarithm of the U.S. industrial production index at time, \( t = 1, ..., T \), and \( X_{t-i} \) is the possibly vector-valued crude oil price (crude oil price volatility) measure of interest at time \( t - i \) for \( i = 1, ..., q \). The autoregressive benchmark model is in turn, given as:

\[ y_t = \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2). \]  

(2.2)

The out-of-sample analysis divides the full-sample data into two subsets. The first \( r \) observations constitute the in-sample period, used for initial model estimation, and the final \( n \) observations constitute the out-of-sample, which we use to evaluate model performance. Point forecasts over the out-of-sample period are recursively constructed\(^{13}\). Here, models are re-estimated each time new observations enter the information set. Once this is done, we compare the accuracy of point forecasts produced under (2.1) with the crude oil price measure of interest relative to (2.2) using the mean square error (MSE) criterion.

Evidently, in order to produce these point forecasts, we must select the lag orders in (2.1) and (2.2). Here, we follow Kilian and Vigfusson (2013), and at each forecast horizon, \( h \), determine the most appropriate lag order for the model of interest based on a forecast accuracy comparison involving all lag orders starting from one to twelve. Furthermore, when forecasting \( h \)-months ahead for \( h > 1 \), we rely on the direct method of forecasting, see Marcellino et al. (2006) for details\(^{14}\).

\(^{13}\) Similar to the mainstream literature, point forecasts are produced using the expanding window estimation approach.

\(^{14}\) An alternative to the direct method of forecasting is the iterated method. For the latter, one considers a one-period model, and iterate it forward for the desired number of periods. On the other hand, direct forecasts are produced using a horizon-specific model, where the dependent variable is the multi-period ahead value being forecasted. An advantage of the latter is that we can avoid using multi-equation systems, see also Ravazzolo and Rothman (2013). Furthermore, in our analysis, the direct forecasting...
Finally, it is important to note that contrary to equation 4 of Herrera et al. (2011) and Kilian and Vigfusson (2013), we do not insert different crude oil price variables simultaneously in (2.1), but instead employ them individually. The is because in a similar fashion as Hamilton (2011b), we observe that parsimony is crucial in the context of out-of-sample analysis, see Section 5. Particularly, Hamilton (2011b) argues that while omitting the percentage change in the price of crude oil from the predictive regression containing a nonlinear crude oil price measure, see for example equation 4 of Herrera et al. (2011) might result in misspecification with regards to in-sample tests, it is nonetheless preferred with regards to out-of-sample forecasting analysis given its greater parsimony. This approach also enables us to better distinguish the predictive power afforded by one model specification relative to another.

2.1 Testing for Out-of-Sample Granger Causality

As mentioned in Section 1, the aim of our study is to evaluate whether the out-of-sample predictive impact of the price of crude oil on industrial production is nonlinear. However, there are multiple potential interpretations of the notion of out-of-sample predictability. This study focuses on analyzing whether there is a predictive impact at the population level. In the literature, this is also referred to as out-of-sample Granger causality. This is effectively equivalent to testing the null hypothesis that \( \beta_1, \ldots, \beta_q \) in (2.1) are jointly equal to zero, see also Ravazzolo and Rothman (2013). In a similar fashion as Ravazzolo and Rothman (2013) among others, if we cannot reject the no out-of-sample population-level predictability null hypothesis for the linear models but do so for certain nonlinear models, then we conclude that the predictive impact is nonlinear. Compared to evaluating the evidence of finite-sample forecast accuracy, this approach better parallels the in-sample analysis of Herrera et al. (2011).

We use the Clark and West (2007) test to evaluate the evidence of out-of-sample population-level predictability. The null hypothesis of this test is equal accuracy of the point forecasts at the population values of the model parameters. The idea of the test is the following: Under the null hypothesis of no population-level predictability, the benchmark model generates the data. Therefore, the population mean square error (MSE) produced under the benchmark is smaller than the population MSE.

Footnote 14 (continued)

approach is the optimal solution because we rely on censored crude oil price variables, and specifying VAR-type models using censored series to perform iterative forecasting can result in misspecification, see Kilian and Vigfusson (2011).

15 It also goes without saying that contrary to Herrera et al. (2011), we do not incorporate contemporaneous crude oil price regressors in (2.1) because we are performing out-of-sample forecasting, i.e. forecasting \( y_{t+1} \) given information up to time \( t \).

16 Testing for population-level predictability is not the same as testing for finite-sample equal predictive ability often performed using the well-known test suggested in Diebold and Mariano (1995). The latter focuses on testing the null hypothesis of equal MSE given the data at hand, whereas the former focuses on population parameter values. We can refer the reader to Inoue and Kilian (2004) and Clark and McCracken (2013) for details regarding the different interpretations.
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 \( \hat{y}_{t+1}^{(b)} \) and \( \hat{y}_{t+1}^{(l)} \) denote the one-month ahead point forecast of \( y_{t+1} \) produced under the benchmark and model \( l \), respectively. The Clark and West (2007) test statistic is given as:

\[
CW = n^{-1} \sum_{j=t+1}^{T} \left( y_j - \hat{y}_j^{(b)} \right)^2 - n^{-1} \sum_{j=t+1}^{T} \left( y_j - \hat{y}_j^{(l)} \right)^2 + n^{-1} \sum_{j=t+1}^{T} \left( \hat{y}_j^{(b)} - \hat{y}_j^{(l)} \right)^2.
\]

(2.3)

In (2.3), \( n^{-1} \sum_{j=t+1}^{T} \left( y_j - \hat{y}_j^{(b)} \right)^2 \) and \( n^{-1} \sum_{j=t+1}^{T} \left( y_j - \hat{y}_j^{(l)} \right)^2 \), are the MSEs produced under the benchmark and model \( l \), respectively. The last term, \( n^{-1} \sum_{j=t+1}^{T} \left( \hat{y}_j^{(b)} - \hat{y}_j^{(l)} \right)^2 \), captures the adjustment for additional noise associated with model \( l \)’s forecasts. 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 (2.3) is therefore also large. Following Clark and West (2007), we compare (2.3) with the one-sided critical values from a standard Normal distribution\(^{17}\). 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.

### 3 Data

We rely on ex-post revised as well as real-time U.S. industrial production index data. The former is extracted from the FRED database: [https://fred.stlouisfed.org/](https://fred.stlouisfed.org/). The real-time data is extracted from the Philadelphia Federal Reserve Bank’s real-time database: [https://www.phillyfed.org/](https://www.phillyfed.org/).

Following results reported in Alquist et al. (2013) among others, there has been a gradual consensus regarding the choice of the crude oil price series, and the time span. The authors advise against using crude oil price data from 1947m1 to 1972m12. The reason is that during this period, the price of crude oil varies very little, and exhibits a pattern resembling a step-function\(^{18}\). As demonstrated by Alquist

---

\(^{17}\) We compute (2.3) by regressing 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. As a robustness check, we also compute the bootstrapped p-values. However, we find that the conclusions are identical to using p-values using the Normal distribution. Likewise, we perform a Monte Carlo analysis to evaluate the size and power of the Clark and West (2007) test under predictive regressions, such as (2.1), where the predictors can be censored. Overall, we find that the Clark and West (2007) test has very good power and is also well-sized.

\(^{18}\) As we mentioned in footnote 2, the discrete pattern of crude oil price changes during this time-period is explained by the specific regulatory structure imposed on the oil industry from 1947 to 1972. The early 1970s sees notable increases in the price of crude oil due to geopolitical tensions in the Middle East and decisions by OPEC. Likewise, as the regulatory controls begin to weaken in the early 1980s, changes in the price of crude oil becomes much more frequent.
et al. (2013), deflating the nominal price by the CPI index still does not solve the problem. The authors also argue that it is inappropriate to combine data before 1973 with data after 1973\textsuperscript{19}. Besides the choice of the time span, Alquist et al. (2013) argue that the West Texas Intermediate (WTI) price of crude oil price \textit{may not} be an accurate measure of the price faced by oil U.S. refiners because the market is regulated well up to the early 1980s\textsuperscript{20}. Instead they argue that U.S. refiners’ acquisition cost (RAC) for imported crude oil is more relevant for analyzing U.S. data. However, contrary to WTI or Brent Blend prices, RAC data are available only at the monthly sampling frequency starting from 1974 and onwards\textsuperscript{21}.

The scope of our study puts some limits on the choice and time span of the crude oil price series. Particularly, as mentioned in Section 1, following results reported in Nonejad (2020a) and Nonejad (2020b), we also explore the out-of-sample predictive impact of crude oil price volatility on the U.S. industrial production index. However, we cannot directly observe volatility, and must therefore measure it. A convenient approach that allows us to use predictive regressions, such as (2.1) relies on an ex-post measurement approach of crude oil price volatility. More precisely, in a similar fashion as Bachmeier et al. (2008), Nonejad (2020a) and Nonejad (2020b), we sum squared daily crude oil price returns to construct an ex-post measure for monthly crude oil price volatility as:

$$RV_t = \frac{\tilde{n}}{n} \sum_{i=1}^{\tilde{n}} r_{i,t}^2,$$

where $\tilde{n}$ denotes the number of trading days in the $t$th month, and $r_{i,t}$ is the day $i$ return on the price of crude oil\textsuperscript{22}. The advantage of this approach is that we can directly use the logarithm of (3.1) as predictors in (2.1)\textsuperscript{23}. Accordingly, to compute (3.1), we need daily crude oil prices. However, RAC prices are not available at the daily sampling frequency. Likewise, since we prefer to have a long historical sample implies that we cannot Brent Blend prices because daily Brent Blend prices are

---

\textsuperscript{19} Dvir and Rogoff (2010) and Alquist et al. (2013) find a structural break in the crude oil price series in 1973.

\textsuperscript{20} WTI refers to a particular grade of crude oil to be delivered to Cushing, Oklahoma. It is priced in U.S. dollars, and is available at the daily sampling frequency from the early 1970s. At the monthly sampling frequency, WTI prices are available going back to 1859\textsuperscript{m10}. This long data has been used in various studies, see for example, Narayan and Gupta (2015).

\textsuperscript{21} Brent Blend is a combination of crude oil from fifteen different oil fields in the North Sea. It is priced in U.S. dollars, and is available at the daily sampling frequency from the late 1980s.

\textsuperscript{22} Evidently, there is a great degree of affinity between (2.1) and the realized volatility literature, employing intraday returns to measure daily volatility, see Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2002) among others.

\textsuperscript{23} Parametric volatility models, such as GARCH or stochastic volatility (SV) could also be used to estimate the conditional volatility of the price of crude oil. However, as argued in Alquist et al. (2013) among others, compared to realized volatility, GARCH or SV generated volatility is more backward-looking and thus a less representative measure of volatility. Likewise, the authors show that conditional volatility processes generated under GARCH and SV models tend to quickly revert to their time-invariant unconditional expectations as the forecasting horizon increases, which in turn, implies that they are not suited to quantify changes in the long-run expected volatility.
only available from the late 1980s. Therefore, we can only rely on the WTI price of crude oil to construct (3.1). In a similar fashion as Nonejad (2020a) and Nonejad (2020b), we extract daily WTI prices from Global Financial Data: https://globalfinancialdata.com/. Furthermore, daily WTI prices are not subject to revision, which makes computing realized volatility measures easier. However, while daily WTI crude oil prices are available from the early 1970s, there are several instances from 1975m1 through 1982m12, where prices are constant across relatively long periods, implying that we estimate (3.1) at zero for certain months. Therefore, in a similar fashion as Bachmeier et al. (2008), Nonejad (2020a) and Nonejad (2020b), our realized volatility series starts at 1983m1.

The final point of discussion regards whether we should use real or the nominal price of crude oil. In the literature, there is no definitive empirical evidence in favor of one specification relative to another. Studies, such as Kilian and Vigfusson (2011) favor using the real price of crude oil, whereas Hamilton (2011b) favors the nominal price of crude oil, especially in forecasting applications. With regards to (3.1), it is well-known that the concept of volatility is based on nominal returns and not real returns. Furthermore, even setting aside this issue, the CPI index cannot be observed at the daily sampling frequency, which complicates things when computing the realized volatility measures. Nevertheless, to bring completeness to our analysis, we also experiment with forecasting industrial production by conditioning on the real price of crude oil and using RAC prices instead of WTI, see Section 5 for details.

4 Out-of-Sample Analysis

This section presents the out-of-sample results. More precisely, in Section 4.1, we provide details regarding how the analysis is carried out. Results using the statistical criterion discussed in Section 2.1 are presented in Section 4.2, whereas out-of-sample evaluation using the economic criterion are presented in Section 4.3.

24 We want that our out-of-sample period starts at 1990m1, such that we parallel the mainstream literature. However, daily Brent Blend prices are only available from June 1987, meaning that after accounting for the necessary number of lags in (2.1), namely, 12, we have less than 20 monthly observations in our in-sample period, which is not enough to perform estimation.

25 For example, from 08/03/1981 to 12/31/1981, we measure the price of crude oil at $35 a barrel each day.

26 Hamilton (2011b) argues that real price induces measurement errors, which can ultimately have implications on forecasting performance. Likewise, the argument for using the nonlinear crude oil price variables suggested in Mork (1989), Hamilton (1996) and Kilian and Vigfusson (2013) is mainly behavioral. Hence, Hamilton (2011b) states that using nominal nonlinear crude oil price measures appear just as reasonable as a real specifications.

27 To construct realized volatility using the real price of crude oil, we could follow Nonejad (2020b) and build a daily CPI index through interpolation. However, similar to Nonejad (2020b), we observe very similar results as using the nominal price.
4.1 Study Design

Given the discussion in the previous section, we make the following decisions. To begin with, we rely on data from 1983 through 2020. Furthermore, we report results using nominal WTI prices. However, later on, we also experiment with other crude oil price series. We use data up to 1989 as the in-sample period, and in a similar fashion as the mainstream literature start the out-of-sample period in 1990.

For each model, we generate forecasts, where . The optimal lag orders in (2.1) and (2.2), namely, and , are chosen as explained in Section 2. The maximum lag lengths are fixed at max( ) = max( ) = 12. The out-of-sample performance of our models is evaluated across three out-of-sample periods, namely, 1990–2020, 1990–2019, and 1990–2007.

Inclusion of the latter two out-of-sample periods allows us to evaluate the sensitivity of results with respect to COVID-19 and the Great Recession, respectively.

Finally, we perform the analysis using ex-post revised as well as real-time industrial production data. However, emphasize that from a statistical viewpoint, complications can arise when evaluating real-time out-of-sample forecasts due to different degrees of data revision across forecast origins. Therefore, to circumvent this problem, we follow Ravazzolo and Rothman (2013), and employ Koenig et al. (2003)’s “strategy 1” for estimation of the predictive regressions. Clark and McCracken (2013) note that under this estimation approach, predictability tests developed for the case of non-revised data, such as Clark and West (2007) can be applied.

4.2 Results

We start by evaluating the out-of-sample predictive power afforded by our linear crude oil price measures. Here, with the term “linear”, we mean that a crude oil price (crude oil price volatility) decrease at time has the same impact on industrial production at time as a crude oil price (crude oil price volatility) increase. Particularly, we consider lagged values of (i): The percentage change in the price of crude oil, computed as the first difference of log-crude oil price: , where is the log-crude oil price at month , and (ii): The logarithm of (3.1) as predictors in (2.1), respectively. In Table 1, we report out-of-sample results for these specifications across eight forecast horizons. In the table, TU (Theil’s U), denotes the MSE produced under the predictive model of interest relative to the MSE produced under the benchmark, and CW denotes the p-values associated with the Clark and West (2007) population-level predictability test. As evidenced by results reported in the table, we cannot observe a uniform pattern of

---

28 We do not optimize the performance of any one predictive model by modifying these settings, in recognition of the concern over data mining discussed extensively in Rossi and Inoue (2012).

29 This strategy consists of using first-release data on the left-hand side and real-time data on the right-hand side.
predictive gains relative to the benchmark model. The former performs best for ex-post revised data at \( h = 1 \) over 1990m1–2020m12 and 1990m1–2019m12 out-of-sample periods, whereas we do not observe any gains over 1990m1–2007m12. Here, the reduction in MSE relative to the benchmark is 6% (2%) and statistically significant at the 10% level over the 1990m1–2020m12 (1990m1–2019m12) out-of-sample period. The gains decrease by 3 percentage points when we use real-time data. The predictive model with log-realized volatility delivers statistically significant forecast improvements mainly over 1990m1–2019m12 for \( h = 1, 2 \) and 3 for ex-post revised and real-time data.

Next, we provide results for the nonlinear measures. The first two measures are Mork’s increase and decrease defined as: 
\[
mork^+ = \max(0, \text{oil}_t - \text{oil}_{t-1})
\]
and
\[
mork^- = \min(0, \text{oil}_t - \text{oil}_{t-1})
\]
respectively. Particularly, in an interesting study using data up to the mid 1980s, Mork (1989) finds that while crude oil price increases tended to precede economic recessions, one could not reject the null hypothesis that crude oil price declines do not lead to expansions. We have included \( \text{mork}^- \) in our analysis because it is a natural counterpart to \( \text{mork}^+ \). It is important to note that these measures involve asymmetries rather than nonlinearities. As displayed in Table 2, the form of asymmetry contained in \( \text{mork}^+ \) does not help obtaining forecast improvements relative to the benchmark. Here, the TUs are consistently close or higher than one. In contrast, the asymmetry contained in \( \text{mork}^- \) does relatively better, delivering statistically significant population-level predictability gains over the 1990m1–2019m12 out-of-sample period across various forecast horizons for both ex-post revised and real-time industrial production data.

In a study published in the mid 1990s, Hamilton (1996) extends the measure suggested in Mork (1989). Particularly, Hamilton (1996) suggests the \( m \)-month net crude oil price increase defined as:
\[
\text{net}^+_t = \max(0, \text{oil}_t - \text{oil}_{t-1})
\]
where
\[
\text{oil}_t = \max(\text{oil}_{t-1}, ..., \text{oil}_{t-m})
\]
This measure contains two distinct types of nonlinearities. The first nonlinearity, which is symmetric reflects the perception that the impact of the price of crude oil is perceived differently, depending on how much it differs from recent historical experience. The second nonlinearity arises from the restriction that only net increases in the price of crude oil matter. This restriction imposes an asymmetry. It is important to mention that Hamilton (1996)’s motivation behind this measure is on the basis of (untested) behavioral arguments rather than economic theory. Nevertheless, this measure has been employed in various studies, including Herrera et al. (2011). Evidently, a natural counterpart to \( \text{net}^+ \) is 
\[
\text{net}^-_t = \min(0, \text{oil}_t - \text{oil}_{t-1})
\]
where
\[
\text{oil}_t = \min(\text{oil}_{t-1}, ..., \text{oil}_{t-m})
\]
In Table 3, we report out-of-sample results using lagged values of \( \text{net}^+ \) and \( \text{net}^- \) as predictors in (2.1), respectively. Here, following Herrera et al. (2011), we set \( m = 12 \), such that \( \text{net}^+ \) corresponds to the one-year net crude oil price increase and so on. In Section 5, we also report results for \( m = 36 \), such that \( \text{net}^- \) corresponds to the three-year net crude oil price decrease. Interestingly, contrary to studies that have successfully used \( \text{net}^+ \) to forecast the real GDP growth rate, we observe that employing \( \text{net}^+ \) in (2.1) does not lead to any MSE gains worth mentioning when forecasting the U.S. industrial production index growth rate out-of-sample. In fact, all the TUs are above one. In contrast, sizable and statistically significant relative point forecast improvements are obtained using the one-year net crude oil price decrease, \( \text{net}^- \). Indeed, it is the only
specification thus far, where the no population-level predictability null hypothesis is rejected across all out-of-sample periods and data at the one-month ahead horizon. Here, we obtain MSE reductions as high as 9% relative to the benchmark over 1990m1–2019m12 and 1990m1–2020m12, whereas over 1990m1–2007m12, the relative gains in MSE are minor. As we increase $h$, we observe statistically significant population-level predictability gains only over the 1990m1–2019m12 out-of-sample period for $h = 2, 3$ and 4. The magnitude of the MSE gains decrease as $h$ increases.

Recently, Kilian and Vigfusson (2013) suggest several new crude oil price measures. For instance, it is not immediately evident whether the second form of nonlinearity contained in $net^+ (net^-)$ is more important than the first. However, one can easily answer this question by comparing out-of-sample point forecasts produced under (2.1) with $net^+ (net^-)$ to a specification that does not involve asymmetry, but accounts for net crude oil price deviations, namely, the $m$-month net change in the price of crude oil, $net = net^+ + net^-$. The $m$-month asymmetric net crude oil price

In their study using U.S. GDP data, Kilian and Vigfusson (2013) establish that the form of nonlinearity that matters the most for obtaining point forecast accuracy gains relative to the benchmark has indeed to do with symmetric price increases and decreases relative to highs and lows in recent memory, i.e. net
change, which compared to net contains a weaker form of asymmetry is defined as: \( \text{anet} = \frac{\text{net}^+}{\text{net}^+ + \text{net}^-} \). Contrary to net, the \( m \)-month net crude oil price increase and the \( m \)-month net crude oil price decrease enter anet with different weights. In Table 4, we report results for these two measures. They both perform well at \( h = 1 \). However, we do not observe the same degree of MSE reductions as we did for net\(^-\), and the evidence of predictability vanishes as we increase the forecast horizon. Furthermore, as somewhat expected, (2.1) with anet delivers point forecasts closest to (2.1) with net\(^-\). This is understandable because contrary to net, anet allows us to put different weights on net\(^+\) and net\(^-\), respectively. Another important finding from Tables 3 and 4 is that by comparing the performance of (2.1) with net\(^-\) with (2.1)
with \( \text{net} \), we find that what drives the performance of the former is the second form of nonlinearity and not the first.

Besides \( \text{net} \) and \( \text{anet} \), Kilian and Vigfusson (2013) suggest three additional measures that focus on crude oil price deviations from recent heights, and large crude oil price increases (changes). First, they suggest the \( \text{m-th} \)-month crude oil price gap given as:

\[
gap_t = \text{oil}_t - \dot{\text{oil}}_t.
\]

Here, we have essentially replaced \( \text{net}^+ \) with its uncensored counterpart. This measure is constructed based on the conjecture that what matters for obtaining forecast improvements has not to do with net increases, but simply the deviation of the price of oil from the highest price in recent memory. Second, they suggest the \( \text{m-th} \)-month large crude oil price change,

\[
large_t = \Delta \text{oil}_t \cdot 1(\Delta \text{oil}_t > \text{std dev}(\Delta \text{oil}_{t-1}, \ldots, \Delta \text{oil}_{t-m})),
\]

where \( 1(A) \) equals 1 if \( A \) is true else 0, and the \( \text{m-th} \)-month large crude oil price increase, \( large^+_t = \Delta \text{oil}_t \cdot 1(\Delta \text{oil}_t > \text{std dev}(\Delta \text{oil}_{t-1}, \ldots, \Delta \text{oil}_{t-m})). \) These measures are constructed based on the assumption that the predictive impact from the price of crude oil on the variable being predicted is limited to large changes (increases). In Table 5,
we report results for these measures. Overall, we observe that they largely fail to deliver TUs less than one for $h > 1$. At $h = 1$, even the best performing specification among them is outperform by net− by anywhere between 2% to 4%.

The final nonlinear measure evaluates the nonlinear predictive impact of crude oil price volatility on the industrial production index. As displayed by (3.1), squaring daily crude oil price returns (or even taking the absolute value of them) eliminates any information contained in their sign. Therefore, to be able to quantify to what degree volatility increases caused by price increases drive forecast performance relative to volatility increases due to price decreases, we follow Patton and Sheppard (2015), and consider the realized semivolatility estimators given as:

$$RV_t^+ = \sum_{i=1}^{\tilde{n}} r_{i,t}^2 1(r_{i,t} > 0) \text{ and } RV_t^- = \sum_{i=1}^{\tilde{n}} r_{i,t}^2 1(r_{i,t} < 0). \quad (4.1)$$

As illustrated in Patton and Sheppard (2015), these estimators provide a complete decomposition of $RV_t$, in that $RV_t = RV_t^+ + RV_t^-$. In our out-of-sample analysis,

| Table 4 | Population-level predictability evaluation results using measures suggested in Kilian and Vigfusson (2013) |
|---------|-----------------------------------------------------------------------------------------------------|
| IP data | Ex-post revised                                                                                   |
|         | Out-of-sample                                                                                     |
|         | 1990m1–2012                                                                                       |
|         | 1990m1–2019                                                                                       |
|         | 1990m1–2007                                                                                       |
|         | Real-time                                                                                         |
|         | 1990m1–2012                                                                                       |
|         | 1990m1–2019                                                                                       |
|         | 1990m1–2007                                                                                       |
| $h$     | TU    | CW    | TU    | CW    | TU    | CW    | TU    | CW    | TU    | CW    | TU    | CW    |
| 1       | 0.97  | 0.09  | 0.96  | 0.00  | 0.99  | 0.02  | 1.01  | 0.30  | 0.96  | 0.00  | 0.98  | 0.01  |
| 2       | 1.01  | 0.80  | 1.00  | 0.43  | 1.00  | 0.48  | 1.02  | 0.77  | 1.00  | 0.17  | 1.02  | 0.81  |
| 3       | 0.99  | 0.25  | 1.01  | 0.57  | 1.00  | 0.29  | 1.01  | 0.26  | 1.01  | 0.60  | 1.02  | 0.87  |
| 4       | 1.00  | 0.26  | 1.01  | 0.71  | 1.00  | 0.81  | 1.01  | 0.61  | 1.00  | 0.17  | 1.02  | 0.80  |
| 5       | 1.00  | 0.71  | 1.01  | 0.71  | 1.00  | 0.74  | 1.00  | 0.31  | 1.01  | 0.52  | 1.01  | 0.76  |
| 6       | 1.00  | 0.18  | 1.01  | 0.53  | 1.01  | 0.85  | 1.00  | 0.30  | 1.00  | 0.63  | 1.00  | 0.62  |
| 7       | 1.00  | 0.32  | 1.00  | 0.61  | 1.00  | 0.80  | 1.00  | 0.87  | 1.01  | 0.86  | 1.00  | 0.77  |
| 8       | 1.00  | 0.11  | 1.00  | 0.24  | 1.00  | 0.90  | 1.00  | 0.23  | 1.00  | 0.66  | 1.00  | 0.54  |

In the table, TU reports the MSE produced under the specification of interest relative to the MSE produced the benchmark. CW reports p-values associated with the null hypothesis of no population-level predictability as specified in Clark and West (2007).
lagged values of the logarithm of $RV_t^+$ and $RV_t^-$ enter the predictive model with different weights, thereby accounting for nonlinearity. In Table 6, we report out-of-sample results for (2.1) using lagged values of the logarithm of (4.1) as predictors. As illustrated in the top part of the table, comparing results produced under (2.1) with (4.1) with the linear counterpart does not lead to any major changes worth mentioning. Furthermore, compared to the more successful nonlinear crude oil price specifications, we fail to reject the no population-level predictability null hypothesis.
over the 1990m1–2020m12 out-of-sample period. In the middle and bottom part of the table, we report results produced under predictive regressions, where we incorporate \(RV^+_t\) and \(RV^-_t\) separately. However, compared to (4.1), we do not find any improvements.

As our final analysis in the section, we look more closely at relative forecast performance throughout the out-of-sample period. In panels (a) to (d) of Fig. 2, we display the cumulative squared point forecast error difference between the autoregressive benchmark and (2.1) with \(net^-\) at the one-month ahead horizon. Periods when the plot line slopes upward represent periods in which the corresponding augmented model outperforms the benchmark, while downward-sloping segments indicate periods when the benchmark forecast is more accurate. Furthermore, in panels (a) and (b), we divide the out-of-sample period into pre 2020 and post 2020 period, such that we can better evaluate the nature of the predictive gains before COVID 19. Although out-of-sample performance is to a certain degree driven by the Great Recession and COVID 19, we observe that the cumulative squared point forecast error difference line has positive slope throughout the out-of-sample period. This in turn, explains why the no population-level predictability null hypothesis is also rejected over 1990m1–2007m12 for the specification. On the other hand, when reporting results for (2.1) with \(\Delta oil\) in Fig. 3, which is representative of the other models, we observe that (a): As indicated by periods with negative slope, there is evidence of overfitting, and (b): The increase in relative forecast accuracy on the onset of the Great Recession and COVID 19 pandemic is not as high as for (2.1) with the one-year net crude oil price decrease.

### 4.3 Economic Evaluation

Thus far, we have explored the magnitude and nature of the predictive impact of the price of crude oil on industrial production form a statistical viewpoint. In this section, we evaluate the nature of the predictive impact from an economic viewpoint. We do this in two different ways. To start with, we concentrate on the ability of the crude oil price (crude oil price volatility) models to predict the left tail of the conditional distribution \(h\)-month ahead. Thereafter, we explore whether one can use point forecasts produced under different models to generate the probability of a downturn in industrial production.

In our first test, we calculate the uncovered losses, defined as the distance between the realized value of \(y_{t+h}\) and the predicted lower confidence band produced using information up to time \(t\) when the realized value of \(y_{t+h}\) falls below the predicted

---

31 At \(h > 1\), the plots show a similar qualitative pattern.

32 Under the null hypothesis of no population-level predictability, the cumulative squared point forecast error difference trends steadily downward, as additional estimation error associated with the more heavily parameterized augmented model increases the cumulative squared point forecast error relative to the benchmark model.

33 Among the predictors, \(net^+\) is the only specification that does not afford any gains on the onset of COVID 19. However, this is because \(net^+ = 0\) since the mid 2019s till then end of the sample.
lower confidence band. More precisely, inspired by the Value-at-Risk literature, we consider the following loss function:

\[ \text{loss}(x) = \begin{cases} 0 & \text{if } x < \text{VaR}_\alpha \\alpha & \text{otherwise} \end{cases} \]

where \( \text{VaR}_\alpha \) is the Value-at-Risk at level \( \alpha \). This loss function is defined as the difference between the realized return and the VaR,天堂 if the realized return is below the VaR,天堂 otherwise.

In the table, TU reports the MSE produced under the specification of interest relative to the MSE produced the benchmark. CW reports \( p \)-values associated with the null hypothesis of no population-level predictability as specified in Clark and West (2007).
where \( \alpha = 0.01 \) is the confidence level, and \( \hat{\Lambda}_{t+h}^{\alpha} \) is the lower confidence band of \( y_{t+h} \) produced under the model of interest given information up to time \( t \). This quantity is simply computed as \( \hat{y}_{t+h} + \Phi^{-1}(\alpha)\hat{\sigma}_{t+h} \), where \( \Phi^{-1} \) is the inverse of the standard Normal cumulative distribution evaluated at \( \alpha \), and \( \hat{\sigma} \) is the conditional innovation volatility estimate from (2.1) or (2.2). As indicated in (4.2), the failure of the generated confidence band to cover the “loss”, i.e. when \( y_{t+h} < \hat{\Lambda}_{t+h}^{\alpha} \), is penalized more heavily than when it is more “conservative”, i.e. when \( y_{t+h} \geq \hat{\Lambda}_{t+h}^{\alpha} \).

In order to save space, we shall not report results across all measures. Rather, we report results for the most representative cases. In Table 7, we report the average (4.2) produced under the model of interest relative to the average (4.2) produced under the benchmark over various out-of-sample periods. Here, numbers lower than one indicate a better performance of the predictive model of interest relative to the benchmark. The linear model with \( \Delta oil \) performs well at \( h = 1 \) over 1990m1–2020m12 and 1990m1–2019m12, delivering accuracy gains of around 5% relative to the benchmark. This is understandable because contrary to 1990m1–2007m12, these out-of-sample periods contain two crises. However, as we increase the forecast horizon, we do not observe gains. Similar to the linear models, gains produced under (2.1) with \( net^- \) are also mostly limited to \( h = 1 \). However, the relative accuracy gains produced under this specifications can be as high as 11%.

Fig. 2 Cumulative squared point forecast error produced under (2.2) minus (2.1) with \( net^- \) at \( h = 1 \). In panels a to d of the figure, we display the cumulative squared point forecast error produced under the benchmark minus (2.1) with the one-year net crude oil price decrease at \( h = 1 \)
Lastly, (2.1) with gap is very representative how the remaining nonlinear models perform. It does not perform as well as (2.1) with net−, but it slightly outperforms the linear model, especially as h increases.

The next analysis focuses on the ability of the models to forecast the growth rate of the industrial production index falling below a certain threshold. Therefore, following Kilian and Manganelli (2008), we define the h-month ahead risk of a y% decrease in the industrial production growth rate as:

\[
DR_{t+h} = \int_{-\infty}^y (y - y_{t+h})^\gamma dF(y_{t+h}), \quad \gamma \geq 0, \tag{4.3}
\]

where \(\gamma\) is a measure of risk aversion and \(F(y_{t+h})\) is the cumulative distribution function of \(y_{t+h}\). Essentially, (4.3) measures the probability-weighted average loss when \(y_{t+1} \leq y\). We set \(\gamma = 2\) and \(y = -2\%\), such that we forecast the probability of a two percent downturn in industrial production, and we are risk adverse. As before, our estimate of \(DR_{t+h}\) is computed based on information up to time \(t\).  

35 The maximum decrease in the monthly industrial production index growth rate on the onset of the Great Recession is 5%, and on the onset of the COVID 19 pandemic is 13%.
Table 7  Average (4.2) produced under the model of interest relative to the average (4.2) produced under the benchmark

| IP data       | Ex-post revised | Real-time       |
|---------------|------------------|-----------------|
|               | 1990m1−2020m12   | 1990m1−2019m12 | 1990m1−2007m12 |
|               | 1990m1−2020m12   | 1990m1−2019m12 | 1990m1−2007m12 |
| Out-of-sample |                  |                 |                |
| $h$           | Relative loss    | Relative loss   | Relative loss  |
| 1             | 0.95             | 1.07            | 0.97           |
| 2             | 1.01             | 1.10            | 1.00           |
| 3             | 1.01             | 0.98            | 1.05           |
| 4             | 1.00             | 1.02            | 1.02           |
| 5             | 1.01             | 1.01            | 0.99           |
| 6             | 1.00             | 1.00            | 1.00           |
| 7             | 1.00             | 1.00            | 1.00           |
| 8             | 1.00             | 1.00            | 1.00           |

First difference of monthly log-crude oil price, $\Delta_{oil}$.

| $h$ | First difference of monthly log-crude oil price, $\Delta_{oil}$ | One-year net crude oil price decrease, $net^-$. | One-year crude oil price gap, $gap$. |
|-----|---------------------------------------------------------------------|---------------------------------------------|----------------------------------|
| 1   | 0.95                                                                | 0.90                                        | 0.95                             |
| 2   | 1.01                                                                | 1.00                                        | 1.00                             |
| 3   | 1.01                                                                | 1.00                                        | 1.00                             |
| 4   | 1.00                                                                | 1.00                                        | 1.00                             |
| 5   | 1.01                                                                | 1.00                                        | 1.00                             |
| 6   | 1.00                                                                | 1.00                                        | 1.00                             |
| 7   | 1.00                                                                | 0.99                                        | 1.00                             |
| 8   | 0.99                                                                | 1.00                                        | 1.00                             |

| $h$ | First difference of monthly log-crude oil price, $\Delta_{oil}$ | One-year net crude oil price decrease, $net^-$. | One-year crude oil price gap, $gap$. |
|-----|---------------------------------------------------------------------|---------------------------------------------|----------------------------------|
| 1   | 0.95                                                                | 0.90                                        | 0.95                             |
| 2   | 1.01                                                                | 1.00                                        | 1.00                             |
| 3   | 1.01                                                                | 1.00                                        | 1.00                             |
| 4   | 1.00                                                                | 1.00                                        | 1.00                             |
| 5   | 1.01                                                                | 1.00                                        | 1.00                             |
| 6   | 1.00                                                                | 1.00                                        | 1.00                             |
| 7   | 1.00                                                                | 0.99                                        | 1.00                             |
| 8   | 0.99                                                                | 1.00                                        | 1.00                             |
| IP data      | Ex-post revised | Real-time       |
|-------------|-----------------|-----------------|
| Out-of-sample | 1990m1–2020m12  | 1990m1–2019m12  | 1990m1–2007m12 |
| h           | Relative loss   | Relative loss   | Relative loss   | Relative loss   | Relative loss   | Relative loss   |
| 4           | 1.00            | 1.00            | 1.00            | 1.00            | 1.00            | 1.01            |
| 5           | 1.00            | 1.00            | 1.01            | 1.00            | 0.99            | 0.99            |
| 6           | 0.99            | 0.98            | 1.00            | 0.99            | 0.99            | 0.99            |
| 7           | 0.99            | 0.98            | 1.01            | 1.00            | 1.00            | 1.00            |
| 8           | 0.98            | 0.96            | 1.01            | 1.00            | 0.99            | 0.99            |

In the table, relative loss reports the average (4.2) produced under the model of interest relative to (4.2) produced under the benchmark over the out-of-sample period of interest.
In Fig. 4, we report (4.3) produced under the benchmark as well as selected specifications for $h = 1$. Accordingly, with regards to predicting downturns in industrial production on the onset of the Great Recession and the COVID 19 pandemic, (2.1) with the one-year net crude oil price decrease performs best. It correctly identifies the downturns towards the end of 2009 and the beginning of 2020. This pattern hold for both ex-post revised as well as real-time data. For example, it correctly forecasts the more than 2% decrease on 2008$m_{9}$ and from 2008$m_{12}$ till 2009$m_{1}$. In contrast, the competitors are less accurate in signaling the downturn in industrial production with respect to the Great Recession. Likewise, the increases in (4.3) in the beginning of 2020 are of less magnitude.

Another interesting feature of (2.1) with $net^-$ is that it also performs well in signaling downturns for $h > 1$. For example, in Fig. 5, we report results for (2.1) with $net^-$ at $h = 4$. Although the increases in (4.3) are not at the same magnitude as $h = 1$, we still observe that the model does a good job of signaling downturns in industrial production with regards to the Great Recession and the COVID 19 pandemic.
5 Robustness Checks

In this section, we carry out a number of robustness checks along several dimensions, which include how we have specified the predictive regression, the choice of the truncation lag of the nonlinear crude oil price measures, and the choice of the crude oil price series itself.

The first analysis focuses on how the predictive regressions are specified. As displayed from the equations in Section 2, our predictive regressions are a function of lagged values of the first difference of the logarithm of the U.S. industrial production index and the crude oil price or the crude oil price volatility measure of interest. Evidently, there is the possibility that our results are sensitive to an omitted variable in the predictive regression. Particularly, when specifying our predictive models as (2.2), we risk leaving out information possibly correlated with \( X_{t-i} \) in the conditional innovation. This omission can lead to biased OLS estimates due to violation of the exogeneity assumption, which can have severe consequences on the quality of the generated forecasts. This in turn, can lead us to wrong conclusions. Therefore, to examine this possibility, we follow Ravazzolo and Rothman (2013), Nonejad (2020b), Nonejad (2020c) and consider the following benchmark model:

\[
y_t = \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \hat{\beta}_i Z_{t-i} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2), \tag{5.1}
\]

where \( Z_{t-i} \) is a financial or macroeconomic variable at time \( t-1 \) not considered in our analysis thus far. As an alternative to (5.1), we consider the following predictive regression:

\[
y_t = \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \beta_i X_{t-i} + \sum_{i=1}^{q} \hat{\beta}_i Z_{t-i} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2), \tag{5.2}
\]

The lag orders in (5.1) and (5.2) are chosen using the same approach as in Section 2. By performing the out-of-sample analysis using (5.1) and (5.2), we can evaluate to what extent our results in Section 4.2 are affected by an omitted variable in the predictive regression. With regards to the choice of \( Z_{t-i} \), Ravazzolo and Rothman (2013) suggest that one should consider variables that are related to business conditions and are arguably also correlated with \( X_{t-i} \). We follow their advice and consider the following variables: The first difference of the logarithm of global crude oil production, the first difference of the logarithm of U.S. ending stocks of crude oil, Moody’s Baa-Aaa spread and Kilian’s real global economic activity index. The first two time-series are extracted from the U.S. Energy Information Administration’s website: https://www.eia.gov/, whereas the remaining variables are extracted from the FRED database: https://fred.stlouisfed.org/36.

\[\text{Kilian (2009)’s index of global economic activity is constructed using data on various bulk dry cargoes shipping freight rates. The basic idea behind this measure is that changes in world economic activity drive demand for shipping and in the short-run this higher demand shows up as an increase in the real cost of shipping. In a recent study, Kilian and Zhou (2018) compare several measures of global real economic activity and find that Kilian’s measure performs very well compared to them.}\]
We conduct our out-of-sample analysis using (5.1) and (5.2), where \( Z_t - i \) corresponds to the mentioned variables one by one. Since results are generally similar across the choice of \( Z_t - i \), we report results for a representative case, namely, where the crude oil price variables are \( \Delta \text{oil} \) and \( \text{net}^{-} \), and \( Z_t - i \) is the first difference of the logarithm of global crude oil production. Results are presented in Table 8. Overall, compared to results reported in Section 2, we do not observe changes worth mentioning, meaning that the models in Section 2 are correctly specified.

Thus far, the nonlinear measures relied on in this study have been based on nonlinear transformations of the price of crude oil. However, there are alternatives ways of accounting for nonlinearities. A very popular approach is Hamilton (1989)’s Markov-switching specification, where the vector of model parameters evolve as a \( m \)-state Markov chain. Therefore, we consider (2.1) with \( \Delta \text{oil} \) (log-RV), and assume that the vector of the regression coefficients, \( \theta = (\phi_0, \phi_1, ..., \phi_p, \beta_1, ..., \beta_q) \), follows a two-state Markov-switching (MS) process as suggested in Hamilton (1989). This in turn, implies that nonlinearity is incorporated by the magnitude of the regression coefficients in different regimes. We conduct our out-of-sample forecasting using this specification, and compare the generated point forecasts with point forecasts produced under the benchmark. As displayed by results reported in Table 9, the MS specifications do not deliver the same degree of forecast improvements as the model with the one-year net crude oil price decrease.

![Fig. 5](image_url) Four-months ahead probability of a 2% downturn in industrial production data for (2.1) with the one-year net crude oil price decrease, \( \text{net}^{-} \). In panels a to d of the figure, we display the four-months ahead probability of a 2% downturn in the U.S industrial production index growth rate produced under the predictive regression with \( \text{net}^{-} \)
The crude oil price (crude oil price volatility)-based predictive regressions considered in this study thus far have been of the form (2.1), where besides the lagged values of $y_t$, we include lagged values of the crude oil price (crude oil price volatility) measure of interest. However, in a similar fashion as Herrera et al. (2011) and Kilian and Vigfusson (2013), we can combine the linear measures with the nonlinear measures. In other words, we can extend (2.1) as follows:

$$y_t = \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} X_{t-i} \beta_i + \sum_{i=1}^{\hat{q}} \triangle oil_{t-i} \tilde{\beta}_i + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2). \quad (5.3)$$

Intuitively, one would expect a better out-of-sample performance from (5.3) relative to (2.1) because we are combining more predictors. However, we observe that this is not the case due to the bias-variance trade-off.\(^{37}\) In fact, we observe that this

\(^{37}\) With the term bias-variance trade-off, we mean that there is both a benefit and a cost of using (5.3) instead of (2.1). The benefit is the additional predictors, $\triangle oil_{t-i}, i = 1, \ldots, \hat{q}$. The cost is higher forecast variability related to the need of estimating the additional parameters, namely, $\beta_1, \ldots, \beta_{\hat{q}}$. Clearly, the lat-

### Table 8 Population-level predictability evaluation results using (5.2) relative to (5.1), where $Z_{t-i}$ is the first difference of the logarithm of global crude oil production

| IP data                  | Ex-post revised | Real-time |
|--------------------------|-----------------|-----------|
|                          | 1990m1–2020m12  | 1990m1–2007m12 |
| Out-of-sample            | TU CW          | TU CW     |
|                          | TU CW          | TU CW     |
|                          | TU CW          | TU CW     |
|                          | TU CW          | TU CW     |
|                          | TU CW          | TU CW     |
|                          | TU CW          | TU CW     |
|                          | TU CW          | TU CW     |
|                          | TU CW          | TU CW     |

First difference of monthly log-crude oil price, $\triangle oil$.

| $h$ | TU | CW | TU | CW | TU | CW | TU | CW | TU | CW | TU | CW |
|-----|----|----|----|----|----|----|----|----|----|----|----|----|
| 1   | 0.94 | 0.07 | 0.98 | 0.03 | 1.00 | 0.19 | 0.97 | 0.10 | 1.00 | 0.18 | 1.01 | 0.72 |
| 2   | 1.00 | 0.45 | 1.00 | 0.30 | 1.02 | 0.86 | 1.02 | 0.81 | 1.00 | 0.29 | 1.02 | 0.88 |
| 3   | 1.01 | 0.77 | 1.01 | 0.43 | 0.99 | 0.10 | 1.03 | 0.90 | 1.01 | 0.58 | 1.04 | 0.91 |
| 4   | 1.00 | 0.50 | 1.00 | 0.27 | 1.01 | 0.56 | 1.02 | 0.83 | 1.00 | 0.22 | 1.03 | 0.81 |
| 5   | 1.00 | 0.54 | 1.01 | 0.62 | 1.00 | 0.28 | 1.01 | 0.86 | 1.01 | 0.68 | 1.01 | 0.63 |
| 6   | 1.00 | 0.67 | 1.01 | 0.72 | 1.01 | 0.79 | 1.00 | 0.72 | 1.01 | 0.67 | 1.00 | 0.38 |
| 7   | 1.00 | 0.94 | 1.01 | 0.89 | 1.01 | 0.95 | 1.00 | 0.67 | 1.00 | 0.78 | 1.01 | 0.77 |
| 8   | 1.00 | 0.56 | 1.00 | 0.53 | 1.01 | 0.69 | 1.00 | 0.72 | 1.01 | 0.88 | 1.01 | 0.77 |

One-year net crude oil price decrease, net$^{-}$.  

| $h$ | TU | CW | TU | CW | TU | CW | TU | CW | TU | CW | TU | CW |
|-----|----|----|----|----|----|----|----|----|----|----|----|----|
| 1   | 0.92 | 0.08 | 0.92 | 0.00 | 0.96 | 0.00 | 1.00 | 0.18 | 0.93 | 0.00 | 0.97 | 0.00 |
| 2   | 1.02 | 0.74 | 0.99 | 0.03 | 1.00 | 0.73 | 1.04 | 0.71 | 0.97 | 0.02 | 1.01 | 0.92 |
| 3   | 1.00 | 0.43 | 1.00 | 0.08 | 1.00 | 0.38 | 1.05 | 0.77 | 0.99 | 0.07 | 1.02 | 0.96 |
| 4   | 1.00 | 0.54 | 1.00 | 0.10 | 1.00 | 0.83 | 1.02 | 0.68 | 1.00 | 0.11 | 1.02 | 0.90 |
| 5   | 1.00 | 0.76 | 1.00 | 0.54 | 1.00 | 0.66 | 1.00 | 0.53 | 1.00 | 0.22 | 1.01 | 0.68 |
| 6   | 1.00 | 0.74 | 1.00 | 0.69 | 1.00 | 0.78 | 1.00 | 0.89 | 1.00 | 0.83 | 1.00 | 0.92 |
| 7   | 1.00 | 0.27 | 1.00 | 0.70 | 1.00 | 0.61 | 1.00 | 0.77 | 1.00 | 0.81 | 1.00 | 0.71 |
| 8   | 0.99 | 0.08 | 0.99 | 0.19 | 1.00 | 0.26 | 1.00 | 0.20 | 1.00 | 0.48 | 1.00 | 0.73 |

In the table, TU reports the MSE produced under the specification of interest relative to the MSE produced the benchmark. CW reports p-values associated with the null hypothesis of no population-level predictability as specified in Clark and West (2007).
is the case. For example, in the top part of Table 10, we report results produced under (5.3), where $X_{t-i}$ corresponds to the one-year net crude oil price decrease relative (2.1). Here, we observe point forecast accuracy deterioration relative to Table 3. In the bottom part, we report result for (5.3), where instead of the crude oil price measures, we consider the logarithm of realized volatility and the logarithm of (4.1), Again, we observe sizable deterioration in relative MSE.

The next analysis focuses on the choice of the truncation lag, $m$. Particularly, in our analysis, we have set $m = 12$. Therefore, we experiment with what happens when we increase the truncation lag. We consider (2.1) with $net^-$, and report out-of-sample results relative to the benchmark, where $m = 36$, such that $net^-$ corresponds to the three-year net crude oil price decrease. Results in Table 11 indicate

| IP data           | Ex-post revised | Real-time         |
|-------------------|-----------------|-------------------|
|                    | 1990m1–2020m1   | 1990m1–2020m2     |
|                   | 1990m1–2019m1   | 1990m1–2019m2     |
|                   | 1990m1–2007m1   | 1990m1–2007m2     |
| $h$               | TU   | CW   | TU   | CW   | TU   | CW   | TU   | CW   |
| 1                 | 0.97 | 0.14 | 0.98 | 0.04 | 1.00 | 0.08 | 0.97 | 0.12 | 1.00 | 0.12 | 1.00 | 0.33 |
| 2                 | 1.02 | 0.80 | 1.01 | 0.66 | 1.01 | 0.84 | 1.04 | 0.90 | 1.01 | 0.82 | 1.01 | 0.96 |
| 3                 | 1.00 | 0.38 | 1.01 | 0.71 | 1.01 | 0.47 | 1.01 | 0.19 | 1.03 | 0.81 | 1.06 | 0.86 |
| 4                 | 1.00 | 0.23 | 1.01 | 0.92 | 1.01 | 0.89 | 1.01 | 0.60 | 1.00 | 0.31 | 1.03 | 0.82 |
| 5                 | 1.00 | 0.65 | 1.01 | 0.48 | 1.00 | 0.13 | 1.00 | 0.34 | 1.01 | 0.58 | 1.01 | 0.75 |
| 6                 | 1.00 | 0.73 | 1.01 | 0.89 | 1.03 | 0.86 | 1.00 | 0.43 | 1.01 | 0.68 | 1.01 | 0.57 |
| 7                 | 1.00 | 0.79 | 1.01 | 0.88 | 1.01 | 0.86 | 1.00 | 0.86 | 1.01 | 0.80 | 1.00 | 0.62 |
| 8                 | 1.00 | 0.26 | 1.00 | 0.91 | 1.00 | 0.45 | 1.00 | 0.42 | 1.00 | 1.00 | 1.00 | 0.98 |

| TU | CW | TU | CW | TU | CW | TU | CW |
|----|----|----|----|----|----|----|----|
| 1  | 1.00 | 0.84 | 1.00 | 0.78 | 1.01 | 0.76 | 1.00 | 0.57 | 1.00 | 0.94 | 1.00 | 0.81 |
| 2  | 1.00 | 0.89 | 1.00 | 0.92 | 1.01 | 0.84 | 1.00 | 0.63 | 1.00 | 0.64 | 1.00 | 0.46 |
| 3  | 1.00 | 0.95 | 1.00 | 0.97 | 1.01 | 0.94 | 1.00 | 0.57 | 1.02 | 0.75 | 1.02 | 0.72 |
| 4  | 1.00 | 0.87 | 1.00 | 0.92 | 1.01 | 0.90 | 1.00 | 0.67 | 1.02 | 0.86 | 1.02 | 0.86 |
| 5  | 1.00 | 0.88 | 1.00 | 0.78 | 1.00 | 0.35 | 1.00 | 0.45 | 1.01 | 0.70 | 1.01 | 0.67 |
| 6  | 1.00 | 0.23 | 1.00 | 0.20 | 1.00 | 0.82 | 1.00 | 0.17 | 1.00 | 0.48 | 1.01 | 0.63 |
| 7  | 1.00 | 0.24 | 1.00 | 0.86 | 1.00 | 0.76 | 1.00 | 0.19 | 1.00 | 0.78 | 1.00 | 0.81 |
| 8  | 1.00 | 0.38 | 1.00 | 0.48 | 1.00 | 0.67 | 1.00 | 0.70 | 1.01 | 0.68 | 1.01 | 0.60 |

First difference of monthly log-crude oil price, $\Delta oil$.

Logarithm of monthly crude oil price realized volatility, $\log (RV)$.

In the table, TU reports the MSE produced under the specification of interest relative to the MSE produced the benchmark. CW reports p-values associated with the null hypothesis of no population-level predictability as specified in Clark and West (2007).
that increasing \( m \) to three-years deteriorates out-of-sample performance relative to the benchmark.

The final analyses have to do with the choice of the crude oil price series, and whether we rely on the real versus the nominal price of crude oil. Again, we consider (2.1) with \( \text{net}^- \), where \( m = 12 \), and report out-of-sample results relative to the benchmark using nominal RAC for imported crude oil. We report results in Table 12. Overall, comparing these results with those reported in Table 3 document that the choice of the crude oil price series has very little bearing on our main conclusions. We reach similar qualitative conclusions for the remaining models. Lastly, we explore to what degree results change if we use the real price of crude oil instead of the nominal price. Therefore, we deflate nominal WTI prices with the U.S. CPI extracted from the FRED database. Results reported in Table 13 show that in accordance with the arguments outlined in Hamilton (2011b), using the real price of crude oil deteriorates forecast performance. The magnitude of the deterioration increases as we increase \( h \).

### Table 10: Population-level predictability evaluation results using (5.3)

| IP data | Ex-post revised | Real-time |
|---------|----------------|-----------|
| Out-of-sample | 1990m1–2020m12 | 1990m1–2007m12 |
|          | 2019m12       | 2007m12   |
| \( h \) | TU       | CW       | TU       | CW       | TU       | CW       | TU       | CW       | TU       | CW       |
| 1       | 0.93  0.08 | 0.94  0.00 | 0.98  0.01 | 1.00  0.19 | 0.95  0.00 | 0.97  0.01 | 1.00  0.19 | 0.95  0.00 | 0.97  0.01 | 1.00  0.19 |
| 2       | 1.01  0.53 | 1.00  0.17 | 1.03  0.88 | 1.03  0.64 | 0.98  0.06 | 1.02  0.75 | 1.03  0.64 | 0.98  0.06 | 1.02  0.75 | 1.03  0.64 |
| 3       | 1.00  0.39 | 1.01  0.37 | 1.00  0.19 | 1.06  0.75 | 1.02  0.20 | 1.05  0.52 | 1.06  0.75 | 1.02  0.20 | 1.05  0.52 | 1.06  0.75 |
| 4       | 1.01  0.59 | 1.00  0.26 | 1.01  0.47 | 1.02  0.63 | 1.01  0.14 | 1.03  0.50 | 1.02  0.63 | 1.01  0.14 | 1.03  0.50 | 1.02  0.63 |
| 5       | 1.01  0.75 | 1.01  0.52 | 1.01  0.30 | 1.01  0.69 | 1.01  0.32 | 1.02  0.40 | 1.01  0.69 | 1.01  0.32 | 1.02  0.40 | 1.01  0.69 |

\( \triangle \text{oil and net}^- \).

\[ \log(RV) \text{ and log-realized semivolatilities.} \]

| 1       | 1.00  0.31 | 0.97  0.00 | 0.99  0.04 | 1.02  0.48 | 0.96  0.00 | 0.99  0.07 | 1.02  0.48 | 0.96  0.00 | 0.99  0.07 | 1.02  0.48 |
| 2       | 1.03  0.71 | 0.98  0.00 | 1.01  0.39 | 1.06  0.74 | 0.99  0.01 | 1.02  0.35 | 1.06  0.74 | 0.99  0.01 | 1.02  0.35 | 1.06  0.74 |
| 3       | 1.03  0.79 | 0.98  0.01 | 1.00  0.20 | 1.04  0.74 | 1.02  0.20 | 1.07  0.84 | 1.04  0.74 | 1.02  0.20 | 1.07  0.84 | 1.04  0.74 |
| 4       | 1.00  0.70 | 1.02  0.99 | 1.03  0.99 | 1.01  0.45 | 1.02  0.76 | 1.05  0.96 | 1.01  0.45 | 1.02  0.76 | 1.05  0.96 | 1.01  0.45 |
| 5       | 1.00  0.58 | 1.00  0.44 | 1.01  0.64 | 1.01  0.49 | 1.00  0.16 | 1.02  0.72 | 1.01  0.49 | 1.00  0.16 | 1.02  0.72 | 1.01  0.49 |
| 6       | 1.00  0.27 | 1.01  0.98 | 1.02  0.98 | 1.00  0.20 | 1.00  0.35 | 1.00  0.33 | 1.00  0.20 | 1.00  0.35 | 1.00  0.33 | 1.00  0.20 |
| 7       | 1.01  0.68 | 1.02  0.78 | 1.03  0.84 | 1.01  0.83 | 1.01  0.77 | 1.02  0.89 | 1.01  0.83 | 1.01  0.77 | 1.02  0.89 | 1.01  0.83 |
| 8       | 1.00  0.33 | 1.01  0.30 | 1.01  0.22 | 1.01  0.63 | 1.01  0.57 | 1.00  0.15 | 1.01  0.63 | 1.01  0.57 | 1.00  0.15 | 1.01  0.63 |

In the table, TU reports the MSE produced under the specification of interest relative to the MSE produced the benchmark. CW reports \( p \)-values associated with the null hypothesis of no population-level predictability as specified in Clark and West (2007).
**Table 11** Population-level predictability evaluation results using the three-year net crude oil price decrease

| IP data          | Ex-post revised | Real-time       |
|------------------|-----------------|-----------------|
|                  | 1990m1–2020m12  | 1990m1–2020m12  |
|                  | 1990m1–2019m12  | 1990m1–2019m12  |
|                  | 1990m1–2007m12  | 1990m1–2007m12  |
|                  | TU   | CW   | TU   | CW   | TU   | CW   | TU   | CW   | TU   | CW   | TU   | CW   |
|                  | 0.98  | 0.13 | 0.94  | 0.01 | 0.96  | 0.00 | 1.00  | 0.21 | 0.98  | 0.00 | 0.98  | 0.01 |
|                  | 0.99  | 0.12 | 1.00  | 0.44 | 1.00  | 0.68 | 1.00  | 0.41 | 1.00  | 0.15 | 1.01  | 0.93 |
|                  | 0.98  | 0.20 | 1.00  | 0.21 | 1.00  | 0.99 | 1.02  | 0.53 | 0.99  | 0.04 | 1.01  | 0.70 |
|                  | 0.99  | 0.18 | 1.00  | 0.37 | 1.00  | 0.86 | 1.01  | 0.48 | 1.00  | 0.16 | 1.02  | 0.59 |
|                  | 1.00  | 0.35 | 1.00  | 0.70 | 1.00  | 0.22 | 0.99  | 0.08 | 1.00  | 0.21 | 1.00  | 0.58 |
|                  | 0.99  | 0.11 | 1.01  | 0.28 | 1.01  | 0.68 | 1.00  | 0.12 | 1.00  | 0.67 | 1.00  | 0.85 |
|                  | 0.99  | 0.07 | 1.00  | 0.19 | 1.00  | 0.45 | 1.00  | 0.18 | 1.00  | 0.58 | 0.99  | 0.91 |
|                  | 1.00  | 0.10 | 1.00  | 0.24 | 1.00  | 0.22 | 1.00  | 0.49 | 1.00  | 0.69 | 1.00  | 0.89 |

In the table, TU reports the MSE produced under the specification of interest relative to the MSE produced the benchmark. CW reports p-values associated with the null hypothesis of no population-level predictability as specified in Clark and West (2007).

**Table 12** Population-level predictability evaluation results using the one-year net crude oil price decrease and nominal RAC for imported crude oil

| IP data          | Ex-post revised | Real-time       |
|------------------|-----------------|-----------------|
|                  | 1990m1–2020m12  | 1990m1–2020m12  |
|                  | 1990m1–2019m12  | 1990m1–2019m12  |
|                  | 1990m1–2007m12  | 1990m1–2007m12  |
|                  | TU   | CW   | TU   | CW   | TU   | CW   | TU   | CW   | TU   | CW   | TU   | CW   |
|                  | 0.92  | 0.05 | 0.94  | 0.00 | 0.96  | 0.00 | 0.92  | 0.06 | 0.94  | 0.01 | 0.97  | 0.01 |
|                  | 1.01  | 0.53 | 0.99  | 0.05 | 1.00  | 0.45 | 1.03  | 0.65 | 0.97  | 0.02 | 1.01  | 0.91 |
|                  | 1.01  | 0.81 | 0.98  | 0.06 | 1.00  | 0.35 | 1.05  | 0.73 | 0.98  | 0.04 | 1.01  | 0.77 |
|                  | 1.01  | 0.77 | 0.99  | 0.05 | 1.00  | 0.86 | 1.02  | 0.73 | 0.99  | 0.01 | 1.00  | 0.76 |
|                  | 1.00  | 0.77 | 1.00  | 0.57 | 1.00  | 0.66 | 1.01  | 0.87 | 1.01  | 0.63 | 1.00  | 0.86 |
|                  | 1.00  | 0.15 | 1.01  | 0.21 | 1.00  | 0.33 | 1.00  | 0.85 | 1.00  | 0.73 | 1.00  | 0.82 |
|                  | 1.00  | 0.39 | 1.01  | 0.58 | 1.00  | 0.22 | 1.00  | 0.84 | 1.00  | 0.72 | 1.00  | 0.56 |
|                  | 1.00  | 0.12 | 1.00  | 0.23 | 1.00  | 0.24 | 1.00  | 0.52 | 1.00  | 0.72 | 1.00  | 0.88 |

In the table, TU reports the MSE produced under the specification of interest relative to the MSE produced the benchmark. CW reports p-values associated with the null hypothesis of no population-level predictability as specified in Clark and West (2007).
Conclusion

More than ten years after Herrera et al. (2011), which focused on evaluating the nonlinear in-sample predictive impact of the price of crude oil on the U.S. industrial production index, this study explores whether the out-of-sample predictive impact of the price of crude oil (crude oil price volatility) on the U.S. industrial production index growth rate is nonlinear, and if so, which forms of nonlinearity help drive forecast performance the most relative to the benchmark. The out-of-sample analysis is conducted using ex-post revised as well as real-time industrial production index data.

Our findings have important implications not only for applied forecasters, but also for economists interested in modeling the transmission of crude oil price shocks on industrial production.
consider the net crude oil price decrease. Hence, it is worth an effort to reproduce the analysis in Herrera et al. (2011), Herrera et al. (2015) and explore whether performing the in-sample predictability tests in Herrera et al. (2011) using the net crude oil price decrease leads to any changes regarding the main conclusions. Finally, it must be mentioned that the nonlinear crude oil price (crude oil price volatility) models used in this study are agnostic about whether the crude oil price movements themselves are due to demand shocks, supply shocks or any other reason. Evidently, it would be beneficial to include such information as it could further help enhance the predictive power contained in the nonlinear variables. In this regard, the recent study of Nonejad (2021), which uses relative predictive performance provides some possible avenues for future research.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Human participants This article does not contain any studies with human participants or animals performed by the author.

References

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.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica, 71*, 529–626.

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.

Barndorff-Nielsen, O. E., & Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B, 64*, 253–280.

Baumeister, C., & Hamilton, J. D. (2019). Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks. *American Economic Review, 109*, 1873–1910.

Bresnahan, T. F., & Ramey, V. A. (1993). Segment shifts and capacity utilization in the U.S. automobile industry. *American Economic Review Papers and Proceedings, 83*, 213–218.

Carlton, A. B. (2010). Oil prices and real-time output growth. Working Paper, University of Houston.

Clark, T. E., & McCracken, M. W. (2009). Tests of equal predictive ability with real-time data. *Journal of Business and Economic Statistics, 27*, 441–54.

Clark, T. E., & McCracken, M. W. (2013). Advances in Forecast Evaluationin A. Timmermann and G. Elliott (editors), Handbook of economic forecasting. Elsevier, Amsterdam.

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

Davis, S. J., & Haltiwanger, J. (2001). Sectoral job creation and destruction responses to oil price changes. *Journal of Monetary Economics, 48*, 465–512.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics, 13*, 253–63.

Dvir, E., & Rogoff, K. (2010). Three epochs of oil. mimeo, Harvard University.

Hamilton, J. D. (1983). Oil and the macroeconomy since world war II. *The Journal of Political Economy, 9*, 228–248.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica, 57*, 357–384.
Hamilton, J. D. (1996). This is what happened to the oil price-macroeconomy relationship. *Journal of Monetary Economics, 38*, 215–220.

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.

Herrera, A. M., Lagalo, L. G., & Wada, T. (2011). Oil price shocks and industrial production: Is the relationship linear? *Macroeconomic Dynamics, 15*, 472–497.

Herrera, A. M., Lagalo, L. G., & Wada, T. (2015). Asymmetries in the response of economic activity to oil price increases and decreases? *Journal of International Money and Finance, 50*, 108–133.

Inoue, A., & Kilian, L. (2004). In-sample or out-of-sample tests of predictability: Which one should we use? *Econometric Reviews, 23*, 371–402.

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

Kilian, L., & Manganelli, S. (2008). The central banker as a risk manager: Estimating the federal reserve’s preferences under greenspan. *Journal of Money, Credit and Banking, 40*, 1103–1129.

Kilian, L., & Vigfusson, R. J. (2011). Nonlinearities in the oil price-output relationship. *Macroeconomics 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 and 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.

Koenig, E. F., Dolmas, S., & Piger, J. (2003). The use and abuse of real-time data in economic forecasting. *Review of Economics and Statistics, 85*, 618–28.

Lopez, J. A. (1998). Testing your risk tests. *Financial Survey*, May–June:18–20.

Lopez, J. A. (1999). Methods for evaluating value-at-risk estimates. *Economic Review-Federal Reserve Bank of San Francisco, 2*, 3–17.

Marcellino, M., Stock, J. H., & Watson, M. W. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics, 135*, 499–526.

Mork, K. A. (1989). Oil and the macroeconomy when prices go up and down: An extension of Hamilton’s results. *Journal of Political Economy, 97*, 740–744.

Narayan, P. K., & Gupta, R. (2015). Has oil price predicted stock returns for over a century? *Energy Economics, 48*, 18–23.

Nonejad, N. (2019). Modeling persistence and parameter instability in historical crude oil price data using a Gibbs sampling approach. *Computational Economics, 53*, 1687–1710.

Nonejad, N. (2020). Crude oil price volatility and short-term predictability of the real U.S. GDP growth rate. Economic Letters. Forthcoming.

Nonejad, N. (2020). Should crude oil price volatility receive more attention than the price of crude oil? An empirical investigation via a large-scale out-of-sample forecast evaluation of U.S. macroeconomic data. Journal of Forecasting. Forthcoming.

Nonejad, N. (2020c). Crude oil price changes and the United Kingdom real gross domestic product growth rate: An out-of-sample investigation. The Journal of Economic Asymmetries. Forthcoming.

Nonejad, N. (2021). The price of crude oil and (conditional) out-of-sample predictability of world industrial production. Journal of Commodity Markets. Forthcoming.

Patton, A. J., & Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics, 97*, 683–697.

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 and Econometrics, 20*, 441–453.

Rossi, B., & Inoue, A. (2012). Out-of-sample forecast tests robust to the choice of window size. *Journal of Business and Economic Statistics, 30*, 432–453.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.